



EDWIN GUIDO BOZA CONDORENA

**“IDENTIFICAÇÃO DE PROCESSOS E CONTROLE
PREDITIVO COM MODELO UTILIZANDO TÉCNICAS DE
INTELIGÊNCIA ARTIFICIAL APLICADAS À PRODUÇÃO
DE BIOETANOL”**

CAMPINAS
2012



UNIVERSIDADE ESTADUAL DE CAMPINAS

FACULDADE DE ENGENHARIA QUÍMICA

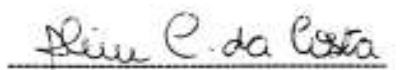
EDWIN GUIDO BOZA CONDORENA

**“IDENTIFICAÇÃO DE PROCESSOS E CONTROLE PREDITIVO
COM MODELO UTILIZANDO TÉCNICAS DE INTELIGÊNCIA
ARTIFICIAL APLICADAS À PRODUÇÃO DE BIOETANOL”**

Orientador: Profa. Dra. Aline Carvalho da Costa.

Tese de Doutorado apresentada à Faculdade de Engenharia Química da Universidade Estadual de Campinas para a obtenção do título de Doutor em Engenharia Química.

Este exemplar corresponde à versão final da Tese defendida pelo aluno Edwin Guido Boza Condorena, e orientada pela Profa. Dra. Aline Carvalho da Costa.


Profa. Dra. Aline Carvalho da Costa.

CAMPINAS
2012

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

B716i Boza Condorena, Edwin Guido, 1958-
Identificação de processos e controle preditivo com
modelo utilizando técnicas de inteligência artificial
aplicadas à produção de bioetanol / Edwin Guido Boza
Condorena. --Campinas, SP: [s.n.], 2012.

Orientador: Aline Carvalho da Costa.
Tese de Doutorado - Universidade Estadual de
Campinas, Faculdade de Engenharia Química.

1. Bioetanol. 2. Inteligência artificial. 3. Redes
neurais (Computação). 4. Processos de fabricação -
Automação. 5. Controle preditivo. I. Costa, Aline
Carvalho da, 1970-. II. Universidade Estadual de
Campinas. Faculdade de Engenharia Química. III.
Título.

Título em Inglês: Process identification and model predictive control using artificial
intelligence techniques applied to bioethanol production

Palavras-chave em Inglês: Bioethanol, Artificial intelligence, Neural networks
(Computers), Manufacturing processes - Automation,
Predictive control

Área de concentração: Desenvolvimento de Processos Químicos

Titulação: Doutor em Engenharia Química

Banca examinadora: Flavio Vasconcelos da Silva, Delba Nisi Cosme Melo, Maria
Isabel Berto, Adilson José de Assis

Data da defesa: 17-09-2012

Programa de Pós Graduação: Engenharia Química

Tese de Doutorado defendida por Edwin Guido Boza Condorena e aprovada em 17 de Setembro de 2012 pela banca examinadora constituída pelos doutores:

Aline C. da Costa

Prof^a. Dr^a. Aline Carvalho da Costa – Orientadora

[Assinatura]

Prof. Dr. Flávio Vasconcelos da Silva

Delba Nisi Cosme Melo

Dr^a. Delba Nisi Cosme Melo – UNICAMP

[Assinatura]

Dr^a. Maria Isabel Berto - ITAL

Adilson José de Assis

Prof. Dr. Adilson José de Assis – FEQ/UFU

AGRADECIMENTOS

Agradeço a Deus por sua presença, providência e graça, que me sustentaram durante todo o tempo.

O presente trabalho também não teria sido possível sem o apoio de pessoas e instituições às quais manifesto meu mais sincero agradecimento.

Agradeço as observações e orientações feitas pela Profa. Dra. Aline Carvalho da Costa quem mostrou muito interesse e dedicação para o melhor desenvolvimento do presente trabalho.

Agradeço aos professores Maria Alvina e José Nunhez pela confiança depositada para conseguir a finalização dos estudos.

A Nina Eleonor companheira da minha vida e Kate minha princesinha pelo apoio constante nesta nova experiência de aprendizagem e pela sua presença oportuna que tornou mais agradável a estadia neste belo país.

À UNICAMP pela oportunidade proporcionada e à CAPES pelo apoio econômico para a conclusão dos estudos e a realização da pesquisa.

Enfim, agradeço a todas as pessoas que de alguma forma contribuíram para que este trabalho fosse realizado.

Muito obrigado!

SUMÁRIO

RESUMO	27
ABSTRACT	28
CAPÍTULO 1. INTRODUÇÃO	29
1.2 . OBJETIVOS.....	31
1.3. ORGANIZAÇÃO DO TRABALHO	32
1.4. REFERÊNCIAS BIBLIOGRÁFICAS	36
CAPÍTULO 2. A PRODUÇÃO DE BIOETANOL.....	38
2.1. INTRODUÇÃO.....	39
2.2. O BIOETANOL DE 1ª GERAÇÃO	40
2.2.1. A produção de álcool a partir da cana-de-açúcar	43
2.2.1.1. <i>Preparo da matéria prima</i>	44
2.2.1.2. <i>Reação</i>	44
2.2.1.3. <i>Separação</i>	48
2.2.1.4. <i>Os processos de fermentação</i>	49
2.2.1.4.1. Processo batelada	49
2.2.1.4.2. Processo batelada alimentada	50
2.2.1.4.3. Processo contínuo	51
2.3. O BIOETANOL DE 2ª GERAÇÃO	57
2.3.1. Pré-tratamento	58
2.3.2. Hidrólise.....	59
2.3.2.1. <i>Hidrólise ácida</i>	59
2.3.2.2. <i>Hidrólise enzimática</i>	60
2.4. A INTEGRAÇÃO DE PROCESSOS NA PRODUÇÃO DE BIOETANOL	61
2.5. INTEGRAÇÃO FERMENTAÇÃO- SEPARAÇÃO	62
2.5.1. Fermentação - remoção de etanol por vácuo.....	63
2.6. A MODELAGEM MATEMÁTICA DO PROCESSO DE FERMENTAÇÃO COM EXTRAÇÃO CONTÍNUA DE ETANOL UTILIZANDO SEPARADOR FLASH	64

2.6.1. Caracterização dos modelos matemáticos utilizados na fermentação para produção de bioetanol	65
2.6.1.1. O tipo dos modelos cinéticos	65
2.6.1.2. Características da modelagem cinética do processo de fermentação	67
2.6.1.2.1. Cinética de crescimento da levedura.....	67
2.6.1.2.2. Cinética de morte dos microorganismos	74
2.6.1.2.3. Cinética de produção de etanol.....	75
2.6.1.2.4. Cinética de consumo de substrato	76
2.6.1.2.5. Efeito da temperatura sobre os parâmetros cinéticos.....	77
2.6.1.3. Características dos modelos utilizados para a separação flash	78
2.6.1.3.1. Modelos para cálculo do equilíbrio L-V.....	78
2.6.1.4. Condições de operação do flash com base em modelos de equilíbrio	80
2.6.2. Observações referidas à modelagem matemática da produção de etanol por via fermentativa.....	82
2.7. CONCLUSÃO	83
2.8. REFERÊNCIAS BIBLIOGRÁFICAS	84
 CAPÍTULO 3. REDES NEURAIS ARTIFICIAIS	97
3.1. INTRODUÇÃO.....	98
3.2. CONCEITO E CARACTERÍSTICAS	99
3.2.1. Tipos de funções de ativação.....	103
3.2.1.1. Função sigmoide.....	103
3.2.1.1.1. Função sigmoide logística.....	104
3.2.1.1.2. Função tangente hiperbólica	104
3.2.1.2. Função linear	105
3.2.1.3. Função de limiar	105
3.2.1.4. Função gaussiana.....	106
3.2.2. Topologia das redes neurais artificiais	107
3.2.2.1. Rede neural artificial tipo alimentada adiante ou feedforward	108
3.2.2.2. Rede neural artificial tipo recorrente ou feedback	108
3.3. PERCEPTRONS DE MÚLTIPLAS CAMADAS.....	108
3.3.1. O algoritmo backpropagation	109
3.4. A GENERALIZAÇÃO	114
3.5. O PAPEL DAS UNIDADES ESCONDIDAS DOS PERCEPTRONS	117

3.6. APLICAÇÕES DAS REDES NEURAIS: PONTOS FORTES E FRACOS	121
3.7. CONCLUSÃO	123
3.9. REFERÊNCIAS BIBLIOGRÁFICAS	124

CAPÍTULO 4. MODELAGEM E IDENTIFICAÇÃO DE PROCESSOS

4.1. INTRODUÇÃO	130
4.2. IDENTIFICAÇÃO DE PROCESSOS	131
4.2.1. Definição do problema	132
4.2.2. Coleta e pré-tratamento dos dados	132
4.2.3. Seleção da estrutura do modelo	136
4.2.4. Estimação dos parâmetros da estrutura selecionada	136
4.2.5. Validação do modelo.....	137
4.3. MODELOS NÃO LINEARES PARA IDENTIFICAÇÃO DE PROCESOS	137
4.3.5. Modelos empíricos com técnicas de inteligência artificial.....	138
4.3.5.1. <i>Modelos empíricos com redes neurais artificiais</i>	138
4.3.5.1.1. Definição do problema	140
4.3.5.1.2. Coleta dos dados de treinamento e de teste	141
4.3.5.1.3. Pré e Pós-processamento de dados	142
4.3.5.1.4. Definição da Topologia da rede.....	143
4.3.5.1.5. Treinamento da rede neural	144
4.3.5.1.5.1. <i>Métodos para melhorar a velocidade de aprendizagem do algoritmo de retropropagação</i>	145
4.3.5.1.5.1.1. <i>O algoritmo Levenberg- Marquardt para o treinamento da rede</i>	146
4.3.5.1.5.2. <i>A regularização bayesiana</i>	148
4.3.5.1.6. Validação e teste do modelo baseado em redes neurais	152
4.3.5.1.6.1. <i>A validação cruzada (cross validation)</i>	152
4.3.5.1.6.2. <i>Variantes da validação cruzada</i>	153
4.3.5.1.6.3. <i>A validação cruzada para seleção de modelos de redes neurais</i>	156
4.3.5.1.6.4. <i>A validação cruzada e parada antecipada (Early stopping)</i>	157
4.4. CONCLUSÕES.....	159
4.5. REFERÊNCIAS BIBLIOGRÁFICAS	160

CAPÍTULO 5. CONTROLE DE PROCESSOS BASEADO EM MODELO 172

5.1. INTRODUÇÃO.....	173

5.2. CONTROLE PREDITIVO COM MODELO (MPC)	173
5.2.1. Características dos controladores preditivos	175
5.2.1.1. <i>Modelos de predição</i>	176
5.2.1.2. <i>A trajetória de referência para a saída do processo</i>	178
5.2.1.3. <i>Funções objetivo</i>	178
5.2.1.4. <i>Procedimentos de minimização</i>	180
5.3. CONTROLE DE MATRIZ DINÂMICA (DMC)	182
5.3.1. Modelo de convolução SISO	182
5.3.2. Algoritmo de controle para sistemas monovariáveis	185
5.3.3. Resposta ao degrau- MIMO	191
5.3.4. Algoritmo de controle para sistemas multivariáveis	192
5.4. CONTROLADOR DE MATRIZ DINÂMICA QUADRÁTICO (QDMC)	197
5.5. CONCLUSÃO	200
5.6. REFERÊNCIAS BIBLIOGRÁFICAS	201

CAPÍTULO 6. ENZYME LOADING OPTIMIZATION IN THE SUGARCANE BAGASSE HYDROLYSIS FOR SECOND GENERATION ETHANOL PRODUCTION USING ARTIFICIAL INTELLIGENCE TECHNIQUES	204
6.1. INTRODUCTION	206
6.2. MATERIAL AND METHODS	209
6.2.1. Substrate	209
6.2.2. Lime pretreatment	210
6.2.3. Enzymatic hydrolysis	210
6.2.4. Dynamic modeling of enzymatic hydrolysis	212
6.2.4.1. <i>ANN configurations</i>	212
6.2.4.2. <i>The data set for dynamic modeling based on ANN</i>	214
6.2.4.3. <i>The experimental data set a</i>	214
6.2.4.4. <i>The interpolated data set b</i>	216
6.2.4.5. <i>The interpolated data set c</i>	216
6.2.4.6. <i>Artificial neural network based interpolation</i>	216
6.2.5. Selection of the best dynamic model based on ann for enzymatic hydrolysis	217
6.2.6. Optimization using the dynamic model based on neural networks	218
6.3. RESULTS AND DISCUSSION	220

6.3.1. Selection of the best neural network for dynamic modeling.....	220
6.3.2. Estimated responses with the dynamic model based on ANN.....	222
6.3.3. <i>Identification of the optimum region using the model based on ANN</i>	223
6.3.4. Optimization by integrating AI Techniques	225
6.4. CONCLUSIONS.....	227
6.5. REFERENCES	228

CAPÍTULO 7. MODELAGEM EMPÍRICA DE PROCESSOS: UM PROCESSO DE FERMENTAÇÃO CONTINUA COM EVAPORADOR FLASH COMO CASO DE ESTUDO

.....	231
7.1. INTRODUÇÃO.....	232
7.2. CASO DE ESTUDO	233
7. 3. IDENTIFICAÇÃO DE PROCESSOS	237
7.3.1. Reconciliação de dados e <i>denoising</i>	237
7.3.2. Caracterização básica dos dados	238
7.3.3. CAracterização dinâmica dos dados	238
7.3.4. Modelos empíricos baseados em redes neurais artificiais.....	239
7.4. RESULTADOS E DISCUSSÃO	243
7.4.1. Pré-tratamento dos dados.....	243
7.4.2. Caracterização básica dos dados	244
7.4.3. Caracterização dinâmica dos dados	246
7.4.4. Comentários finais do comportamento dinâmico das variáveis do processo.	264
7.4.5. Modelagem empírica baseada em redes neurais para monitoramento da concentração de etanol.....	265
7.4.5.1. <i>MODELO 1: modelo empírico com variáveis cuja determinação é off-line</i>	265
7.4.5.2. <i>MODELO 2: modelo empírico com variáveis que não requerem determinação off- line</i>	267
7.4.5.3. <i>Gráficos de paridade com dados experimentais de teste e estimados pelos modelos empíricos</i>	268
7.4.5.4. <i>Análise exploratória de modelos empíricos com diferentes entradas que não requerem determinação off-line</i>	270
7.5. CONCLUSÕES.....	271
7.6. REFERÊNCIAS BIBLIOGRÁFICAS	273

CAPÍTULO 8. ANÁLISE DAS RELAÇÕES CAUSA- EFEITO EM MODELOS EMPÍRICOS DE REDES NEURAIS: INTERPRETAÇÃO DA INFORMAÇÃO CONTIDA NA CAIXA

PRETA	276
Prefácio	278
8.1. PART 1: A METHOD OF TRANSFER FUNCTIONS AND BLOCK DIAGRAMS TO STUDY THE CONTRIBUTION OF VARIABLES IN ARTIFICIAL NEURAL NETWORK PROCESS MODELS	280
Abstract	280
8.1.1. INTRODUCTION	280
8.1.2. THE TRANSFER FUNCTION METHOD	281
8.1.2.1. Basic concepts	281
8.1.2.1.1. Transfer functions	281
8.1.2.1.2. Block diagrams	281
8.1.2.1.3. Basic rules of block diagram algebra	282
8.1.2.2. Generalized equations for the effect of inputs on the outputs (e_{oi})	283
8.1.3. NUMERICAL APPLICATION	286
8.1.3.1. PARTICULAR CASE: Relative contributions expressed in terms of absolute values of the weights as transfer function	286
8.1.3.1.1. Example 1: Neural network with 5: 2: 1 topology, with five inputs, two hidden neurons and one output neuron	288
8.1.3.1.2. Example 2: Neural network with 11:5:5:1 topology, to model a batch fermentation process for bioethanol production	293
8.1.3.1.2.1. Introduction	293
8.1.3.1.2.2. ANN model topology	295
8.1.3.1.2.3. Comparison of the results for the importance of the effects of input variables on the outputs	295
8.1.4. CONCLUSION	297
8.1.5. REFERENCES	297
8.2. PART 2: A NEW FRAMEWORK TO STUDY CAUSE AND EFFECT RELATIONSHIPS WITH NEURAL NETWORKS AND ITS APPLICATION TO BIOPROCESS MODELS ...	300
Abstract	300
8.2.1. INTRODUCTION	300

8.2.1.1. <i>Transfer functions</i>	302
8.2.1.2. <i>Block diagrams</i>	303
8.2.2. THE METHOD OF TRANSFER FUNCTIONS TO STUDY THE EFFECTS OF INPUTS ON THE OUTPUTS IN ARTIFICIAL NEURAL NETWORKS: SHORTCUT PROCEDURE.....	305
8.2.3. NUMERICAL APLICATIONS	312
8.2.3.1. PARTICULAR CASE: Relative contributions expressed in terms of absolute values of the weights and biases as transfer functions.....	312
8.2.3.1.1. Case 1: Neural network with 5: 2: 1 topology, with five inputs, two hidden neurons and one output neuron	313
8.2.3.1.2. Case 2: Neural network with 13:7:1 topology, to model a continuous extractive fermentation process for bioethanol production from sugarcane molasses	316
8.2.3.1.2.1. <i>Analyzing the results using Principal Component Analysis</i>	322
8.2.4. CONCLUSIONS.....	324
8.2.5. LITERATURE CITED	326
8.3. PART 3: USE OF THE TRANSFER FUNCTION METHOD TO REDUCE THE NUMBER OF INPUTS IN ANN PLANT MODELS FOR EXTRACTIVE FERMENTATION WITH FLASH UNIT	329
8.3.1. INTRODUCTION	329
8.3.2. Alternatives for plant modeling using different number of variables	330
8.3.3. Use of variable contributions to the ann output to identify the importance of variables	331
8.3.4. Use of principal component analysis to identify the importance of variables and to reduce the number of inputs	334
8.3.5. Description of the ann plant models	338
8.3.5.1. <i>Alternative I</i>	338
8.3.5.2. <i>Alternative II</i>	339
8.3.6. Selection of manipulated variables from the input data set.....	343
8.3.7. CONCLUSIONS.....	345
8.3.8. REFERENCES	345

CAPÍTULO 9. DYNAMIC MATRIX CONTROL WITH INTERNAL MODEL BASED ON ANN OF A CONTINUOUS EXTRACTIVE PROCESS FOR BIOETHANOL PRODUCTION

.....	347
Abstract	349
9.1. INTRODUCTION	351
9.2. ARTIFICIAL NEURAL NETWORK MODELING	353
9.3. EXTRACTIVE FERMENTATION PROCESS FOR BIOETHANOL PRODUCTION	355
9.4. DYNAMIC MATRIX CONTROL WITH INTERNAL MODEL BASED ON ANN	356
9.4.1. Dnamic matrix control (DMC)	356
9.4.2. Internal model based on ANN	358
9.5. RESULTS AND DISCUSSION	361
9.5.1. ALTERNATIVE A.	361
9.5.1.1. ANN internal plant model of the controller	361
9.5.1.2. Non linear predictive control	362
9.5.2. ALTERNATIVE B	367
9.5.2.1. ANN internal plant model of the controller	368
9.5.2.2. Controller performance	370
9.5.2.2.1. Regulator performance	370
9.5.2.2.2. Servo performance	375
9.5.3. SUBTITUTION OF THE ANN INTERNAL MODEL OF THE CONTROLLER FROM ALTERNATIVE B IN THE CONTROLLER OF ALTERNATIVE A	376
9.6. CONCLUSION	379
9.9. REFERENCES	379

CAPÍTULO 10. NON- LINEAR PREDICTIVE CONTROL OF A FERMENTOR IN A CONTINUOUS REACTION-SEPARATION PROCESS.....382

Abstract	383
10.1. INTRODUCTION	383
10.2. PROCESS DESCRIPCION	384
10.2.1. Experimental stage	384
10.3. NEURAL NETWORK MODEL PREDICTIVE CONTROL (NNMPC) ALGORITHM	386
10.4. RESULTS AND DISCUSSION	388
10.4.1. ALTERNATIVE I	388

10.4.1.1. Process modeling.....	388
10.4.1.2. Dynamic behavior study.....	389
10.4.1.3. Internal model training.....	390
10.4.1.4. Controller performance	392
10.4.1.4.1. Regulator performance.....	392
10.4.1.4.2. Servo performance.....	394
10.4.2. ALTERNATIVE II	396
10.5.2.1 Regulator performance	398
10.5.2.2. Servo performance.....	400
10.5.3. SUBSTITUTION OF THE ANN INTERNAL MODEL OF THE CONTROLLER FROM ALTERNATIVE II IN THE CONTROLLER OF ALTERNATIVE I	403
10.6. CONCLUSION	406
10.7. REFERENCES	406

CAPÍTULO 11. DECENTRALIZED NON-LINEAR PREDICTIVE CONTROL FOR MIMO REACTION-SEPARATION PROCESSES: A CONTINUOUS EXTRACTIVE FERMENTATION AS STUDY CASE.....	409
Abstract	410
11.1. INTRODUCTION	410
11.2. PROCESS DESCRIPCION.....	412
11.2.1. EXPERIMENTAL STAGE	412
11.2.2. VARIABLE SELECTION	415
11.3. MODELING AND CONTROL.....	418
11.3.1. ARTIFICIAL NEURAL NETWORK PROCESS MODELING	418
11.3.2. NON-LINEAR MODEL PREDICTIVE CONTROL ALGORITHM.....	419
11.4. RESULTS AND DISCUSSION.....	421
11.4.1. PROCESS MODELING	421
11.4.2. DYNAMIC BEHAVIOR STUDY	423
11.4.3. PAIRING OF VARIABLES AND RESILIENCY	425
11.4.4. INTERNAL MODEL TRAINING.....	427
11.4.5. DESCENTRALIZED CONTROL SYSTEM PERFORMANCE.....	428
11.4.5.1. Regulator performance	428
11.4.5.2. Combined servo and regulator performances. Case A.	429

11.4.4.3. Combined servo and regulator performances. Case B.	431
11.5. CONCLUSIONS.....	432
11.6. REFERENCES	433
 CAPÍTULO 12. CONCLUSÕES GERAIS	435
SUGESTÕES PARA TRABALHOS FUTUROS	438
 APÊNDICES.....	439
APÊNDICE 1A.....	439
DESCRIÇÃO COMPLEMENTAR DE PROCEDIMENTOS PARA REMOÇÃO DO ETANOL DESDE O CALDO DE FERMENTAÇÃO.	440
APÊNDICE 1.B.	446
Balancos de massa na fermentação contínua com evaporador flash.....	446
separação do produto por Flash	447
APÊNDICE 1.C.....	450
Modelos para o equilíbrio líquido-vapor: NRTL e UNIFAC	450
 APÊNDICE 2.A.	454
Processos de aprendizagem das redes neurais.....	454
 APÊNDICE 3.A.	461
CONSIDERAÇÕES MAIS AMPLAS PARA SELEÇÃO DA ESTRUTURA DO MODELO E ESTIMAÇÃO DOS PARÂMETROS DA ESTRUTURA SELECIONADA.....	461
APÊNDICE 3.B.	464
MODELOS NÃO LINEARES PARA IDENTIFICAÇÃO DE PROCESSOS.....	464
APÊNDICE 3.C.....	477
Outros métodos que podem ser aplicados para obter a topologia ótima nos modelos de redes neurais.....	477
APÊNDICE 3.D.....	486
Outros métodos para melhorar a velocidade de aprendizagem do algoritmo de <i>backpropagation</i>	486

ÍNDICE DE TABELAS

CAPÍTULO 2

Tabela 2.1. Inibição por subprodutos na fermentação com <i>Saccharomyces cerevisiae</i> para a produção de etanol.....	48
Tabela 2.2. Resultados médios da análise composicional do bagaço.	57
Tabela 2.3. Termos de inibição pelo produto, substrato e biomassa nos modelos que expressam a velocidade específica de crescimento da levedura.....	69
Tabela 2.4. Modelos utilizados para fermentação contínua e com separação do produto por evaporação flash.	74
Tabela 2.5. Parâmetros cinéticos e sua dependência com a temperatura.....	78
Tabela 2.6. Condições de operação do tanque flash.....	80

CAPÍTULO 6

Table 6.1. Composition of raw material and bagasse pretreated with lime.....	210
Table 6.2. Runs of the experimental design central composite 2^2 plus star (coded levels in parenthesis) and glucose (g L^{-1}) at three sampling times during the enzymatic hydrolysis.....	211
Table 6.3. Values of the criteria parameters to select the best neural network model using the validation data set 1.	220
Table 6.4. Weight and biases in the optimal ANN structure	221
Table 6.5. Estimated global optimum using different methods.	226

CAPÍTULO 7

Tabela 7.1. Matriz de correlações para as variáveis do processo de produção de etanol por fermentação contínua extrativa com evaporador flash.	245
Tabela 7.2. MODELO 1: Número de neurônios na camada oculta, erro quadrático	266
Tabela 7.3. MODELO 1: Número de neurônios na camada oculta, erro quadrático médio (mse), R^2 e CORR no grupo de teste com 300 épocas de treinamento.	266
Tabela 7.4. MODELO 2: Número de neurônios na camada oculta, de erro quadrático médio (mse), R^2 e CORR no grupo de teste com 300 épocas de treinamento.	267
Tabela 7.5. MODELO 2: Número de neurônios na camada oculta, erro quadrático médio (mse), R^2 e CORR no grupo de teste com 500 épocas de treinamento.	267

Tabela 7.6. Modelos de redes neurais artificiais (ANN) com diferentes entradas determináveis <i>on-line</i> e saídas etanol no fermentador Ef e etanol condensado Ec.....	271

CAPÍTULO 8

Table 8.1.1. Example 1: ANN weights associated to each	291
Table 8.2.1: General matrix equations for ANN models	311
Table 8.2.2. Case 1: Weights associated to each connection between neurons	313
Table 8.2.3. Case 1: Comparison of the importance of the inputs' effects using different methods.	315
Table 8.2.4. Case 2: Weights associated to each connection between neurons	319
Table 8.2.5. Principal Component contributions to the total variation in the input data set.....	322
Table 8.3.1. Input variable contribution to the output variable.....	332
Table 8.3.2. Principal Component contributions to the total variation in the input data set: Operation time 40 - 400 h.....	335
Table 8.3.3. Principal Component contributions to the total variation in the input data set : Operation time 150 - 400h.....	335
Table 8.3.4. Alternative I: ANN model selection, mean squared error (mse) and R^2 values in the validation test.....	339
Table 8.3.5. Alternative 2: ANN selection, mean squared error (mse) and R^2 values in the validation test.....	341

CAPÍTULO 9

Table 9.1. ANN internal model of the controller: values of the criteria used for selecting the model when is used the purge flow rate (F_{pu}) as manipulated variable.....	339

CAPÍTULO 10

Table 10.1. ANN internal model of the controller: values of the criteria used for selecting the model when is used the feed flow rate to the fermentor (F_o) as manipulated variable.....	397

CAPÍTULO 11

Table 11.1. Indicators of effective pairing of variables and resiliency.....	425
Table 11.2. Control structure tested to implement the Decentralized Control System	428

ÍNDICE DE FIGURAS

CAPÍTULO 2

Figura 2.1. Diferentes materias primas e processos para produção de etanol	40
Figura 2.2. Produtividade de Etanol (Litros por Hectare).....	41
Figura 2.3. Balanço energético: Energia contida no combustível/Energia Fóssil	42
Figura 2.4. Diagrama de blocos da produção de bioetanol de cana-de-açúcar	43
Figura 2.5. Tipos de modelos cinéticos usados em fermentações	65

CAPÍTULO 3

Figura 3.1. Modelo de um neurônio	103
Figure 3.2. a) underfitting, b) fitting , c) overfitting	115

CAPÍTULO 4

Figura 4.1. Identificação de sistemas utilizando redes neurais.....	140
Figura 4.2. Método V-fold para validação cruzada, com V=3.....	155

CAPÍTULO 5

Figura 5.1. O conceito de controle preditivo (Soertemboek, 1990).....	175
Figura 5.2. Resposta de um sistema em malha aberta frente a uma perturbação degrau	183
Figura 5.3. Resposta dinâmica do sistema quando são aplicadas três	184

CAPÍTULO 6

Figure 6.1. Experimental behaviour of glucose (g L^{-1}) for runs in the design matrix on different times	215
Figure 6.2. Structure of the basic genetic algorithm	218
Figure 6.3. Modeling and optimization based on AI.....	219
Figure 6.4. Optimal ANN structure	221
Figure 6.5. ANN output and target output after training and validation. (a) , (b) training,223	

Figure 6.6. Optimal region for the maximum glucose production in the enzymatic hydrolysis of lime pretreated sugarcane bagasse: (a), (c) and (e) response surfaces and (b), (d) and (f) contour plots of the ANN based dynamic model. Constant factors: (a) and (b) time = 60 h (3600 min), (c) and (d) β -glucosidase = 750 CBU/L, (e) and (f) cellulase = 475 FPU/L; pH = 4.8 and T = 50 °C for all the experiments.....224

CAPÍTULO 7

Figura 7.1. Diagrama esquemático de processo de fermentação com unidade de separação flash.	235
Figura 7.2. Esquema do modelo empírico do processo com variáveis de	239
Figura 7. 3. Esquema do modelo empírico do processo com variáveis de	240
Figura 7.4. Eliminação do ruído do sinal para a pressão de operação do tanque flash, ...	243
Figura 7.5. Eliminação do ruído do sinal para a vazão de retorno do filtro, ...	244
Figura 7.6. Media móvel da variável concentração de etanol condensado, Ec.....	247
Figura 7.7. Variância móvel da variável concentração de etanol condensado, Ec.....	247
Figura 7.8. Media móvel da variável concentração de etanol no fermentador, Ef.....	248
Figura 7.9. Variância móvel da variável concentração de etanol no fermentador, Ef.....	248
Figura 7.10. Media móvel da variável concentração de biomassa no fermentador, X....	249
Figura 7.11. Variância móvel da variável concentração de biomassa no fermentador, X.....	249
Figura 7.12. Media móvel da variável concentração de substrato no fermentador, S....	250
Figura 7.13. Variância móvel da variável concentração de substrato no fermentador, S.....	250
Figura 7.14. Media móvel da variável vazão de purga, Fpu.....	251
Figura 7.15. Variância móvel da variável vazão de purga, Fpu.....	251
Figura 7.16. Media móvel da variável vazão de permeado, Fp.....	252
Figura 7.17. Variância móvel da variável vazão de permeado, Fp.....	252
Figura 7.18. Media móvel da variável vazão de alimentação, Fo.....	253
Figura 7.19. Variância móvel da variável vazão de alimentação, Fo.....	253
Figura 7.20. Media móvel da variável temperatura no tanque flash, Tflash.....	254
Figura 7.21. Variância móvel da variável temperatura no tanque flash, Tflash.....	254
Figura 7.22. Media móvel da variável pressão de operação no tanque flash, Pflash.....	255

Figura 7.23. Variância móvel da variável pressão de operação no tanque flash, Pflash.	255
Figura 7.24. Media móvel da variável temperatura no fermentador, Tferm.	256
Figura 7.25. Variância móvel da variável temperatura no fermentador, Tferm.	257
Figura 7.26. Media móvel da variável vazão de entrada no tanque flash, Fin.	257
Figura 7.27. Variância móvel da variável vazão de entrada no tanque flash, Fin.	258
Figura 7.28. Media móvel da variável taxa de diluição no fermentador, D.	258
Figura 7.29. Variância móvel da variável taxa de diluição no fermentador, D.	259
Figura 7.30. Media móvel da variável concentração de etanol condensado, Ec.	259
Figura 7.31. Variância móvel da variável turbidez no fermentador, Tb.	260
Figura 7.32. Media móvel da variável vazão de retorno ao fermentador, Fr.	261
Figura 7.33. Variância móvel da variável vazão de retorno ao fermentador, Fr.	261
Figura 7.34. Media móvel da variável viabilidade celular, Cv.	262
Figura 7.35. Variância móvel da variável viabilidade celular, Cv.	262
Figura 7.36. Media móvel da variável concentração de glicerol no fermentador, G.	263
Figura 7.37. Variância móvel da variável concentração de glicerol no fermentador, G.	263
Figura 7.38. Media móvel da variável pH no fermentador.	264
Figura 7.39. Variância móvel da variável pH no fermentador.	264
Figura 7.40. Representação gráfica das estruturas de rede neural:	
a) Modelo 1 com variáveis que precisam de determinação <i>off-line</i> .	
b) Modelo 2 com variáveis que não precisam de determinação <i>off-line</i> .	268
Figura 7.41. MODELO 1 : Valores da concentração de etanol no fermentador experimentais e estimados pela rede neural.	269
Figura 7.42. MODELO 1: Valores da concentração de etanol no fermentador experimentais e estimados pela rede neural.	269

CAPÍTULO 8

Figure 8.1.1. Transfer function	281
Figure 8.1.2. Blok diagram for transfer functions between input x_1 and output y .	282
Figure 8.1.3. Rule 1: blocks in series	282
Figure 8.1.4. Rule 2: blocks in parallel	282
Figure 8.1.5. Reduced block diagram	283

Figure 8.1.6. Notation used for transfer function the effect of the input on the output (E_{oi})	283
Figure 8.1.7. Block diagram to show transfer functions in the neural network structure.	284
Figure 8.1.8. Transfer functions in neural network with two inputs, three neurons.	284
Figure 8.1.9. Algorithm to apply the method of transfer functions.	286
Figure 8.1.10. Notation used for weights in neural networks	286
Figure 8.1.11. Comparison of the importance of the inputs' effects using different methods.	292
Figure 8.1.12. Profiles of the variation of the output variable.	294
Figure 8.1.13. Comparison of the importance of the input's effect.	296
Figure 8.2.1. Transfer function.	303
Figure 8.2.2. Rule 1. Bloks in series.	303
Figure 8.2.3. Rule 2 . Bloks in parallel.	303
Figure 8.2.4. Blok diaggram for transfer functions between input x and output y	304
Figure 8.2.5. Blok diagram reduction for Figure 8.2.4.	304
Figure 8.2.6. Notation used for transfer functions.	305
Figure 8.2.7. Block diagram reduction in the SISO subsystem between the neurons a and b .	306
Figure 8.2.8. Blok diagram to show transfer functions in the neural network structure.	307
Figure 8.2.9. Transfer functions in neural network A.	308
Figure 8.2.10. Transfer functions in neural network B.	310
Figure 8.2.11. Algorithm to apply the shorcut procedure for the method of transfer functions.	311
Figure 8.2.12. Notation used for ANN weights.	312
Figure 8.2.13. Influence of the bias value on the estimation of the importance of the input effects on the output.	316
Figure 8.2.14. Schematic diagram of the fermentation pilot plant.	317
Figure 8.2.15. Experimental target values of the validation data set and network outputs.	320
Figure 8.2.16. Input effects on the output (E_f) with the method of transfer functions.	321
Figure 8.2.17. Correlation coefficients r between each input and the principal components $Cp1$ and $Cp2$.	323

Figure 8.3.1. Correlation coefficients r between each input factor and the principal components Cp1 and Cp2. Operation time: 50 – 400 h.....	336
Figure 8.3.2. Correlation coefficients r between each input factor and the principal components Cp1 and Cp2. Operation time: 150 – 400 h.....	336
Figure 8.3.3. Alternative I : ANN selected model topology.....	338
Figure 8.3.4. Experimental and ANN predicted values of the fermentor ethanol concentration in the pilot plant with extractive fermentation process.....	339
Figure 8.3.5. Topology of the ANN plant model to be controlled. Alternative II, structure A.....	342

CAPÍTULO 9

Figure 9.1. Flow diagram of the ethanol production pilot plant with extractive fermentation process.....	355
Figure 9.2. Model predictive control process.....	359
Figure 9.3. Inputs and output of the internal model of the DMC controller.....	360
Figure 9.4 Alternative A: Topology of the internal ANN plant model of the controller.....	362
Figure 9.5. Controlled variable responses against time for step disturbance.....	364
Figure 9.6. DMC controller actions to reach the steady state for an initial step.....	364
Figure 9.7. Alternative A, controller behavior for <i>set-point</i> changes:.....	366
Figure 9.8. Damping effect for different values of f the weighting factor for Δu 's in the controller objective function.....	367
Figure 9.9. Changes in the manipulated variable for different values of f the weighting factor	367
Figure 9.10. General topology of the ANN plant model of the controller.	368
Figure 9.11 Alternative B: Topology of the internal ANN plant model of the controller....	369
Figure 9.12. Case 1: Alternative B, a) response of the controlled variable, E_f , with time to perturbations of $\pm 10\%$ in the permeate flow rate (F_p). b) manipulated variable: purge flow rate (F_{pu}).....	371
Figure 9.13. Case 2: Alternative B, a) response of the controlled variable, E_f , with time to perturbations of $\pm 10\%$ in the permeate flow rate (F_p), b) manipulated variable: purge flow rate (F_{pu}).....	372
Figure 9.14. Case 3. Alternative B, a) response of the controlled variable, E_f , with time to perturbations of $\pm 5\%$ in the feed flow rate to the fermentor (F_o), b) manipulated variable: purge flow rate (F_{pu}).....	372

Figure 9.15 Case 4. Alternative B, a) response of the controlled variable, E_f , with time to perturbations of $\pm 5\%$ in the feed flow rate to the fermentor (F_o), b) manipulated variable: purge flow rate (F_{pu}).....	372
Figure 9.16 Alternative B, servo performance: a) controlled variable E_f and setpoint changes, b) manipulated variable purge flow rate (F_{pu}).....	375
Figure 9.17 Use of the internal model from alternative B in alternative A: controller regulator performance, response of the controlled variable, E_f , with time to perturbations of $\pm 5\%$ in the feed flow rate to the fermentor (F_o). Manipulated variable: purge flow rate (F_{pu}).....	377
Figure 9.18 Use of the internal model from alternative B in alternative A: controller servo performance: a) Controlled variable E_f and setpoint changes. b) Manipulated variable purge flow rate (F_{pu}).....	378

CAPÍTULO 10

Figure10.1. Diagram of the extractive fermentation process with vacuum flashing.....	385
Figure 10.2. Normalized values of inputs and outputs:.....	390
Figure10.3 Controller ANN plant model topology	391
Figure 10.4. Control structure 1A: Response of the controlled variable, E_f , with time to perturbations of $\pm 5\%$ in the feed flow rate to the fermentor, F_o . Manipulated variable: purge flow rate, F_{pu}	393
Figure10.5. Control structure 2A: Response of the controlled variable, E_f , with time to perturbations of $\pm 20\%$ in the purge flow rate, F_{pu} . Manipulated variable: feed flow rate to the fermentor, F_o	393
Figure 10.6. Servo performance of the control structure 1A:	395
Figure 10.7. Servo behavior of control structure 2A:	395
Figure10.8. Control structure 1B: Response of the controlled variable, E_f , with time to perturbations of $\pm 5\%$ in the feed flow rate to the fermentor, F_o . Manipulated variable: purge flow rate, F_{pu}	399
Figure10.9. Control structure 2B: Response of the controlled variable, E_f , with time to perturbations of $\pm 20\%$ in the purge flow rate, F_{pu} . Manipulated variable: feed flow rate to the fermentor, F_o	400
Figure 10.10. Servo performance of the control structure 1B:	401
Figure 10.11. Servo performance of the control structure 2B:	402

Figure 10.12. Regulator performance of the control structure 2A from alternative I with controller internal model IM1-II: a) Response of the controlled variable, E_f , with time to perturbations of $\pm 20\%$ in the purge flow rate, F_{pu} . b) Manipulated variable: feed flow rate to the fermentor, F_o	404
Figure 10.13. Servo performance of the control structure from	405

CAPÍTULO 11

Figure 11.1. Diagram of the extractive fermentation process with vacuum flashing.....	413
Figure 11.2. Experimental behavior of the selected variables in the.....	417
Figure 11.3 General scheme of the ANN-MPC predictive controller.....	419
Figure 11.4. General scheme of decentralized control system.	421
Figure 11.5. ANN model topology	422
Figure 11. 6. Output of ANN process model and experimental values of the validation data set.....	423
Figure 11.7. Effects of the input random step disturbances. a) Input: permeate flow rate, F_p . Output: fermentor ethanol concentration, E_f . b) Input: purge flow rate, F_{pu} . Output: condenser ethanol concentration, E_c	424
Figure 11.8. Controlled variables in the regulator performance to disturbances of $\pm 10\%$ in the feed flow rate to the fermentor, F_o , and of $\pm 10\%$ in the dilution rate, D . a) Control loop 1, controlled variable: Fermentor ethanol concentration, E_f . b) Control loop 2, controlled variable: Ethanol concentration in the condensed stream, E_c	429
Figure 11.9. Manipulated variables in the regulator performance to disturbances of $\pm 10\%$ in the feed flow rate to the fermentor, F_o , and of $\pm 10\%$ in the dilution rate, D . a) Control loop 1, manipulated variable: Permeate flow rate, F_p . b) Control loop 2, manipulated variable: Fermentor purge flow rate, F_{pu}	429
Figure 11.10. Controlled variables in the servo and regulator performance, case A: a) Servo performance: Control loop1, controlled variable: Fermentor ethanol concentration, E_f . b) Regulator performance: Control Loop 2, controlled variable: Ethanol concentration in the condensed stream, E_c	430
Figure 11.11. Manipulated variables in the servo and regulator performance, case A: a) Control loop 1, manipulated variable: Permeate flow rate, F_p . b) Control Loop 2, manipulated variable: Fermentor purge flow rate, F_{pu}	430

Figure 11.12. Controlled variables in the servo and regulator performance, case B: a) Servo performance: Control loop 2, controlled variable: Ethanol concentration in the condensed stream, E_c . b) Regulator performance: Control Loop 1, controlled variable: Fermentor ethanol concentration, E_f 431

Figure 11.13. Manipulated variables in the servo and regulator performance, case B: a) Control loop 1, manipulated variable: Permeate flow rate, F_p . b) Control Loop 2, manipulated variable: Fermentor purge flow rate, F_{pu} 431

RESUMO

Na presente tese foram abordados problemas no contexto da produção de etanol de primeira e segunda geração. No caso de etanol de segunda geração foi abordado o problema de melhorar a eficiência da hidrólise enzimática, com as seguintes contribuições: (1) otimização da carga enzimática utilizando técnicas de modelagem por redes neurais. (2) substituição de um modelo com limitações na representatividade dos pontos experimentais obtido por desenho experimental pela superfície de resposta de um modelo de redes neurais que permite a exploração da região do ótimo no espaço de fatores e respostas. Os resultados obtidos mostram (1) boa precisão para a localização das coordenadas da região ótima de trabalho. (2) um mapeamento da evolução do processo com localização do ótimo global utilizando algoritmos genéticos e técnicas de computação evolutiva. No caso dos processos de produção de etanol de primeira geração foi estudado o problema do controle com modelo de um processo de fermentação contínua com extração de etanol utilizando vácuo quando não se tem modelo fenomenológico disponível com as seguintes contribuições e resultados: (1) foi desenvolvida uma abordagem que integrou redes neurais artificiais, com o controle preditivo com modelo (MPC). (2) foram desenvolvidos modelos empíricos do processo com redes neurais artificiais (3) foi desenvolvida uma abordagem que utiliza o “conhecimento aprendido” pelas redes neurais o qual é armazenado em pesos sinápticos e bias. (4) foram desenvolvidos vários modelos empíricos de redes neurais para o monitoramento das concentrações de etanol que podem ser utilizados para desenvolver software sensors. (5) foram implementadas diferentes estruturas de controle preditivo com diferentes modelos internos de redes neurais para os controladores, com otimizador linear e não linear para o caso de estudo. Nos diferentes capítulos em que foram implementadas estruturas de controle integrando as redes neurais com a tecnologia MPC, se mostrou que, a abordagem desenvolvida é eficiente para os projetos dos sistemas de controle com modelo empírico.

Palavras-Chave: Produção de bioetanol, Inteligência artificial, Redes neurais, Identificação de processos, Controle preditivo com modelo.

ABSTRACT

In this thesis were studied problems in the context of ethanol production of first and second generation. In the case of second generation ethanol was studied the problem of improving the enzymatic hydrolysis efficiency, with the following contributions: 1) optimization of enzyme loading by using modeling techniques based on neural networks, 2) substitution of a model with limited representation capacity of the experimental points which was obtained by using experimental design by the response surface model of a neural network that allows the exploration of the space of factors and responses. The results show: 1) good accuracy in locating the coordinates of the optimum working region, 2) the mapping of the evolution of the process with the location of the global optimum by using genetic algorithms and evolutionary computing techniques. In case of the production of first generation ethanol was studied the problem of the control with process model of a fermentation process with continuous extraction of ethanol by using a vacuum system when there is no a phenomenological model available with the following contributions and results: 1) It was developed an approach in which were integrated artificial neural networks with model predictive control (MPC). 2) It were developed empirical process models by using artificial neural networks 3) It was developed an approach that uses the “learned knowledge” by neural networks which is stored in synaptic weights and bias. 4) It were developed several empirical models of neural networks for monitoring concentrations of ethanol that can be used to develop software sensors. 5) It were implemented different predictive control structures with different internal models based on neural networks for controllers with linear and non-linear optimizer to be applied on the case of study. In the different chapters in which were implemented control structures by integrating neural networks with MPC technology, it was showed that the developed approach is efficient to be applied in the designs of control systems with empirical model.

Keywords: Bioethanol production, artificial intelligence, neural networks, process identification, model predictive control.

CAPÍTULO 1. INTRODUÇÃO

A utilização de combustíveis fósseis afeta negativamente a qualidade e o equilíbrio do meio-ambiente, pela elevada taxa de emissão de CO₂, que está relacionada diretamente com o problema do efeito estufa. Além disso, existem expectativas para esgotamento dos combustíveis fósseis nos próximos 100 anos (SANTOS e MOTHE, 2007), e de esgotamento de petróleo nos próximos 41 anos. Nota-se então, a necessidade de mudar a matriz energética baseada quase que exclusivamente em combustíveis fósseis. Neste contexto, a pesquisa e desenvolvimento foram intensificados em uma forma mais diversificada para a utilização de fontes renováveis de energia em substituição às fontes fósseis (PEREIRA Jr. *et al.*, 2008).

O bioetanol tem sido considerado uma alternativa para diminuir problemas ambientais e energéticos no mundo. Comparado aos combustíveis fósseis, o bioetanol apresenta as vantagens de ser uma fonte renovável de energia, e de contribuir para a redução das emissões de dióxido de carbono (EMBRAPA, 2011) e por tanto na diminuição do impacto ao meio ambiente. Essas características têm contribuído ao aumento da demanda internacional pelo bioetanol.

No Brasil, devido à importância do bioetanol para a sua utilização como combustível, as instituições de pesquisa, que já tradicionalmente dedicavam suas investigações ao setor sucroalcooleiro, reforçaram suas atividades no sentido de uma maior compreensão da fermentação alcoólica.

O etanol que se acumula na fermentação acima de uma determinada concentração inibe a atividade metabólica dos microrganismos (ATALA, 2006), de forma que a remoção parcial do etanol produzido até níveis em torno de 40 g/L na fermentação reduz significativamente o efeito inibitório, e isto pode ser alcançado anexando um evaporador flash ao fermentador. As vantagens do processo contínuo de retirada do produto da fermentação utilizando vácuo são mencionados

por Costa, et al. (2002), Silva (1998) e Schimidell e Facciotti (2001), que sugerem que *obtem-se* um melhor desempenho do processo, se reduzem tempos mortos, aumenta-se a produtividade e com o uso de estratégias eficazes de controle se podem reduzir os custos de produção e impactos ambientais, operando sob condições ótimas.

No entanto, os processos contínuos de fermentação são processos cuja complexidade é reconhecida na literatura (RIVERA et al., 2010). Consequentemente, o controle do fermentador no ponto ótimo de operação é tarefa desafiadora (HENSON e SEBORG, 1992; BIEGLER e RAWLINGS, 1991).

A utilização de controladores clássicos PID apresenta desempenho ineficiente quando o processo tem comportamento complexo, apresenta não linearidades, é multivariável e variante no tempo (DECHECHI, 1998; LEITE, 2011). Por estas razões, surgiu o interesse em desenvolver estratégias para controle de processos com ferramentas de inteligência artificial: algoritmos genéticos, redes neurais e lógica *fuzzy*. Neste trabalho foi dada ênfase às redes neurais.

As redes neurais são consideradas aproximadores universais que permitem modelar processos complexos altamente não lineares com múltiplas entradas e múltiplas saídas (MIMO), sendo a sua aplicação extensa em muitas áreas das engenharias de processos. As redes neurais são apropriadas para a criação de modelos de processo a partir de um conhecimento implícito contido em um conjunto de dados, enquanto os sistemas *fuzzy* são adequados para a criação de modelos a partir de um conhecimento explícito originário de especialistas humanos (LEITE, 2011).

Outro componente para a abordagem a ser desenvolvida no presente trabalho é uma tecnologia de controle que possa ser aplicada conjuntamente com as técnicas de inteligência artificial, sendo a alternativa de interesse o Controle Preditivo com Modelo (MPC), que é uma tecnologia de controle avançado que tem tido grande impacto na prática industrial, com aplicações principalmente em

refinarias de petróleo, plantas petroquímicas, plantas químicas, e indústrias de papel e celulose. Nessas indústrias o MPC tornou-se o método de escolha para problemas difíceis de controle multivariável que incluem restrições de desigualdade (OGUNNAIKE e RAY, 1994; QUIN e BADGWELL, 2003). Entretanto, relativamente poucas aplicações têm sido reportadas em outras indústrias de processo, mesmo que o MPC seja uma abordagem bastante geral e não limitada a um tipo de indústria (KWONG, 2005).

O presente trabalho busca integrar técnicas de inteligência artificial com técnicas de controle preditivo de processos. Para isso, a informação contida nos dados reais do processo é recolhida em modelos empíricos de processo, se interpreta-se a informação contida nesses modelos, e se utiliza para projetar estratégias de controle de processos nas quais se integram redes neurais em algoritmos de controle preditivo com modelo (MPC), para manter o sistema em condições desejadas de funcionamento, mesmo que ocorram perturbações dos tipos servo ou regulatório durante o processo de produção de bioetanol. A metodologia proposta é útil quando se tem abundância de dados reais de operação do processo, mas não se tem um modelo matemático preciso que o descreva.

1.2 . OBJETIVOS

O presente trabalho tem como objetivo principal, desenvolver uma abordagem de projeto de sistemas de controle preditivo com modelos empíricos de redes neurais para a operação estável de um processo fermentativo extrativo de produção de bioetanol; e como objetivos específicos:

- 1) Avaliar o uso de redes neurais para a construção de modelos empíricos nos processos de produção de bioetanol e sua aplicação conjunta com algoritmos genéticos como ferramenta para otimização desses processos.
- 2) Desenvolver algoritmos baseados em funções de transferencia para o estudo das relações causa-efeito em modelos empíricos baseados em redes neurais.

- 3) Implementar algoritmos baseados na integração da modelagem empírica de processos com redes neurais e as técnicas de controle baseadas em controle preditivo com modelo (MPC), para o controle da operação de um processo de produção de bioetanol: fermentativo, extrativo e contínuo com sistema de destilação flash à vácuo.
- 4) Implementar a abordagem de projeto de sistemas de controle proposta e avaliar o seu desempenho por simulação no controle da operação do processo de produção de bioetanol considerado como estudo de caso.

1.3. ORGANIZAÇÃO DO TRABALHO

O presente trabalho foi organizado em 12 capítulos, os quais têm o intuito de apresentar o conteúdo deste trabalho de forma clara e coerente.

No Capítulo 1 são apresentadas: a introdução fazendo referência à motivação para realizar o trabalho, os objetivos e a organização do trabalho, ressalta-se a tarefa desafiadora de utilizar volumosos conjunto de dados reais do processo contendo uma distribuição aleatória de erros naturais e grosseiros e ruído sistemático em alguns sinais, para finalmente projetar estruturas de controle que, com grande precisão levam o sistema para as condições de operação desejada. Outro componente é o uso de ferramentas denominadas na literatura como caixa preta. A organização, portanto, revela a necessidade inicial de conhecer melhor as ferramentas a serem utilizadas, especificamente as redes neurais e os algoritmos genéticos.

Nos Capítulos 2 a 5 são apresentados conceitos teóricos fundamentais ao entendimento do trabalho desenvolvido.

No Capítulo 2 são apresentados os conceitos do processo de produção de bioetanol no que se refere às tecnologias de produção de primeira e segunda gerações (1G e 2G), com ênfase na tecnologia 1G, onde o bioetanol é obtido por fermentação a partir do caldo e/ou melaço de cana-de-açúcar. É dada ênfase à construção de modelos fenomenológicos, baseados nas cinéticas de crescimento

de biomassa, consumo de substrato e formação de produto, e nas características dos modelos de equilíbrio líquido-vapor que são utilizados na previsão das condições de operação do tanque flash utilizado no processo de fermentação extrativa contínua.

No Capítulo 3 são apresentados os conceitos das redes neurais, uma das ferramentas fundamentais para o desenvolvimento do trabalho. A informação contribui ao melhor entendimento dos processos de aprendizagem das redes, assim como dos fenômenos de *overtraining* e *overfitting*, que têm importante efeito na capacidade de generalização das redes neurais.

No Capítulo 4 é realizada uma revisão bibliográfica do estado da arte na identificação de processos onde são identificadas importantes etapas nesta tarefa e as diferentes alternativas existentes. Nesse capítulo são desenvolvidas etapas específicas para a identificação de processos com redes neurais, sendo que a informação permitiu identificar alternativas de ação mais adequadas para a realização do presente trabalho.

No Capítulo 5 são apresentados os conceitos do controle de processos com modelo, que é utilizado na abordagem de projeto de sistemas de controle no presente trabalho. A ênfase é dada aos conceitos e características dos controladores SISO e MIMO: MPC, DMC, QDMC, com otimização linear e não linear.

Nos Capítulos 6 a 11 são apresentados desenvolvimentos e resultados das diferentes aplicações realizadas.

No Capítulo 6 são integradas redes neurais e os algoritmos genéticos para realizar tarefas de modelagem e otimização em etapas do processo de produção de bioetanol. O estudo de caso é um processo de hidrólise enzimática para produção de etanol 2G e é demonstrado como com o uso de redes neurais e algoritmos genéticos é possível determinar a carga ótima de enzimas no processo.

No Capítulo 7 são apresentados resultados da identificação de processos tendo como estudo de caso o processo de fermentação extrativa contínua com evaporador flash, desde o pré-tratamento dos dados até a modelagem. O caso escolhido para esse Capítulo permite fazer uma contribuição ao desenvolvimento de “software sensors” melhorando resultados da literatura, ao introduzir variáveis adicionais e demonstrar que o ruído do sinal limita a precisão dos “software sensors”.

No Capítulo 8 desenvolve-se um novo método para interpretar a informação contida na caixa preta (modelo de rede neural), o método é aplicado e orienta a utilização de variáveis de entrada nos capítulos seguintes da tese.

No Capítulo 9 é desenvolvido o controle DMC (Dynamic Matrix Control). O conhecimento da importância das variáveis desenvolvido no Capítulo 8 é aproveitado e utilizam-se técnicas complementares para reduzir o número de variáveis de entrada nos modelos, é feito um estudo comparativo com uso de modelos com diferentes números de variáveis nos projetos de sistema de controle DMC linear.

No Capítulo 10 é desenvolvido o QDMC (Quadratic Dynamic Matrix Control) utilizando o conhecimento do processo obtido nos capítulos anteriores. São estudadas diversas alternativas de modelos de planta internos dos controladores e modelos de processo com diferentes números de variáveis de entrada para o controle da concentração de etanol no fermentador. Nesse capítulo projetam-se com bons resultados sistemas de controle preditivo não linear (QDMC) para malhas simples.

No Capítulo 11 se apresentam-se resultados da abordagem utilizada para o projeto de um sistema de controle descentralizado. Com base na aplicação do método desenvolvido no Capítulo 8, é obtido um modelo MIMO do processo com número reduzido de entradas que foi utilizado para treinar os modelos internos dos controladores das malhas de controle. Nesse Capítulo foram utilizados diversos critérios para emparelhamento de variáveis nas malhas de controle e com sucesso

foi desenvolvido um sistema de controle preditivo com modelo não linear descentralizado para controle simultâneo das concentrações de etanol no fermentador e de etanol condensado na corrente de saída do tanque flash.

No Capítulo 12 são apresentadas as conclusões gerais sobre o trabalho realizado e, em seguida, são mencionadas algumas sugestões para trabalhos futuros.

1.4. REFERÊNCIAS BIBLIOGRÁFICAS

1. Atala, D.I.P., & Maciel, R. (2006). Experimental development, optimization and control of a Continuous extractive production process for the ethanol production. Postdoctoral research. Process 151335/2005-2, FAPESP, Brazil.
2. Biegler, L.T. and Rawlings, J.B. *Optimization approaches to nonlinear model predictive control*. In Y. Arkun and W.H. Ray, editors, Chemical Process Control-CPCIV, pp. 543-571. CACHE, 1991.
3. Costa, A.C.; Meleiro, L.A.C.; Maciel Filho, R. (2002). Non-linear predictive control of an extractive alcoholic fermentation process. *Process Biochemistry*, v. 38(5), pp. 743-750.
4. Dechechi, E.C., **Controle avançado preditivo adaptativo"DMC multivariável adaptativo"**, Tese de Doutorado, FEQ/UNICAMP, Campinas, SP, 1998.
5. EMBRAPA. **Produção de Etanol: Primeira ou Segunda Geração?**. Circular Técnica, Brasília, 2011.
6. Henson, M. A., & Seborg, D. E. (1992). Nonlinear control strategies for continuous fermenters. *Chemical Engineering Science*, v. 47, pp. 821- 835.
7. Kwong, W.H. **Introdução ao Controle Preditivo com Matlab**. São Carlos, SP. Brasil: EdUFSCar, 2005
8. Leite, M.S. **Projeto, montagem e instrumentação de um protótipo experimental de sistema de polimerização para o desenvolvimento e implementação de diferentes técnicas de controle inteligentes**. Tese de Doutorado. UNICAMP. Brasil. 2011.
9. Ogunnaike, B.A., Ray, W.H. (1994). Process **Dynamic Modeling and Control**. New York NY, USA: Oxford University Press. Oxford.
10. Pereira Jr., N., Couto, M.A.P.G., Santa Anna, L.M.M. Biomass of lignocellulosic composition for fuel ethanol production and the context of biorefinery. In Séries on Biotechnology. Rio de Janeiro, Brasil: Ed. UFRJ (2008).
11. Qin, S.J. and Badgwell, T.A. (2003). A Survey of Industrial Model Predictive Control Technology, *Control Engineering Practice*, v. 11, pp. 733-764.

12. Santos, M. G. R. S. e Mothe, Ch. G. (2007). Fontes alternativas de energia. *Analytica*. n. 32, pp. 56-70.
13. Schmidell, W. and Facciotti, M. C. R.. **Biorreatores e Processos Fermentativos**. In: Schmidell, Willibaldo *et al.* (Coord.). *Biotecnologia Industrial: Engenharia Bioquímica*. São Paulo: Edgar Blücher, v. 2, pp. 179-192. 2001
14. Silva, F.L.H., **Modelagem, Simulação e Controle de Fermentação Alcoólica Contínua Extrativa**, Tese de Doutorado, FEA/UNICAMP, Campinas, SP, 1998.
15. Rivera, E.C.; Atala, D. I. P.; Maugeri, F.; Costa, A. C.; Maciel Filho, R. (2010). Development of real-time state estimators for reaction-separation processes: A continuous flash fermentation as a study case. *Chem. Eng. Process*, v. 49, pp. 402-409.

CAPÍTULO 2. A PRODUÇÃO DE BIOETANOL

2.1. INTRODUÇÃO

O principal combustível utilizado como substituto do petróleo para os veículos de transporte é o bioetanol. O bioetanol é produzido no Brasil a partir da cana-de-açúcar por fermentação. Comparado com combustíveis fósseis como o petróleo e a gasolina, o etanol provoca menos poluição ambiental e aumenta a octanagem do combustível utilizado. Quando se mistura com a gasolina, oxigena a mistura de modo que ela pode-se queimar mais completamente, reduzindo assim as emissões de gases causadores do efeito de estufa. Além disso, o etanol é obtido de uma fonte de energia renovável e não de um recurso finito como os combustíveis fósseis.

A produção de bioetanol é beneficiada pela aplicação de leis de proteção ambiental em diversos países para controle da poluição, e sua investigação é de interesse dos governos porque permite a utilização de fontes de energia alternativa renováveis.

Em 2005, os Estados Unidos estabeleceram um padrão nacional de combustíveis renováveis. Essa legislação estabelece uma meta de duplicar o uso de combustíveis renováveis no país em sete anos, para 2012 a meta é de produzir 7,5 bilhões de galões. O principal combustível renovável a ser utilizado será o etanol. Os Estados Unidos e outros países industrializados do mundo são dependentes de petróleo importado e a importação de petróleo continua a aumentar. Aproximadamente 97% da energia requerida pelo setor dos transportes é derivada do petróleo. O etanol pode reduzir a poluição veicular em cerca de 54% (BYUNG-HWAN, 2007).

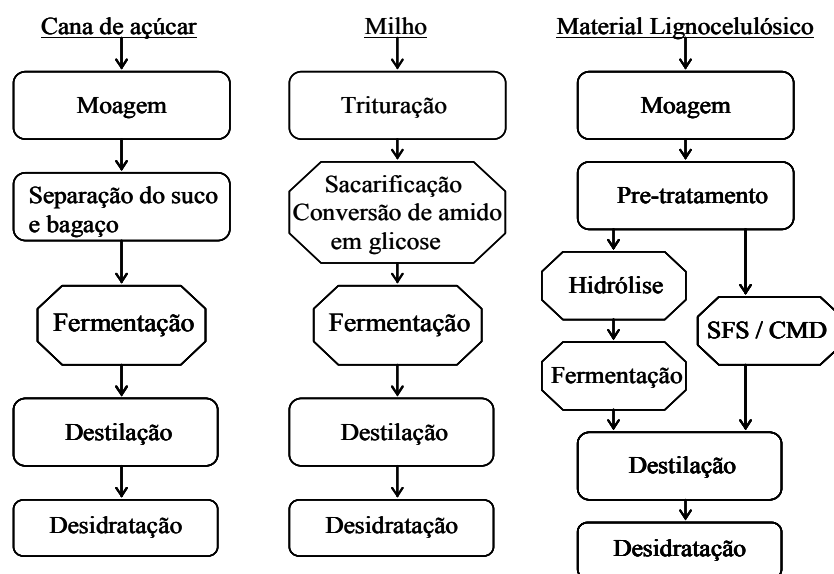
Segundo a OECD-FAO¹ (2011) os preços do etanol no mundo aumentaram mais de 30% em 2010 e a produção mundial de etanol está projetada para continuar se expandindo nos próximos 10 anos rapidamente. Portanto, as tendências mostram que o mercado do bioetanol é um mercado em crescimento.

¹ OECD: Organização para Cooperação Econômica e Desenvolvimento.
FAO: Organização das Nações Unidas para Agricultura e Alimentação.

O bioetanol pode ser produzido tendo como fonte matérias-primas:

- 1) açucaradas, ex.: cana-de-açúcar, melaço, beterraba açucareira.
- 2) amiláceas, ex.: milho, mandioca, batata, trigo, cevada.
- 3) lignocelulósicas: ex.: bagaço, palha de cana de açúcar.

A tecnologia utilizada com matérias-primas açucaradas e amiláceas é chamada de tecnologia de primeira geração, 1G; e a utilizada com matérias-primas lignocelulósicas é denominada tecnologia de segunda geração, 2G. A Figura 2.1 mostra etapas importantes nos diferentes processos de produção.



SFS: Sacarificação e Fermentação Simultâneas.

CMD: Conversão Direta pelo Microrganismo.

Figura 2.1. Diferentes matérias-primas e processos para produção de etanol
Elaboração do autor com base na bibliografia citada nas referências.

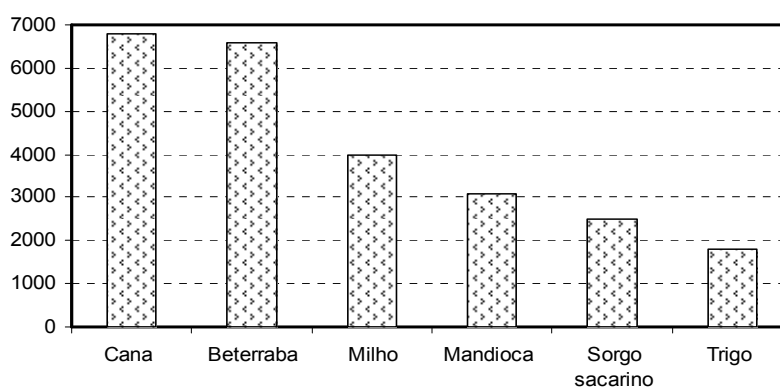
2.2. O BIOETANOL DE 1ª GERAÇÃO

Os maiores produtores de bioetanol no mundo são os EUA, que utilizam o milho como matéria-prima e o Brasil, que produz bioetanol a partir da cana-de-açúcar. Em 2009, os Estados Unidos (com 54%) e o Brasil (com 34%) responderam por 88% de todo o etanol produzido no mundo. O interesse pela produção de bioetanol nos EUA é conhecido, no período 2007 – 2009 a sua produção aumentou em 63,8%, enquanto para o mesmo período a produção do

Brasil aumentou em 16,9% (BIOFUELS PLATFORM, 2010). No ano 2010, os Estados Unidos tornaram-se, pela primeira vez, exportadores de etanol (OECD-FAO, 2011).

Segundo Bastos (2007), no Brasil a produção de cana-de-açúcar está concentrada na região centro-sul (85%). O estado de São Paulo responde por 60% da produção brasileira de cana. A produção é trabalho-intensiva, com cerca de 80% da cana cortada a mão (em São Paulo, a colheita é mecanizada apenas em 30% da área plantada). A cultura da cana-de-açúcar é semi-perene, pois pode ser colhida sem necessidade de replantio por cinco a sete safras anuais consecutivas. Segundo o Ministério do Desenvolvimento, Indústria e Comércio Exterior do Brasil (2012), na safra 2010/2011 foram colhidas 624 milhões de toneladas de cana-de-açúcar, e produzidos 27,6 bilhões de litros de álcool (8 bilhões de litros de álcool anidro e 19,6 de hidratado), com 421 unidades produtoras, sendo 168 produtoras de álcool (40%) e 253 de açúcar e álcool (60%). A produção brasileira de álcool evoluiu de 11,5 bilhões de litros na safra 1990/1991 para 15,4 bilhões de litros em 2004/2005 (VIEIRA, 2007) e 27,6 bilhões de litros em 2010/2011. O que representa 34% de aumento nos primeiros 15 anos e 79,2% de aumento nos últimos 6 anos.

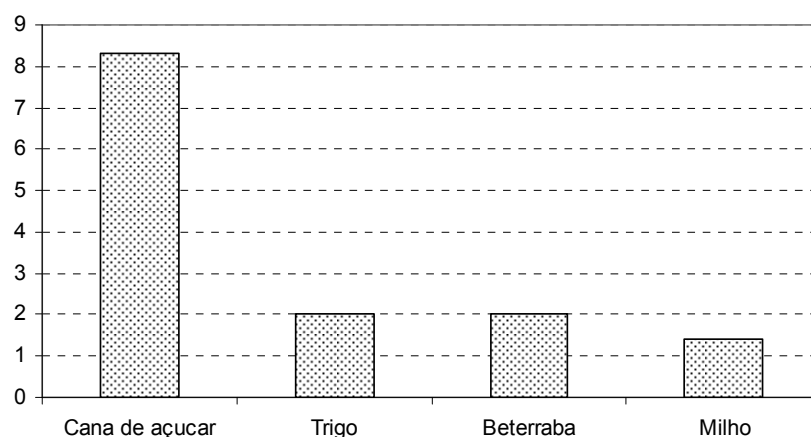
Algumas vantagens comparativas do bioetanol da cana-de-açúcar do Brasil, em comparação ao bioetanol obtido usando outras matérias primas são apresentadas nas figuras seguintes.



Fonte: BNDES, 2008.

Figura 2.2. Produtividade de Etanol (Litros por Hectare)

A Figura 2.2 mostra que a produtividade (Litros por Hectare) do bioetanol da cana-de-açúcar do Brasil é maior em uma proporção de 1,7/1 do que a do bioetanol de milho dos EUA, que é o maior produtor de bioetanol no mercado mundial. A beterraba também tem boa produtividade, no entanto, o processo industrial para produção de etanol a partir da beterraba depende do fornecimento de energia externa (eletricidade e combustível), enquanto a eletricidade para o etanol de cana-de-açúcar é fornecido pela queima do bagaço (WATANABE, 2009).



Fonte: Worldwatch Institute (2006)

Figura 2.3. Balanço energético: Energia contida no combustível/Energia Fóssil usada para produzi-lo.

A Figura 2.3 mostra que, na produção do bioetanol de milho utiliza-se uma quantidade muito maior (em proporção 6/1, aproximadamente) de energia fóssil do que a requerida para produzir a mesma quantidade de bioetanol utilizando cana-de-açúcar como matéria-prima.

O bioetanol da cana-de-açúcar do Brasil, também tem grande vantagem comparativa nos custos de produção. A International Energy Agency (IEA) (2007) afirma que os custos têm alta suscetibilidade para vários fatores, entre eles: matéria-prima, processo, terra, custo de mão-de-obra, etc. Os dados de várias fontes mostram que, o custo estimado de produção do bioetanol de cana-de-açúcar do Brasil é de 22 a 35 centavos de dólar por litro, e do bioetanol de milho

dos EUA é de 40 a 60 centavos de dólar por litro (HENNIGES e ZEDDIES, 2005; IEA, 2007).

2.2.1. A PRODUÇÃO DE ÁLCOOL A PARTIR DA CANA-DE-AÇÚCAR

A Figura 2.4 mostra um diagrama de blocos do processo de produção de bioetanol a partir de cana-de-açúcar com os tratamentos convencionais.

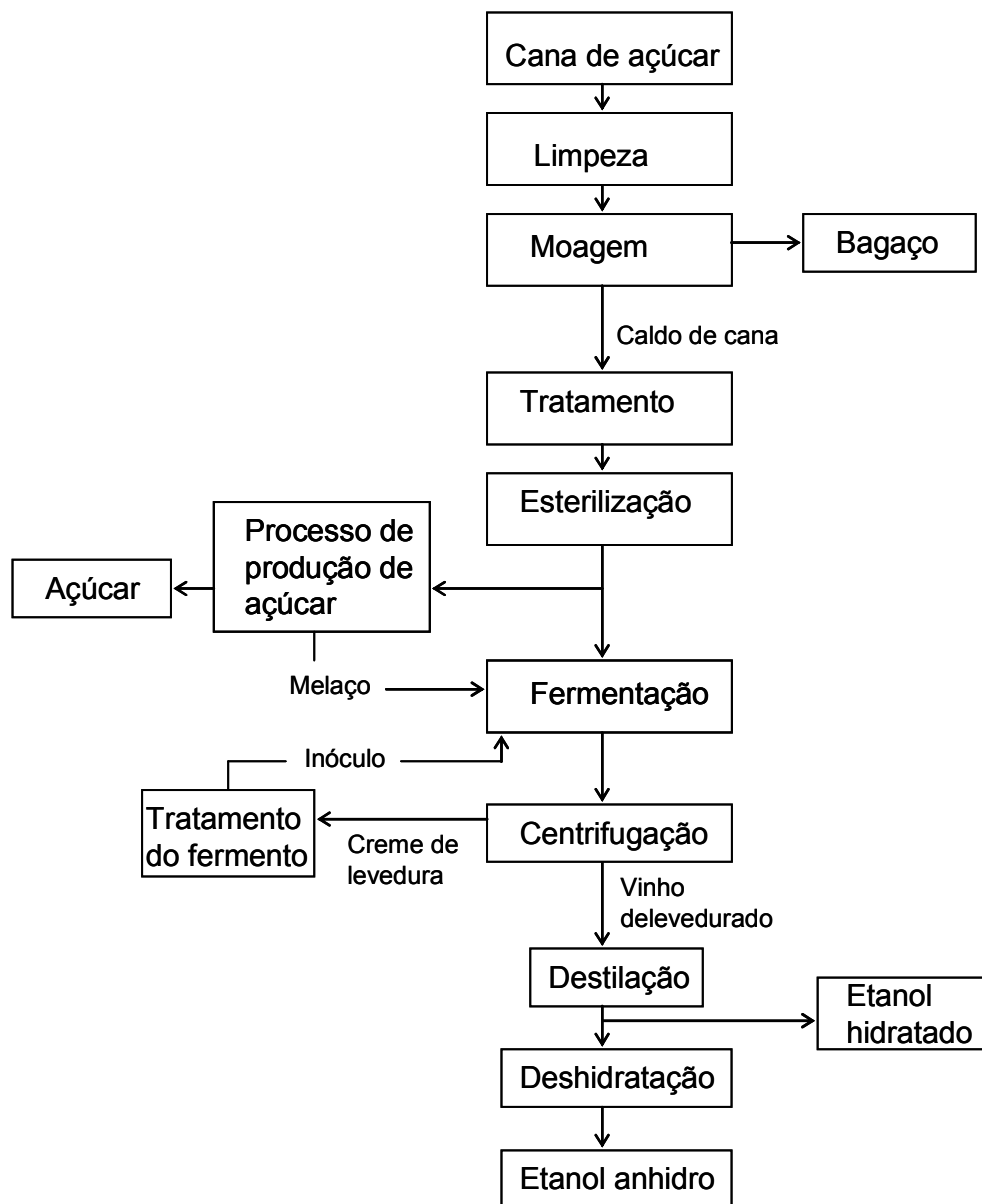


Figura 2.4. Diagrama de blocos da produção de bioetanol de cana-de-açúcar
Elaboração do autor com base na bibliografia citada nas referências.

O processo de produção de etanol mostrado na Figura 2.4, pode ser dividido em três etapas: a) preparo da matéria-prima, b) reação e c) separação.

2.2.1.1. PREPARO DA MATÉRIA PRIMA

Após o processo de limpeza, a cana desfibrada é submetida a moagem para separar o caldo da cana rico em açúcar do bagaço, que é a parte fibrosa. O bagaço obtido no processo de extração é utilizado como matéria prima para produção de energia nas caldeiras. O caldo de cana é peneirado e logo é submetido a processos de aquecimento, clarificação e resfriamento. O motivo desses tratamentos é que nesta etapa deve ser obtido um caldo com a menor quantidade de impurezas e contaminantes possível para o bom andamento do processo fermentativo. O caldo devidamente esterilizado pode ser misturado com o melaço, para obter o mosto que é uma solução de açúcar pronta para fermentar. O melaço é um produto residual do processo de produção de açúcar que ainda tem alto teor de sacarose e açúcares redutores (glicose e frutose como resultante da desagregação da sacarose) e pode ser utilizado como matéria-prima para a produção de bioetanol, através da fermentação. Assim, a produção de bioetanol a partir da cana-de-açúcar pode ser baseada em fermentação diretamente a partir do suco de cana, ou das misturas de suco e melaço (BNDES, 2008; DIAS *et al.*, 2011).

A obtenção do bioetanol a partir do bagaço de cana ou de outros subtratos lignocelulósicos requer hidrólise ácida ou enzimática para se obter a solução açucarada para fermentar. Embora seja possível, o processo não é ainda economicamente viável.

2.2.1.2. REAÇÃO

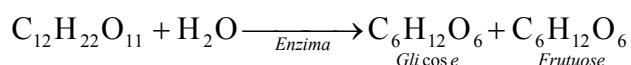
O mosto é fermentado nas dornas de fermentação utilizando microrganismos, usualmente fungos unicelulares da espécie *Saccharomyces cerevisiae*, que se alimentam do açúcar liberando gás carbônico e álcool. O mosto

é fermentado por um período de 8 a 13 horas dependendo do tamanho da dorna e da temperatura na qual é conduzida a fermentação, dando origem ao vinho com um teor alcoólico entre 7% e 12 % (v/v). As leveduras podem ser recuperadas do vinho por centrifugação. Desta forma, após a fermentação são guardadas para posterior utilização, e o vinho é enviado para as colunas de destilação (BNDES, 2008; COHEN, 2011; DIAS *et al.*, 2011).

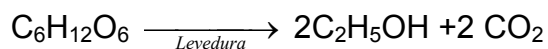
A fermentação é um bioprocesso complexo no qual os microrganismos, para viver, reproduzir e formar os produtos obtêm nutrientes a partir de seus arredores, necessitam de açúcar ou de outros carboidratos como fontes de carbono, e obtêm energia por quebra de ligações químicas. Amônia ou um sal de amônia (NH₄Cl, (NH₄)₂SO₄) é usado como fonte de nitrogênio, o hidrogênio pode ser obtido a partir de água ou substratos orgânicos. Outras substâncias que promovem o crescimento são P, S, K, Ca, Mg, que são adicionados como sais inorgânicos.

Segundo Basso (2011) existem duas vias pelas quais a sacarose é metabolizada por *Saccharomyces cerevisiae*. Uma delas é conhecida como a via extracelular, na qual a sacarose é hidrolisada no ambiente extracelular pela enzima invertase, produzindo glicose e frutose, que são subsequentemente transportadas para o interior das células e posteriormente consumidas na via glicolítica. A segunda via é a via intracelular, na qual a sacarose é primeiramente transportada através da membrana plasmática, e em seguida hidrolisada no ambiente intracelular por ação da enzima invertase intracelular ou por maltases. Dado que, a fração extracelular é a fração majoritária da atividade total de invertase (>90%), postula-se que a via extracelular seja a rota principal no metabolismo da sacarose por *Saccharomyces cerevisiae*.

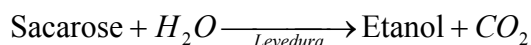
A hidrólise da sacarose em glicose e frutose é representada pela equação:



O etanol é produzido a partir das hexoses segundo a equação seguinte:



A transformação da sacarose em etanol como resultado das reações anteriores pode ser representada por:



Outros produtos podem ser formados nas reações de fermentação, entre eles, o glicerol, os ácidos orgânicos e os outros álcoois.

Segundo Guo *et al.* (2011) o glicerol é um importante sub-produto na produção anaerobia de etanol. Para sua síntese, a levedura *Sacharomyces cerevisiae* poderia consumir até 4% da provisão de açúcar em processos industriais típicos de produção de etanol. A levedura *Sacharomyces cerevisiae* é considerada a mais importante levedura produtora de glicerol. O glicerol tem importantes funções nos processos fisiológicos desse microrganismo, como a manutenção do equilíbrio redox $\text{NAD}^+ / \text{NADH}$ (a forma oxidada é reduzida da coenzima nicotinamida adenina dinucleótido) e a resposta ao estresse osmótico, que é causado por elevadas concentrações de açúcares ou sais no mosto, contaminação bacteriana, presença de sulfito no mosto e temperatura elevada (DIAS, 2008). Uma das consequências do estresse hiperosmótico é a rápida difusão de água da célula para o meio circundante (SCANES *et al.*, 1998).

O ácido succínico é outro subproduto nas reações de fermentação e possui atividade antibacteriana (DIAS, 2008).

A fermentação alcoólica é afetada por vários fatores, entre eles estão:

- 1) A temperatura.
- 2) As inibições da taxa de crescimento celular
- 3) Outros fatores.

A temperatura na qual é conduzida a fermentação tem efeitos críticos sobre a atividade dos microorganismos. Em geral, as leveduras são capazes de promover a fermentação alcoólica eficientemente entre 28°C e 35°C.

Temperaturas elevadas afetam o comportamento da levedura e diminuem o teor alcoólico do vinho, o que aumenta o consumo de energia durante a etapa de destilação e a razão volume de vinhaça/ volume de álcool produzido (DIAS, 2008). O efeito do aumento da temperatura é diverso e afeta a composição dos ácidos graxos na membrana lipídica, de maneira que, os ácidos graxos saturados se incrementam às custas dos insaturados, modificando a fluidez da membrana como resposta da célula de levedura que tenta se adaptar à mudança na temperatura (D'AMORE, 1991). Por outro lado a proteína nativa é mantida por um delicado balanço de forças não covalentes, como ligações de hidrogênio, pareamento de íons, interações hidrofóbicas e força de Van der Waals. Com o aumento da temperatura, essas interações são rompidas e a proteína se desdobra, produzindo-se a desnaturação, que na maioria dos casos é irreversível (GOMES *et al.*, 2007).

Outros fatores que afetam o desempenho da levedura são: qualidade da matéria-prima utilizada na fermentação pH do processo; aditivos químicos, tais como ácidos, biocidas, antibióticos, íons Ca e P; contaminação microbiana, que leva à formação de ácidos e outros compostos; floculação celular; qualidade da centrifugação do fermento; tipo de processo fermentativo (batelada alimentada, contínuo); dimensionamento da planta; limpeza e assepsia (NETO, 2008).

As inibições da taxa de crescimento celular são devidas ao produto, ao substrato e aos subprodutos. Segundo Maiorella *et al.* (1983), na fermentação com *Sacharomyces cerevisiae*, a inibição pelo produto etanol começa em torno de 25 g/l e é total em 95 g/l. Os danos causados pelo etanol à levedura resultam em alterações na permeabilidade e organização da membrana plasmática das células e das membranas de organelas celulares, resultando no vazamento de cofatores essenciais e coenzimas. Outros mecanismos propostos para os efeitos inibitórios do etanol incluem a alteração dos fluxos de prótons, a inibição e desnaturação de várias proteínas intracelulares e enzimas glicolíticas, e a inibição do transporte de açúcar e de aminoácidos (D'AMORE, 1991).

Em referência à inibição pelo substrato, Maiorella *et al.* (1983) reporta que um alto nível de inibição (80% de redução na massa celular) ocorre com concentrações de glicose de 380 g/l. As altas concentrações de substrato inibem o crescimento da levedura e o seu desempenho na fermentação como resultado da pressão osmótica elevada e baixa atividade de água (D'AMORE, 1991).

A inibição produzida por vários subprodutos da fermentação com extração contínua de produto utilizando vácuo (vacuum fermentation) foi estudada por Maiorella *et al.* (1983) A Tabela 2.1 mostra as concentrações dos subprodutos para um estado inicial de inibição que corresponde a 20% de redução na produção de massa celular, e para um estado de alta inibição identificado por 80% de redução na produção de massa celular.

Tabela 2.1. Inibição por subprodutos na fermentação com *Sacharomyces cerevisiae* para a produção de etanol.

Subproduto	Estado de inibição inicial. Concentração (g/l)	Estado de alta inibição. Concentração (g/l)
2,3-Butanodiol	55	90
Glicerol	200	450
Acetaldeído	2,8	5,0
Ácido fórmico	1,2	2,7
Ácido acético	0,9	7,5
Ácido láctico	17	38
1-Propanol	2,0	12
2-Metil-1-butanol	2,8	3,5

Fuente: Maiorella *et al.* (1983).

Os resultados da Tabela 2.1 foram obtidos sob condições experimentais detalhadas em Maiorella *et al.* (1983), que correspondem ao processo de fermentação com extração contínua de produto utilizando vácuo. Este é um processo que seletivamente concentra produtos não voláteis no caldo de fermentação, e onde os produtos secundários podem concentrar-se até o ponto que eles chegam a ser tóxicos para a levedura, afetando a produtividade do processo de produção de etanol.

2.2.1.3. SEPARAÇÃO

A separação do vinho advindo da fermentação se realiza em unidades de destilação e desidratação para se obter o etanol hidratado ou o etanol anidro.

Durante a destilação é recuperado o bioetanol na forma hidratada, com aproximadamente 96 ° GL (porcentagem em volume), correspondendo a cerca de 6% de água, em peso, deixando a vinhaça ou "vinhoto" como resíduo em uma proporção média de 10 a 13 litros por litro de bioetanol hidratado produzido. O bioetanol hidratado pode ser armazenado como um produto final ou pode ser enviado para a coluna de desidratação. Mas como é uma mistura azeotrópica, seus componentes não podem ser separados por uma destilação simples. A tecnologia mais utilizada no Brasil é de desidratação por meio da adição de ciclohexano, quando ele faz uma mistura azeotrópica ternária, com um ponto de ebulição inferior ao do bioetanol anidro. Na coluna de desidratação, o ciclohexano é adicionado na parte superior e é removido pela parte inferior o bioetanol anidro, com aproximadamente 99,7 ° GL, ou seja, 0,4% de água, em peso. Além disso, a desidratação do bioetanol pode ser feita por destilação extrativa com monoetilenoglicol ou pela adsorção com peneiras moleculares, onde são usados sólidos porosos (Zeólita) seletivos que aprisionam preferencialmente água e liberam etanol anidro. (MEIRELLES, 2006; BNDES, 2008).

2.2.1.4. OS PROCESSOS DE FERMENTAÇÃO

Os processos fermentativos podem ser classificados conforme a maneira através da qual o substrato é adicionado e o producto retirado, assim temos:

- 1) Processo batelada
- 2) Processo batelada alimentada
- 3) Processo contínuo

2.2.1.4.1. Processo batelada

O processo batelada é um processo descontínuo em que todos os componentes necessários, com exceção das substâncias para regular o pH e dos agentes antiespumantes são colocados no reator no início do processo de

fermentação. Durante o processo, não há entrada nem saída de fluxos. O ciclo completo de operação compreende as etapas de carga, inoculação, fermentação, descarga e limpeza do equipamento. O mosto a ser fermentado é inoculado por uma cultura nova, recém-ativada em laboratório ou por um corte em estado final de fermentação na dorna. A fermentação é considerada concluída quando cessa a atividade da levedura por falta de substrato ou excesso de produto inibidor (etanol). O etanol é recuperado por destilação e o reator é preparado para a nova batelada (FERREIRA, 2005). Segundo Schmidell e Facciotti (2001) esse processo é o mais seguro quando se tem problema de manutenção e condições de assepsia, pois ao final de cada batelada o reator pode ser esterilizado juntamente com o meio de cultura, e se podem fazer todos os controles necessários no inóculo, a fim de assegurar a presença única do micro-organismo responsável pelo processo. Segundo Maiorella *et al.* (1981) o processo é lento pelo tempo que é utilizado em limpeza e preparação do reator para cada batelada.

2.2.1.4.2. Processo batelada alimentada

O processo batelada alimentada é conhecido desde 1900, sendo utilizado para regulação do crescimento de *Saccharomyces cerevisiae* (CARVALHO *et al.* 1975) por meio de um processo de fermentação descontínua continuamente alimentada de nutrientes; é uma técnica na qual um ou mais nutrientes são fornecidos gradualmente ao bioreator durante o cultivo e no qual os produtos permanecem no mesmo até o final do processo. Nesse processo, a adição de substrato permite controlar a velocidade de reação e diminuir o efeito inibitório causado pelo excesso de substrato (LESSMANN, 1993).

No Brasil, o processo de batelada-alimentada com reciclo de células é chamado de processo Melle-Boinot, esse processo surgiu na França na década de 1930, e sua principal característica é a recuperação das leveduras do vinho por centrifugação. Segundo Viegas (2003), até o início dos anos 60, o processo batelada simples era utilizado na fermentação do melaço. A utilização do processo Melle-Boinot nas plantas de produção de etanol brasileiras se generalizou no final da década de 60 e nos anos 70. Quando o Programa Nacional do Alcool

(Proálcool) foi implantado (1976), as destilarias instaladas foram equipadas com esse processo (ZARPELLON E ANDRIETTA, 1992).

O proceso se baseia na alimentação de substrato sob condições controladas até atingir o volume do fermentador (PORTO, 2005) e na reciclagem do fermento utilizado, que é separado do vinho por centrifugação. Após ser separado, o fermento é submetido a tratamento adequado e encaminhado para a dorna de fermentação como inóculo.

Segundo Almeida (1960) citado por Andrietta (1994), o processo Melle-Boinot tem as seguintes vantagens: economia de açúcar devido a menor reprodução celular elevando o rendimento em etanol, eliminação de contaminantes pela centrifugação do vinho (separação de células de levedura), fermentação mais pura devido ao tratamento de leite de levedura (tratamento ácido), eliminação da necessidade de cultura pura no preparo de pé-de-cuba, prática exigida no processo clássico, diminuindo, portanto, a complexidade das operações da planta.

2.2.1.4.3. Proceso contínuo

No processo contínuo os equipamentos passam a ser utilizados ininterruptamente, o meio é adicionado de forma contínua e os produtos de fermentação também são continuamente removidos a uma vazão igual à de alimentação, o que permite que o volume seja mantido fixo ou constante e se tenha um fluxo contínuo do produto.

O meio de alimentação (mosto) é constituído de caldo tratado, nas unidades produtoras que só produzem álcool (destilarias autônomas), e de uma mistura de caldo tratado e melaço proveniente da fabricação de açúcar, nas unidades que produzem açúcar e álcool.

As células de leveduras, após terem sido submetidas ao tratamento ácido (fermento tratado) são misturadas com o meio de alimentação (mosto). Esta mistura é então enviada aos fermentadores.

Após ocorrida a transformação dos açúcares em etanol, o vinho fermentado contendo células de levedura (vinho bruto) é enviado para a unidade de separação. Esta unidade é constituída de separadoras centrífugas de prato que trabalham a uma velocidade angular em torno de 8000 r.p.m. Estas separadoras recebem o vinho bruto contendo de 30 a 45 g de células por litro de vinho e fornecem dois produtos: uma fase leve e outra pesada. A fase pesada, contendo entre 160 a 200 g de células por litro, é chamada de leite ou creme de levedura. Este creme é enviado para a unidade de tratamento ácido, onde o pH do creme diluído é ajustado a 2,0 – 2,5 com ácido sulfúrico concentrado. A fase leve, praticamente isenta de células de levedura (a concentração máxima permitida é 3 g de células por litro) é chamada de vinho delevedurado e é enviado à unidade de destilação (ANDRIETTA, 1994).

A eficiência média global de recuperação de açúcar situa-se em torno de 86%, se a usina opera com cana-de açúcar com conteúdo de sacarose de 14%. No caso do uso exclusivo para a produção de etanol, com uma tonelada de cana é possível obter 86 litros de bioetanol hidratado, ao passo que, se o objetivo é a produção de açúcar, além de 100 kg desse produto ainda é possível produzir 23 litros de bioetanol hidratado por meio do melaço (BNDES, 2008).

Segundo Borzani (1960) citado por Andrietta (1994) as vantagens dos processos fermentativos contínuos são: maior produtividade, maior uniformidade do produto e maior adaptabilidade ao controle automático.

Tosseto (2002) e Hendler (2011) mencionam as seguintes vantagens do sistema contínuo, em relação à batelada alimentada:

- 1) A operação em estado estacionário permite que a levedura trabalhe em condições ótimas sem perturbações.
- 2) Reduz o tempo não produtivo, menos operações de carga, descarga e limpeza das dornas, portanto aumenta a produtividade.
- 3) Menor custo de investimento, porque o volume total dos reatores e a necessidade de trocador de calor são menores.

- 4) Menor custo operacional, a automatização é mais barata e reduz gastos em mão de obra.

Siqueira (1997) menciona como pontos negativos na fermentação contínua:

- 1) A manutenção de um meio estéril por muito tempo é bastante difícil.
- 2) Pode haver ocorrência de mutantes não interessantes ao processo de fermentação, especialmente quando este é conduzido por longos períodos de tempo.

Além da fermentação contínua, segundo Silva (1998), outros itens que incrementam a produtividade de álcool, são:

- Utilização de cepas de leveduras mais resistentes ao etanol;
- Utilização de sistemas de múltiplos estágios, com ou sem aeração;
- Emprego de reatores tubulares com leveduras imobilizadas;
- Retirada do etanol do meio à medida que ele é produzido.

Na literatura há diversos estudos desenvolvidos em escala de laboratório, planta piloto e industrial visando melhorar a produtividade do processo de produção de álcool, a maioria deles dão ênfase a algum dos itens mencionados acima.

Silva (1998) estudou por meio da modelagem e simulações um processo de fermentação alcoólico contínuo conectado a um sistema de extração de etanol produzido enquanto ele é formado (tanque flash a vácuo), tentando, desta forma, aumentar o rendimento e a produtividade pela minimização da inibição celular pelo produto. A modelagem do processo baseou-se em cinética fermentativa e balanços de massa e energia. A otimização da fermentação alcoólica extrativa foi estudada através de simulação em computador, do método de planejamento fatorial e análise de superfície de resposta. Reportou que foi possível otimizar o processo para uma conversão de 99,2% dos açúcares do meio de alimentação, e uma produtividade de 21 g Etanol/L.h.

Andrietta (1994) propôs um sistema de 4 reatores de mistura perfeita ligados em série determinando as condições de operação de maior produtividade para que a conversão de açúcar a produto atinja valores da ordem de 99,3% no último reator com uma produtividade de 7,93 g Etanol/L.h.

Viegas (2003) estudou em escala de laboratório o arranjo ótimo de um sistema de fermentação constituído de reatores tipo torre para produção de etanol utilizando leveduras com características floculantes, sem a necessidade de se utilizar unidade de separação de células, visando assim à redução do custo de produção de etanol em torno de 15%. Reportou que o arranjo constituído de três reatores ligados em série com distribuição da alimentação 50% para o primeiro reator e 50% para o segundo reator, apresentou produtividade máxima de 27,39 g Etanol/L.h.

Atala (2004) desenvolveu estudos em escala de bancada de um processo fermentativo contínuo acoplado a um evaporador que funciona a vácuo do tipo Flash, utilizando a levedura *Saccharomyces cerevisiae* e melaço de cana-de-açúcar como substrato. Reportou resultados do processo de fermentação contínua extrativa com uma produtividade em etanol do sistema entre 8,25 e 10 g/L.h. Em trabalho posterior (ATALA, 2006) pesquisou as condições operacionais para a maximização do rendimento e produtividade em etanol e conversão total do substrato, para o mesmo sistema de fermentação alcoólica contínua extrativa, e reportou uma produtividade em etanol de 25 g/L.h, valor este superior ao encontrado no processo tradicional, em torno de 7,5 g/L.h.

Galassi (2007) realizou um estudo do processo de fermentação alcoólica contínua em que são utilizados reatores tubulares tipo torre e células auto-imobilizadas. Buscou-se nesse projeto a seleção entre cepas com características floculantes e perfis metabólicos adequados ao trabalho em tais reatores. O método proposto busca reduzir o custo de produção do etanol e obter melhorias no sistema de fermentação contínua com utilização de células auto-imobilizadas, devido a que nesse sistema inexistente a necessidade da utilização de unidades de

separação de células (centrifugas), reduzindo assim em torno de 15% o custo de produção de etanol. As linhagens utilizadas foram cepas do gênero *Saccharomyces cerevisiae* com características floculantes e formadoras de partículas esféricas (pellets) quando cultivadas em mesa agitadora circulatória. Reportou que as melhores médias para a produção de etanol foram de 22,29 g etanol/L.h. Segundo o autor é importante a manutenção de um ponto ótimo de operação para o trabalho com biorreatores tubulares tipo torre em sistema contínuo de produção de etanol utilizando linhagens de levedura floculantes.

Tang *et al.* (2010) desenvolveram um processo de fermentação contínua de etanol de dois estágios sem o uso de pré-tratamento com ácido sulfúrico para controle da contaminação bacteriana. Foi utilizado um método com alimentação de melaço diluído como meio, a cepa de levedura floculante KF-7 e reatores de tanque agitado. Segundo os autores o processo foi operado com sucesso por mais de um mês, com uma concentração de etanol no produto de 80 g / L e uma produtividade de etanol de 6,6 g / L. h.

Liu *et al.* (2009) estudaram a fermentação do etanol por células de *Saccharomyces cerevisiae* imobilizada em partículas magnéticas em um reator de leito fluidizado estabilizado magneticamente (MSFBR). A produção de etanol teórico, a produtividade e a concentração de álcool atingiram 95,3%, 26,7 g/L. h e 66 g/L, respectivamente, com uma concentração de glicose de 150 g/L na alimentação.

Basso (2011) assinala que como o etanol é produto derivado diretamente do metabolismo central do carbono, são possíveis estratégias que alterem o metabolismo da sacarose em *Saccharomyces cerevisiae* para se aumentar o fator de conversão deste dissacarídeo em etanol. Para este objetivo propõe fazer uso de uma estratégia de melhoramento genético pela qual se favorece a hidrólise de sacarose intracelular produzindo glicose e frutose. A hipótese é que este cenário leva à formação de mais etanol por mol de sacarose consumido em anaerobiose do que uma situação de hidrólise extracelular, que é a forma mais conhecida pela

qual a sacarose é metabolizada. Nos resultados de laboratório relata que conseguiu gerar uma linhagem de *Saccharomyces cerevisiae* com fator de conversão de sacarose a etanol 11% maior que a linhagem–referência.

Segundo Dias *et al.* (2011) algumas melhorias no processo que, do ponto de vista dos autores poderiam ter impacto para melhorar a situação atual dos sistemas industriais de produção de bioetanol no Brasil são as seguintes:

- 1) O uso de um sistema de limpeza da cana ao seco, ao invés de uma baseada no uso da água, pode diminuir as perdas de açúcar, bem como o consumo de água.
- 2) Um tratamento de caldo de cana com adição de ácido fosfórico e cal, pode aumentar a pureza do caldo e, conseqüentemente, melhorar a fase de fermentação.
- 3) A concentração de suco em evaporadores de efeito múltiplo pode diminuir o consumo de vapor no processo.
- 4) Uma menor temperatura de fermentação (28 ° C), para a produção de vinho com maior teor de etanol, diminuindo as perdas de açúcar e de etanol, bem como a geração de vinhaça.
- 5) Um processo de destilação de efeito duplo para a integração térmica entre condensadores das colunas e refeedores.
- 6) Otimização do processo de destilação extrativa com MEG para produção de bioetanol anidro, e economia de energia suprimindo a etapa de vaporização do bioetanol hidratado antes do processo de destilação extrativa.

Outras melhorias para o processo podem ser identificadas, tendo em conta o seu impacto na redução de custos na produção de etanol. Dale (1995) manifesta que os custos de produção de etanol podem ser reduzidos via: 1) redução de custos de substrato, 2) aumentar a eficiência da conversão de substrato em etanol, 3) redução dos custos de energia para a purificação e desidratação do etanol, 4) redução da quantidade de águas residuais efluentes que devem ser

tratadas. 5) redução dos custos de capital para a planta de processamento de etanol, 6) redução de mão de obra para a operação da planta de etanol, e 7) o aumento do valor dos subprodutos da planta de etanol.

2.3. O BIOETANOL DE 2ª GERAÇÃO

O bioetanol é produzido a partir da biomassa lignocelulósica (bagaço, palha de cana de açúcar, palha do milho, etc.). A conversão bioquímica da celulose a etanol envolve três etapas básicas:

1. Pré-tratamento, (que causa o rompimento da hemicelulose, a remoção e/ou transformação da lignina e o aumento do potencial de hidrólise da celulose),
2. Hidrólise da celulose (empregando ácidos ou enzimas celulolíticas para a obtenção de uma solução rica em glicose) e
3. Fermentação alcoólica (processos fermentativos alcoólicos da biotecnologia convencional).

Alguns resultados da análise composicional do bagaço são mostrados na Tabela 2.2.

Tabela 2.2. Resultados médios da análise composicional do bagaço.

Componente (% base seca)				Referência
Celulose	Hemicelulose	Lignina	Cinzas	
37 26-47 38	28 19-33 33	21 14-23 22	n.d. 1-5 3	Bon (2007) (*) Paturau (1989) (*) Trickett and Neytzell-de Wilde (1982) (*)

(*) trabalhos citados em Walford, 2008.

A celulose é um polímero linear formado pela união de até 150 000 unidades moléculas de D-glicose (BANERJEE E PANDEY, 2002). Perto de 50-90% da celulose em materiais lignocelulósicos é unida lateralmente por ligações de

hidrogênio e forma uma estrutura cristalina. A porção residual e menos ordenada é em geral nomeada como celulose amorfa (BRIGHAM *et al.*, 1996).

A lignina é uma macromolécula constituída de moléculas orgânicas aromáticas, a unidade básica tem uma estrutura fenilpropanóide. Três monômeros que formam quase toda a lignina encontrada na natureza são os álcoois p-cumarílico, coniferílico e sinapílico. Os grupos funcionais principais na estrutura da lignina são: metoxila, hidroxila, eter e carbonila (SALIBA *et al.*, 2001). O objetivo principal da lignina é dar à planta apoio estrutural, a lignina protege os polissacarídeos da parede celular da degradação microbiana, assim comunicando resistência à decadência, é também um dos principais fatores limitantes na conversão de biomassa vegetal a biocombustíveis (VALHOLME *et al.*, 2010).

A hemicelulose consiste de cadeias ramificadas de açúcares cujas unidades incluem principalmente aldopentoses, como xilose e arabinose, e algumas aldohexoses, como glicose, manose, e galactose. Além de um alto grau de polimerização, o imero de hemicelulose normalmente tem substituintes na cadeia principal ou em seus ramos. A variedade de vínculos, ramificação e diferentes unidades monoméricas contribuem para a complexa estrutura da hemicelulose. Dentro da biomassa, hemicelulose está ligada a lignina e celulose por ligações covalentes, mas porque poucas são as ligações de hidrogênio envolvidas, é muito mais facilmente quebrada que a celulose cristalina (BRIGHAM *et al.*, 1996).

2.3.1. PRÉ-TRATAMENTO

Para o uso de materiais lignocelulósicos como o bagaço de cana para a produção de biocombustíveis, é necessário primeiramente submeter o material a uma etapa de pré-tratamento para facilitar o posterior ataque enzimático. Esta é uma das etapas mais custosas devido à utilização de reagentes de preço elevado, além da geração de produtos inibidores (caso de pré-tratamentos ácidos e explosão de vapor) para as etapas posteriores de hidrólise e fermentação.

Diferentes métodos de pré-tratamento têm sido estudados, tais com os pré-tratamentos: físicos (FAN *et al.*, 1982; CHANG E HOLTZAPPLE, 2000;

RAMOS *et al.*, 1992); físico-químicos (DALE e MOREIRA, 1982; TEYMOURI *et al.*, 2005; GOLLAPALLI, 2002); químicos com agentes ácidos (JACOBSEN e WYMAN, 2000, MOISER *et al.*, 2005; HSU, T. A., 1996); químicos com agentes básicos (CHANG *et al.*, 1997; CHANG *et al.*, 1998; KAAR e HOLTZAPPLE, 2000; KIM e HOLTZAPPLE, 2005; KIM e HOLTZAPPLE, 2006; RABELO *et al.*, 2008, RABELO, 2010; FUENTES, 2009); químicos com solventes orgânicos (CHUM *et al.*, 1988; MOISER *et al.*, 2005); ozonólise (VIDAL e MOLINIER, 1988); biológicos (LEE, 1997; KIRK e FARRELL, 1987) e combinados (PLAYNE, 1984; HSU, 1996).

2.3.2. HIDRÓLISE

A hidrólise, como o próprio nome descreve, envolve a quebra de ligações por adição de água. A etapa de hidrólise atualmente é realizada por duas rotas: a hidrólise ácida ou enzimática.

2.3.2.1. HIDRÓLISE ÁCIDA

Os açúcares fermentescíveis provenientes dos materiais lignocelulósicos podem ser obtidos por hidrólise com ácido concentrado ou com ácido diluído. Na hidrólise com ácido concentrado a principal desvantagem é que ela requer equipamentos de alta qualidade em aço inoxidável, o que aumenta consideravelmente o custo de capital.

A fração hemicelulose é mais fácil de ser hidrolisada do que a fração de celulose, e os açúcares resultantes da hemicelulose são expostos à degradação de derivados. Esses subprodutos diminuem a produção de açúcar e podem causar inibição na fermentação de etanol. A recuperação do ácido utilizado, que é geralmente exigente em energia, é importante a fim de tornar o processo economicamente viável.

Em contraste com a hidrólise ácida concentrada, uma baixa concentração de ácidos é aplicada no processo de ácido diluído, mas em altas temperaturas. O processo de ácido diluído normalmente consiste em duas etapas de hidrólise distintas. Na primeira etapa, a fração hemicelulose é degradada a açúcares

solúveis em água. Na segunda etapa, a celulose é convertida em glicose. Isso minimiza a formação de subprodutos devido ao desnecessariamente longo tempo de residência. No entanto, quantidades consideráveis de compostos tóxicos podem ser encontradas nesses hidrolisados por causa da temperatura mais alta aplicada (VON SIVERS e ZACCHI, 1995, SZENGYEL *et al.*, 2000).

Segundo Fuentes (2009) a técnica hidrólise ácida apresenta grandes desvantagens devido à formação de compostos inibidores produzidos pela degradação da celulose e da lignina, entre eles estão os derivados do furano, o furfural, os ácidos alifáticos, acético, fórmico, derivados fenólicos, esteróis, cera e outros.

2.3.2.2. HIDRÓLISE ENZIMÁTICA

As enzimas são utilizadas como catalisadores para separar ou hidrolisar as cadeias de hemicelulose e celulose, formando seus respectivos açúcares redutores, inclusive a glicose. Ao contrário dos catalisadores comuns, as enzimas apresentam uma elevada especificidade em relação ao substrato e sua utilização reduz a obtenção de subprodutos indesejáveis na reação, diminuindo assim os custos de separação dos produtos (SEGEL, 1975).

Diversos fatores devem ser considerados importantes para realizar a hidrólise enzimática, Chang e Holtzaple (2000) indicam que o conteúdo de lignina e o índice de cristalinidade do material tem o maior impacto na digestibilidade da biomassa. Sewalt *et al.*, (1997), além dos fatores indicados mencionam a porosidade (área superficial acessível) e o conteúdo de hemicelulose do material. Segundo Morrison *et al.* (2011) a lignina representa o maior obstáculo para o acesso das enzimas à celulose e hemicelulose.

As celulasas são as enzimas encarregadas de hidrolisar a celulose, que é degradada em monômeros de glicose. O complexo enzimático das celulasas é integrado pelas celobiohidrolases (CBHs) (exo-1,4- β -D-glicanases), as endo-1,4- β -D-glicanases (EGs); e as β -glicosidases (BG), as quais trabalham em conjunto e em forma sinérgica com as CBHs e EGs.

2.4. A INTEGRAÇÃO DE PROCESSOS NA PRODUÇÃO DE BIOETANOL

A intensificação dos processos de produção de etanol combustível se tornou uma prioridade para o projeto de configurações tecnológicas com melhor desempenho técnico, econômico e ambiental. A integração de processos alcança a intensificação através da realização, em simultâneo de várias operações unitárias e processos (separação, mistura, reação química, transformação biológica, transporte de fluidos) em uma mesma unidade, ou através de seu acoplamento, ou de um melhor aproveitamento dos fluxos de energia (CARDONA *et al.*, 2008).

A integração de processos oferece muitas vantagens em comparação com processos não integrados: maior conversão de reagentes e, portanto um melhor uso das matérias primas; maior seletividade que permite a redução na quantidade de produtos não desejados; processos mais compactos com menores necessidades de vapor e de água de resfriamento, portanto com menores necessidades de consumo de energia. Além disso, o tamanho compacto de alguns esquemas integrados permite a redução na quantidade de fluxos de reciclagem externa, que são substituídos por ciclos internos (CARDONA *et al.*, 2008).

Na produção de etanol, a integração de processos pode ser estudada com base nos seguintes itens:

- a) Integração reação-reação, com várias aplicações na tecnologia do etanol de segunda geração, que incluem a co-fermentação de hidrolisados lignocelulósicos, a sacarificação e fermentação simultâneas e sacarificação e co-fermentação simultâneas. Nestos casos duas ou mais transformações químicas ou bioquímicas ocorrem de um modo sinérgico na mesma unidade do processo.
- b) Integração separação-separação, este é outro modo de realizar intensificação de processos, a combinação de duas ou mais operações unitárias, seja simultaneamente, ou de modo conjugado é a principal característica deste tipo de integração. O caso típico é o acoplamento da destilação com outras

operações de separação: com a extração (destilação extrativa), com o uso de membranas (pervaporação) etc., para melhorar desempenhos na separação do produto.

- c) Integração reação-separação. Esse caso envolve a integração fermentação-separação a qual é mais relacionada com a fermentação contínua com separação de produto por flash, processo que tem maior interesse para o desenvolvimento do presente trabalho, pelo que será tratado com mais detalhes.

Cardona *et al.* (2008) afirmam que a integração reação-separação permite a remoção do produto da zona onde a transformação acontece, oferecendo várias oportunidades para aumentar o rendimento do produto e, conseqüentemente, reduzindo os custos do produto, especialmente no caso de reações reversíveis ou processos biocatalíticos. A integração reação-separação não implica somente a realização da reação química e separação da mistura formada em uma única unidade, mas também a realização simultânea de ambos os processos. A realização simultânea permite a superação das limitações inerentes a cada processo.

2.5. INTEGRAÇÃO FERMENTAÇÃO- SEPARAÇÃO

Segundo Silva (1998) uma grande restrição do processo de fermentação alcoólica convencional é o álcool produzido, pois o álcool é um inibidor das células de leveduras e, desta forma, reduz o rendimento e a produtividade, motivo pelo qual vários caminhos combinando fermentação com sistema de separação foram propostos como meios de minimizar a inibição pelo produto no processo de fermentação alcoólica. Para Cardona *et al.* (2008) os esquemas que integram a fermentação com operações de separação são uma alternativa atrativa para a intensificação dos processos de fermentação alcoólica, a remoção do etanol desde o caldo de fermentação diminui e neutraliza seus efeitos inibitórios sobre taxa de crescimento celular. Esquemas deste tipo são:

- a) Fermentação - remoção de etanol por vácuo.
- b) Fermentação - remoção de etanol por gás- stripping.
- c) Fermentação - remoção de etanol por membranas.
- d) Fermentação - remoção de etanol por extração líquida.

2.5.1. FERMENTAÇÃO - REMOÇÃO DE ETANOL POR VÁCUO

O acoplamento do fermentador com uma câmara de vácuo para a extração do etanol desde o caldo de fermentação permite a remoção parcial do produto e o aumento da produtividade global do processo. Segundo Cardona *et al.* (2008), em geral, a pesquisa sobre a fermentação de vácuo tem recebido menos atenção do que outras formas de remoção de etanol. Na literatura alguns estudos realizados neste tipo de integração são os descritos a seguir.

Cysewski e Wilke (1977) propuseram um sistema de reciclo de células e fermentação de vácuo para produção contínua de etanol, utilizando a levedura *Saccharomyces cerevisiae* (ATCC No. 4126). Segundo os autores, à pressão total de 50 mmHg e usando 33,4% de glicose na alimentação, produtividades de 82 e 40 g/L.h foram atingidas com o sistema de vácuo com e sem reciclo de células, respectivamente.

Lee *et al.* (1983) utilizaram cepas de *Zymonomas mobilis* em um sistema projetado com dois estágios com o fermentador a pressão atmosférica e um recipiente externo operando a vácuo (50 mmHg). A operação seqüencial da câmara de vácuo esteve sob controle por microprocessador. Os autores relataram que o sistema permite superar a dificuldade de manter condições estéreis no fermentador a pressões baixas, e que a produtividade atingida foi de 85 g/L.h.

Silva (1998) estudou através de modelagem e simulações um processo de fermentação alcoólica contínua conectada a um sistema de extração de etanol produzido enquanto ele é formado (tanque flash a vácuo) tentando, desta forma, aumentar o rendimento e a produtividade pela minimização da inibição celular pelo produto.

Atala (2004, 2006) desenvolveu estudos em escala de bancada de um processo fermentativo contínuo acoplado a um evaporador que funciona a vácuo do tipo Flash. Segundo o autor o álcool evaporado sai com uma concentração de aproximadamente 50°GL após condensação, o que possibilita a retirada de uma coluna de destilação, bem como a redução dos trocadores de calor no reator, já que o álcool extraído diminui a temperatura de operação da fermentação. Por isso, a sua aplicação nos processos industriais poderia gerar grande economia de energia e de custo.

Existem outras formas citadas na literatura de se remover o etanol continuamente, uma descrição complementar pode ser encontrada no apêndice 1.A.

2.6. A MODELAGEM MATEMÁTICA DO PROCESSO DE FERMENTAÇÃO COM EXTRAÇÃO CONTÍNUA DE ETANOL UTILIZANDO SEPARADOR FLASH

A modelagem, neste caso, inclui os modelos que se referem aos seguintes componentes:

- 1) Processo de fermentação - A modelagem matemática da fermentação é realizada com base em modelos cinéticos acoplados com as equações de balanço de massa para as células (biomassa), substrato e produto e o balanço energético.
- 2) Operações de separação:
 - a) do produto - Se utilizam os modelos para cálculo do equilíbrio líquido-vapor e equações aplicáveis à separação *flash* de fases em equilíbrio em condições de vácuo.
 - b) auxiliares - Incluem a separação de componentes do caldo de fermentação com propósitos específicos, por exemplo: filtração, purga, centrifugação. O modelo global usualmente integra o efeito destas operações nos balanços de massa.

2.6.1. CARACTERIZAÇÃO DOS MODELOS MATEMÁTICOS UTILIZADOS NA FERMENTAÇÃO PARA PRODUÇÃO DE BIOETANOL

2.6.1.1. O TIPO DOS MODELOS CINÉTICOS

Segundo Bayley e Ollis, citados por Andrietta (1994), os modelos cinéticos usados em fermentações podem ser:

- a) Não estruturados e não segregados, nos quais as células de microrganismos são consideradas como soluto;
- b) Estruturados e não segregados, onde as células são tratadas como indivíduos de múltiplos componentes, porém com composição média semelhante;
- c) Não estruturados e segregados, onde as células são tratadas como seres individuais distintos, porém descritos por um único componente; e
- d) Estruturados e segregados, onde as células de microrganismos são consideradas como indivíduos distintos e formados com múltiplos componentes.

A Figura 2.5 mostra as diferenças entre esses tipos de modelos.

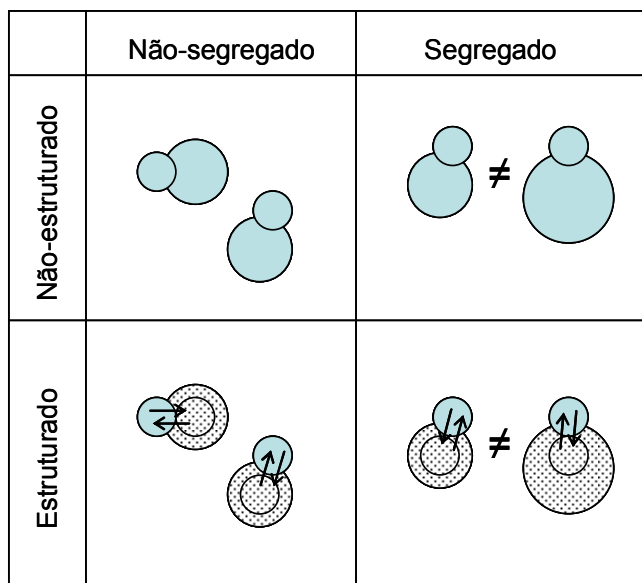


Figura 2.5. Tipos de modelos cinéticos usados em fermentações

Dunn *et al.* (2003) afirma que, os modelos cinéticos não estruturados apresentam as cinéticas celulares desprovidas de qualquer interpretação baseada em mecanismos. Os modelos cinéticos estruturados consideram com algum detalhe os processos e reações das células individuais, a fim de obter uma descrição mais realista. Os mais simples modelos segregados consideram diferentes estágios no desenvolvimento das células e, portanto uma distribuição de estágios de células em uma cultura sem estruturar a composição celular. Os modelos estruturados e segregados apresentam uma descrição mais realística e complexa da biomassa (distribuição de células em diferentes estágios de desenvolvimento) e da estrutura intracelular.

Andrietta (1994) afirma que os modelos cinéticos que são mais comumente utilizados para descrever a fermentação alcoólica são os do tipo não estruturados e não segregados, ou seja, modelos onde as células dos microrganismos são tratadas como soluto e que não proporcionam uma interpretação mecanística das cinéticas celulares.

Andrietta (1994) explica que o processo de produção de etanol por via fermentativa é muito complexo pelo grande número de reações envolvidas na transformação, tendo como agravante o fato de que estas ocorrem no interior de um organismo vivo, o que ilustra o nível de dificuldade para a modelagem deste sistema se as células forem consideradas como indivíduos distintos constituídos de vários componentes. Esta complexidade torna pouco aconselhável o uso de modelos muito elaborados, sendo a escolha mais indicada modelos do tipo não estruturado e não segregado, cujas simplificações reduzem o número de parâmetros que podem ser obtidos com precisão com ensaios relativamente simples. No entanto, relata algumas restrições que devem ser observadas na utilização dos modelos não estruturados, para seu uso no estudo dinâmico do processo:

- 1) Os modelos derivados do de Monod só podem ser aplicados quando o crescimento balanceado é assumido. Sendo o crescimento balanceado

definido como um estado biológico, durante o qual a cadeia de reações metabólicas intracelular está ocorrendo essencialmente em estado estacionário (DUNN e MOR, 1981, citados por Adrietta (1994)).

- 2) A utilização deste tipo de modelos está vinculada com a manutenção do crescimento balanceado. É muito difícil determinar o ponto exato de quando este deixa de existir por perturbações externas e o prejuízo nos resultados se estas condições não forem mantidas.

2.6.1.2. CARACTERÍSTICAS DA MODELAGEM CINÉTICA DO PROCESSO DE FERMENTAÇÃO

Na literatura, os modelos do processo de fermentação para produção de bioetanol, utilizam todos ou alguns dos seguintes modelos cinéticos:

1. Cinética de crescimento da levedura
2. Cinética de morte dos microorganismos
3. Cinética de formação de etanol
4. Cinética de consumo de substrato

2.6.1.2.1. Cinética de crescimento da levedura

Os fatores que interferem na velocidade específica de crescimento da levedura (μ), segundo os modelos cinéticos, são as concentrações de: ART (açúcares redutores totais) (S), etanol (P) e células de levedura (X); e a temperatura (T), portanto:

$$\mu = f(S, P, X, T) \quad (2.1)$$

O efeito da temperatura não é considerado diretamente nos modelos, segundo Adrietta (1992). Tratando-se de uma grandeza física, é aconselhável considerar a sua influência sobre a velocidade específica de crescimento da levedura, através do seu efeito sobre determinados parâmetros cinéticos, então tem-se que:

$$\mu = f(S, P, X) \quad (2.2)$$

Sem a influência destes fatores considerados como interferentes, a velocidade específica de crescimento tem um valor máximo: $\mu = \mu_{\max}$ ($\mu_{\max}$ é a velocidade específica de crescimento máxima), então considerando termos para expresar os efeitos destes fatores:

$$\mu = \mu_{\max} E_S E_P E_X \quad (2.3)$$

Na literatura, os termos que expresam os efeitos diferem segundo cada autor, assim temos:

E_S = Termo de limitação e/ou inibição pelo substrato. Na limitação o termo tende a zero quando acaba o substrato e na inibição o termo tende a zero quando aumenta a concentração de substrato.

E_P = Termo de inibição pelo produto.

E_X = Termo de inibição pela concentração celular de levedura.

Estando os valores de E_S , E_P e E_X entre 0 (máximo efeito) e 1 (sem efeito).

Quando utilizamos:

- a) $E_S = \frac{S}{K_s + S}$, da equação empírica de Monod $\mu = \mu_{\max} \frac{S}{K_s + S}$,
- b) $E_P = (1 - \frac{P}{P_{\max}})^n$, proposta por Levenspiel (1980), e
- c) $E_X = (1 - \frac{X_t}{X_{\max}})^m$, proposta por Lee *et al.* (1983).

A velocidade específica de crescimento da levedura resulta expressa segundo o modelo de Lee *et al.* (1983):

$$\mu = \mu_{\max} \frac{S}{K_s + S} (1 - \frac{P}{P_{\max}})^n (1 - \frac{X_t}{X_{\max}})^m \quad (2.4)$$

Onde:

$\mu_{\max}$: Velocidade (taxa) específica de crescimento máxima do microrganismo.

K_s : Constante de saturação do substrato, ou constante de afinidade do microrganismo com o substrato (ANDRIETTA, 1994).

$P_{\max}$: Concentração máxima de etanol na qual o microrganismo deixa de se desenvolver.

$X_{\max}$: Concentração máxima de células na qual o microrganismo deixa de se desenvolver.

n : potência do termo de inibição pelo produto.

m : potência do termo de inibição pela concentração celular.

Na literatura existe uma variedade de alternativas para uso de termos que expresam os efeitos dos fatores na equação (2.3), algumas delas se mostram na Tabela 2.3.

Tabela 2.3. Termos para expresar os efeitos dos fatores: produto, substrato e biomassa nos modelos que expressam a velocidade específica de crescimento da levedura.

Inibição pelo Produto	Referência	Limitação e/ou inibição pelo Substrato	Referência	Inibição pela Biomassa	Referência
$\mu = \mu_{\max} \left(1 - \frac{P}{P_{\max}}\right)$ $\mu = \mu_{\max} - k_1 P$	Luong (1985) Linear	$\mu = \mu_{\max} \frac{S}{k_s + S + \frac{S^2}{k_t}}$	Modelo de Edwards Citado por Rao (2010)	$\mu = \mu_{\max} \left(1 - \frac{X}{X_{\max}}\right)$	Andrietta (1994)
$\mu = \mu_{\max} e^{-K_2 P}$	Exponencial	$\mu = \mu_{\max} \left(1 - e^{-\frac{S}{K_s}}\right)$	Modelo de Teissier Citado por Trejos <i>et al.</i> (2009)	$\mu = \mu_{\max} e^{-K_1 P} \left(1 - \frac{X}{X_1}\right)$	Chataway <i>et al.</i> (1988)
$\mu = \mu_{\max} \frac{1}{1 + \frac{P}{k_3}}$	Hiperbólica	$\mu = \mu_{\max} \frac{S}{K_s S_0 + S}$ S_0 = concentração inicial de substrato	Monod modificado Dunn <i>et al.</i> (2003)	$\mu = \mu_{\max} \frac{S}{K_s X + S}$	Modelo de Contois. Citado por: Dunn <i>et al.</i> (2003)
$\mu = \mu_{\max} \left(1 - \frac{P}{P_{\max}}\right)^{\frac{1}{2}}$	Parabólica	$\mu = \mu_{\max} \frac{S^n}{K_m + S^n}$	Modelo de Moser Citado por Trejos <i>et al.</i> (2009)	$\mu = \mu_{\max} \left(1 - \frac{X_t}{X_{\max}}\right)^m$	Lee <i>et al.</i> (1983)
$\mu = \mu_{\max} \left(1 - \frac{P}{P_{\max}}\right)^n$	Não linear generalizada	$\mu = \mu_{\max} \frac{S}{K_s + S + K_1 S^2}$	Modelo de Haldane Citado por Trejos <i>et al.</i> (2009)	$\mu_{\max} = A * X^{-B}$	Krishnan <i>et al.</i> (1999)

Na Tabela 2.3, o significado dos parâmetros $\mu_{\max}$, K_s , $P_{\max}$, $X_{\max}$, n , m é o mesmo que para o modelo de Lee *et al.* (1983).

Ertola *et al.*, (2008) salientam que a inibição pode ser competitiva ou não competitiva, neste caso o inibidor é uma outra substância diferente do substrato, de modo que se existe inibidor a afinidade do microorganismo pelo substrato é diminuída. Os modelos descritos são:

1) Inibição competitiva

$$\mu = \mu_{\max} \frac{S}{K_s \left(1 + \frac{I}{K_I} \right) + S} \quad (2.5)$$

2) Inibição não competitiva

$$\mu = \frac{\mu_{\max}}{\left(1 + \frac{I}{K_I} \right)} \frac{S}{K_s + S} \quad (2.6)$$

Onde:

I = concentração do inibidor.

K_I = Constante cujo valor está inversamente relacionado com a afinidade do microorganismo por el inibidor.

Na inibição não competitiva o termo de inibição $\left(1 + \frac{I}{K_I} \right)$ não influencia na constante K_s , mas sim afeta a taxa específica máxima de crescimento $\mu_{\max}$.

Outros autores propõem modelos que incluem termos de limitação e/ou inibição pelo substrato (S), inibição pelo produto (P) e alguns também inibição pela biomassa celular (X):

$$\mu = \mu_{\max} \left(\frac{S}{K_s + S} \right) (1 - K_p P) \quad \text{Hinshelwood, citado por Trejos *et al.* (2009)}$$

$$\mu = \mu_{\max} \frac{S}{k_s + S + \frac{S^2}{k_I}} \left(1 - \frac{P}{P_{\max}} \right) \quad \text{Ghose e Tyagi (1979)}$$

$$\mu = \mu_{\max} \frac{S}{k_s + S + \frac{S^2}{k_I}} \left(1 - \frac{P}{P_{\max}}\right)^\beta \quad \text{Krishnan et al. (1999)}$$

$$\mu = \mu_{\max} \frac{S}{K_s + S} e^{-K_r S} \left(1 - \frac{P}{P_{\max}}\right)^n \quad \text{Magazoni et al. (2010)}$$

$$\mu = \mu_m \frac{S}{K_s + S} \left(1 - \left(\frac{P}{P_{\max}}\right)^{A1}\right) \left(1 - \left(\frac{X}{X_{\max}}\right)^{A2}\right) \quad \text{Jarzebki et al. (1989)}$$

$$\mu = \mu_{\max} \frac{S}{K_s + S + \frac{S^2}{K_I}} \left(1 - \frac{P}{P_{\max}}\right)^n \left(1 - \frac{X}{X_{\max}}\right)^m \quad \text{Ferreira (2005)}$$

Žeravic and Stevanovic (2007) propõem o uso de termos para inibição por substrato e produto, e de um parâmetro de manutenção para o crescimento celular m , que inclui no modelo, o efeito de mecanismos não associados ao crescimento celular na biossíntese de etanol a partir da glicose.

$$\mu = \mu_{\max} \frac{S}{K_M + S + \frac{S^2}{K_{IS}} + \frac{S \cdot P}{K_{IP}}} - m \quad (2.7)$$

Onde:

K_{IS} é a constante de inibição pelo substrato (S).

K_{IP} é a constante de inibição pelo produto (P).

K_M é a constante de Monod da velocidade de crescimento celular.

Todas as equações anteriores se aplicam em casos em que existe um só substrato, para dois substratos algumas alternativas são listadas a seguir.

Pode ser utilizado o Modelo cinético de Monod de substrato múltiplo (DUNN *et al.*, 2003) para descrever a influência de muitos substratos, para dois substratos toma a forma:

$$\mu = \mu_{\max} \left(\frac{S_1}{K_1 + S_1} \right) \left(\frac{S_2}{K_2 + S_2} \right) \quad (2.8)$$

Para dois substratos o efeito pode também ser escrito como reações paralelas, de acordo com:

$$\mu = \mu_{\max} \left(\frac{k_1 S_1}{k_1 + S_1} + \frac{k_2 S_2}{k_2 + S_2} \right) \left(\frac{1}{k_1 + k_2} \right) \quad (2.9)$$

Krishnan et al.(1999) para uma mistura de substratos Glicose (G) e Xilose (Xe) propôs utilizar um modelo que incorpora inibições de substrato e produto de acordo com:

$$\mu_G = \mu_{mG} \frac{G}{K_S + G + \frac{G^2}{K_{IG}}} \left\{ 1 - \left(\frac{P}{P_{\max}} \right)^\beta \right\} \quad (2.10)$$

$$\mu_{Xe} = \mu_{mXe} \frac{Xe}{K_S + Xe + \frac{Xe^2}{K_{IXe}}} \left\{ 1 - \left(\frac{P}{P_{\max}} \right)^\beta \right\} \quad (2.11)$$

Onde:

μ_{mG} e μ_{mXe} representam a taxa específica de crescimento máximo em glicose e xilose, respectivamente.

K_{IG} e K_{IXe} representam as constantes de inibição por glicose e xilose, respectivamente.

K_S é a constante da equação de Monod.

A proposta, inclui o efeito do tamanho de inóculo (X [g/L]) no parâmetro μ_m , utilizando duas equações com parâmetros determinados experimentalmente:

Fermentação de Glicose: $\mu_{mG} = 0,52 * X^{-0,461}$

Fermentação de Xilose: $\mu_{mXe} = 0,75 * X^{-0,438}$

O modelo proposto para crescimento celular e fermentação em mistura de açúcares é:

$$\mu_{mistura} = \frac{G}{G + Xe} \mu_G + \frac{Xe}{G + Xe} \mu_{Xe} \quad (2.12)$$

μ_G e μ_{Xe} representam a taxa específica de crescimento em glicose e xilose, respectivamente.

Tosetto (2002) propôs utilizar para dois substratos no meio de alimentação, Glicose (G) e Frutuose (F), as seguintes equações:

$$\mu = \mu_{\max} \frac{G}{K_{SG} + G + \frac{G^2}{K_{IF}}} \left(1 - \frac{P}{P_{\max G}} \right)^n \quad (2.13)$$

$$\mu = \mu_{\max} \left(1 - \frac{G}{F} \right) \frac{F}{K_{SF} + F + \frac{F^2}{K_{IF}}} \left(1 - \frac{P}{P_{\max F}} \right)^n \quad (2.14)$$

Onde: $\left(1 - \frac{G}{F} \right)$ é o termo que inclui no modelo o efeito da diauxia, que é a característica que um microorganismo tem de consumir preferencialmente um substrato (Glicose).

A Tabela 2.4 mostra os modelos utilizados para fermentação contínua e os termos de limitação e/ou inibição incluídos neles.

Tabela 2.4. Modelos utilizados para fermentação contínua e com separação do produto por evaporação flash.

Referência	Modelo	Processo
Adrieta (1994)	$\mu = \mu_{\max} \frac{S}{K_s + S} \left(1 - \frac{P}{P_{\max}}\right)^n \left(1 - \frac{X_t}{X_{\max}}\right)^m$ (sugerido por Lee <i>et al.</i>, 1983)	Fermentação contínua
Siqueira (1997)	$\mu = \mu_{\max}(T) \frac{S}{K_s(T) + S} \left(1 - \frac{P}{P_{\max}(T)}\right)^n \left(1 - \frac{X_t}{X_{\max}(T)}\right)^m$ (sugerido por Lee <i>et al.</i>, 1983)	Fermentação contínua
Silva (1998)	$\mu = \mu_{\max} \frac{S}{K_s + S} \left(1 - \frac{P}{P_{\max}}\right)^n$ (sugerido por Levenspiel, 1980)	Fermentação contínua extrativa com flash
Atala (2000, 2004, 2006)	$\mu = \mu_{\max} \frac{S}{K_s + S} e^{-KS} \left(1 - \frac{P}{P_{\max}}\right)^n \left(1 - \frac{X_t}{X_{\max}}\right)^m$ Contribuições: Inibição de substrato: Monod modificado Inibição de produto: Levespiel, 1980. Inibição de biomassa: Lee <i>et al.</i>, 1983.	Fermentação contínua extrativa com flash
Atarassi (2001) Duarte (2007) Cohen (2011)	$\mu = \mu_{\max} \frac{S}{K_s + S} e^{-KS} \left(1 - \frac{P}{P_{\max}}\right)^n \left(1 - \frac{X_t}{X_{\max}}\right)^m$ Modelo utilizado: Atala (2000)	Fermentação contínua extrativa com flash

2.6.1.2.2. Cinética de morte dos microorganismos

Para a fermentação alcoólica contínua extrativa com evaporador flash, Atala (2006) utiliza a equação:

$$r_d = (K_{dT} \exp(K_{dP}P))X \quad (2.15)$$

Onde:

r_d é a velocidade cinética de morte, K_{dT} e K_{dP} são constantes de morte celular devido a temperatura e concentração de etanol, respectivamente, X é a concentração de células viáveis, e P é a concentração do produto.

2.6.1.2.3. Cinética de produção de etanol

A formação de produto r_p como uma função da concentração de biomassa X , em sua forma mais simples pode ser expressa por:

$$r_p = k.X \quad (2.16)$$

No entanto, a cinética da formação do produto pode ser extremamente complexa, já que o produto às vezes é formado durante o crescimento das células e, às vezes depois, que todo o crescimento parou (DUN *et al.*, 2003).

As equações para a formação do produto utilizadas nos modelos cinéticos da fermentação com separador flash incluem a formação associada ao crescimento e a formação associada à manutenção, sendo a equação mais utilizada:

$$r_p = Y_{P/X} r_x + m_x X \quad (2.17)$$

Onde: a taxa de formação do produto r_p , depende linearmente da concentração de biomassa X ; $Y_{P/X}$ é o rendimento de produto baseado em células, r_x é a taxa cinética de crescimento celular, e m_x é um coeficiente de manutenção.

O seguinte procedimento mostra que a equação (2.17) é derivada do modelo de Luedeking-Piret (1959) que expressa a geração metabólica do produto pela equação:

$$\frac{dP}{dt} = \alpha \frac{dX}{dt} + \beta X \quad (2.18)$$

A equação (2.19) permite relacionar a taxa de consumo de substrato com a concentração celular X e a taxa específica de crescimento celular μ (GÓDIA CASABLANCAS e LÓPEZ SANTIN, 1998):

$$-\frac{dS}{dt} = \frac{1}{Y_{P/S}} \frac{dP}{dt} = \frac{1}{Y_{X/S}} \mu X + mX \quad (2.19)$$

A taxa de formação de produto portanto é:

$$\frac{dP}{dt} = \frac{Y_{P/S}}{Y_{X/S}} \mu X + mX \quad (2.20)$$

Comparando as equações (2.18) e (2.20)

$$\alpha = \frac{Y_{P/S}}{Y_{X/S}}, \quad \frac{dX}{dt} = \mu X, \text{ e } \beta = m$$

Segundo Pauletti *et al.* (2004)

$$Y_{P/S} = \frac{-\frac{dP}{dt}}{\frac{dS}{dt}}, \quad Y_{X/S} = \frac{-\frac{dX}{dt}}{\frac{dS}{dt}}$$

Então, o coeficiente α pode ser expresso como:

$$\alpha = \frac{-\frac{dP}{dt}}{\frac{dS}{dt}} = \frac{\frac{dP}{dt}}{\frac{dX}{dt}} = Y_{P/X}$$

Substituindo na equação (2.20) tem-se:

$$\frac{dP}{dt} = Y_{P/X} \frac{dX}{dt} + mX \quad (2.21)$$

Como:

$$r_p = \frac{dP}{dt}, \text{ e } r_x = \frac{dX}{dt}$$

Obtemos a equação derivada de Luedeking-Piret utilizada pelos modelos da fermentação extrativa contínua com evaporador flash (ATALA, 2006):

$$r_p = Y_{P/X} r_x + mX$$

A equação tem termo para formação de produto associada ao crescimento e para formação associada à manutenção.

2.6.1.2.4. Cinética de consumo de substrato

Para o processo de fermentação extrativa contínua com evaporador flash, segundo Silva (1998), a taxa de consumo de substrato é representada pela equação:

$$r_s = -\mu \frac{X}{Y_{X/S}} \quad (2.22)$$

Onde:

μ é a velocidade específica de crescimento do microrganismo.

$Y_{X/S}$ é o rendimento teórico em biomassa.

X é a concentração de biomassa.

Atala (2006) utiliza a seguinte equação:

$$r_s = \frac{r_x}{Y_x} + m_x X \quad (2.23)$$

Onde:

r_s é a taxa de consumo de substrato, r_x é a taxa cinética de crescimento celular.

X é a concentração de biomassa, m_x é o parâmetro de manutenção, e

Y_x é o rendimento celular limite.

2.6.1.2.5. Efeito da temperatura sobre os parâmetros cinéticos

Os parâmetros cinéticos são componentes empíricos muito importantes que dão precisão às predições dos modelos, sem a determinação experimental cuidadosa e precisa desses parâmetros os modelos não teriam utilidade. Portanto, estão muito relacionados às matérias-primas utilizadas, condições da experimentação, dados experimentais do processo e aos métodos matemáticos utilizados para sua determinação. Andrietta (1997) afirma que, por serem características do microrganismo e do meio de cultivo utilizado nos ensaios de obtenção, os valores dos parâmetros cinéticos encontrados na literatura variam bastante de um trabalho para outro. Um número de 16 parâmetros cinéticos com 50% deles dependentes da temperatura é usual nos modelos cinéticos da produção de etanol por via fermentativa.

Alguns parâmetros cinéticos são apresentados na Tabela 2.5. Os valores dos parâmetros cinéticos foram obtidos em experimentos de produção de álcool por via fermentativa com *Saccharomices cerevisiae*, as condições experimentais estão detalhadas em Alves (1996), Atala (2000) e Andrade (2007). Devido a grande diversidade de valores dos parâmetros cinéticos na literatura, Silva (1998) sugere que para um uso adequado dos valores dos parâmetros é de grande importância o conhecimento do procedimento utilizado nos levantamentos destes

parâmetros. Os resultados de Andrade (2007) salientam que otimizar os parâmetros dos modelos cinéticos para seu uso num intervalo de temperaturas não é suficiente, se ocorrem mudanças na composição do meio de alimentação ou nas condições operacionais, os parâmetros deverão ser reestimados.

Tabela 2.5. Parâmetros cinéticos e sua dependência com a temperatura

Parâmetro	Expressão (Temperatura em °C)	Referência
$\mu_{\max}$	$4,181 \times 10^4 \exp(-3535/T) - 5,397 \times 10^{69} \exp(-50\,475/T)$ 28 °C < T < 38 °C	Alves (1996)
$P_{\max}$	638,1 exp (-0,05741T) 30 °C < T < 38 °C	
$\mu_{\max}$	$1,57 \exp(-41,4714/T) - 12\,948,91 \exp(-431,39/T)$ 28 °C < T < 40 °C	Atala (2000)
$P_{\max}$	$-0,44206T^2 + 26,41032T - 279,74825$ 28 °C < T < 40 °C	
$\mu_{\max}$	A exp (B/T) + C exp (D/T), onde: A= $-2,98 \times 10^5$ B=-302,44 C= $2,82 \times 10^5$ D=-302,18 30 °C < T < 38 °C	Andrade (2007)
$P_{\max}$	A exp (B/T) +C exp (D/T), onde: A= $6,01 \times 10^{-22}$ B=1588,05 C= 106,33 D=-12,02 30 °C < T < 38 °C	

Uma descrição dos balanços de massa na fermentação contínua com evaporador flash para a produção de etanol pode ser encontrada no apêndice 1.B.

2.6.1.3. CARACTERÍSTICAS DOS MODELOS UTILIZADOS PARA A SEPARAÇÃO FLASH

O tanque flash é um equipamento onde se promove um contato íntimo entre as fases líquida e vapor, com o objetivo de separar componentes com base na diferença de volatilidade de um componente em relação ao outro. Assim, o líquido se enriquece com o componente menos volátil e o vapor com o componente mais volátil. O tanque flash é uma coluna de destilação com apenas um estágio.

2.6.1.3.1. Modelos para cálculo do equilíbrio L-V.

Segundo Smith & Van Ness (1980) e Sandler (1989), para que o equilíbrio se estabeleça em uma mistura multicomponente com duas fases, às mesmas

condições de temperatura e pressão (P), a fugacidade de cada componente deve ser igual em todas as fases:

$$f_i^L = f_i^V \quad (i=1,2,\dots,n) \quad (2.38)$$

As fugacidades em cada fase são expressas pelas equações:

$$f_i^L = x_i \cdot \gamma_i \cdot f_i^0 \quad (\text{fase líquida}) \quad (2.39)$$

$$f_i^V = y_i \cdot \Phi_i \cdot P \quad (\text{fase vapor}) \quad (2.40)$$

Onde:

γ_i é o coeficiente de atividade do componente i na fase líquida.

f_i^0 é a fugacidade do componente i puro.

Φ_i é o coeficiente de fugacidade do componente i na fase vapor.

P é a pressão do sistema.

Substituindo na equação tem-se:

$$y_i \cdot \Phi_i \cdot P = x_i \cdot \gamma_i \cdot f_i^0 \quad (2.41)$$

A constante de equilíbrio k_i para um componente i é definida por:

$$k_i = \frac{y_i}{x_i} = \frac{\gamma_i f_i^0}{P \Phi_i} \quad (2.42)$$

Para as condições de operação do evaporador flash (pressões menores a 1 atm), pode-se considerar que a fase vapor se comporta como uma fase ideal.

Desta forma, não é necessário o uso do componente que descreve o desvio da idealidade (Φ_i), e f_i^0 pode ser substituído por P_i^{sat} , que é a pressão de vapor saturado do componente i . Assim, a equação (2.42), torna-se:

$$k_i = \frac{y_i}{x_i} = \frac{\gamma_i P_i^{sat}}{P} \quad (2.43)$$

O valor de P_i^{sat} pode ser conhecido através da equação de Antoine, que relaciona a pressão de vapor com a temperatura e as constantes de cada espécie química (A, B, C), ajustadas com base nos dados experimentais.

O maior problema para o cálculo do equilíbrio de fases de uma mistura multicomponente é encontrar os valores dos coeficientes de atividade γ_i . A literatura apresenta estudos que mostram modelos e métodos que calculam os coeficientes de atividade, sendo eles:

- Método ASOG (Analytical Solutions of Groups)
- Modelo UNIQUAC (Universal Quase Chemical)
- Método UNIFAC (Uniquac Functional Group Activity Coefficients)
- Modelo NRTL (Nonrandom Two – Liquid).

Os modelos mais utilizados para fermentação extrativa com evaporador Flash são o modelo NRTL e o método UNIFAC. Um detalhamento desses modelos pode ser encontrado no apêndice 1.C.

2.6.1.4. CONDIÇÕES DE OPERAÇÃO DO FLASH COM BASE EM MODELOS DE EQUILÍBRIO

A Tabela 2.6 resume o resultado do uso de modelos de equilíbrio em estudos de simulação do processo fermentativo contínuo com evaporador flash para retirada de produto do caldo de fermentação e se mostram também para comparação resultados dos estudos experimentais feitos por Atala (2004 e 2006).

Tabela 2.6. Condições de operação do tanque *Flash*

Referência	Método utilizado	Modelo de equilíbrio	Condições de operação do Flash			
			Pressão Atm	(mmHg)	Temperatura K	(°C)
Silva (1998)	Simulação	NRTL (binário etanol-água)	0,045	(34,2)	302,15	(29)
Atarassi (2001)	Simulação	UNIFAC (*)	0,047	(35,5)	303,15	(30)
Costa (2001)	Simulação	NRTL (binário etanol-água)	0,045	(34,2)	303,15	(30)
Atala (2004)	Experimental	-----	0,158	(120)	306,15	(33)
Atala (2006)	Experimental	-----	0,158	(120)	306,15	(33)
Cohen (2011)	Simulação	NRTL-HOC(**)	0,039	(29,6)	302,15	(29)

(*) modificado por Peres & Macedo (1997). (**) NRTL-HOC (Non-Random Two-Liquid - Hayden-O'Connell).

Na Tabela 2.6, segundo os resultados de simulação, as condições de operação do tanque *Flash* são: pressão no intervalo 29,6 mmHg – 35,5 mmHg e temperatura no intervalo 29 °C – 30 °C. No entanto, os resultados experimentais (ATALA, 2004 e 2006) são bem diferentes: pressão entre 90 mmHg e 150 mmHg (meia de 120 mmHg) e temperatura de 33 °C.

Cohen (2011) fez uma análise das simulações com o modelo NRTL e contribui para uma melhor ideia da discrepância entre os resultados teóricos de simulação e os experimentais, quando afirma, que para a faixa de pressões entre 90 mmHg e 150 mmHg a temperatura de operação correspondente é superior a 47 °C, temperatura muito diferente dos 33 °C obtidos experimentalmente.

Nos estudos de simulação foram utilizadas misturas binárias etanol-água para simular o comportamento do caldo de fermentação na separação *Flash*, excepto no caso de Atarassi (2001) que utilizou uma mistura multicomponente, essa informação fornece um elemento importante para explicar a diferença entre resultados teóricos e experimentais.

Atala (2006) explica que o desvio das condições de operação do sistema deve-se a diferença entre a solução modelada (etanol-água) e a realmente usada (caldo de fermentação com resíduos de açúcar, glicerol, biomassa, gás carbônico, sais, ácidos, metabólitos secundários e outros). Os resultados obtidos permitiram ao autor concluir que, no tanque *Flash*, o modelo NRTL obtido com dados de equilíbrio da mistura binária etanol-água não representa bem o processo fermentativo extrativo.

Atarassi (2001) utilizou o método UNIFAC para o cálculo do *Flash* e considerou uma solução complexa composta de sacarose, glicose, frutose, glicerol, gás carbônico e etanol. Na comparação com dados experimentais, houve uma diferença relativamente grande, aponta que a diferença entre o valor simulado e o valor experimental provavelmente é devido a outros compostos

presentes no meio fermentado que não foram analisados. Concluiu que o CO_2 afeta o equilíbrio do sistema e, que talvez o modelo proposto deva incluir mais componentes para representar melhor o processo.

Em consequência os procedimentos teóricos utilizados até agora não permitem representar bem o processo real no tanque *Flash*.

Cohen (2011) salienta a necessidade de fazer em trabalhos futuros, estudos experimentais do comportamento termodinâmico dos componentes que intervêm no processo de produção de etanol, pois atualmente muitos destes parâmetros não estão disponíveis na literatura, dificultando a caracterização termodinâmica e a seleção do modelo termodinâmico.

2.6.2. OBSERVAÇÕES REFERIDAS À MODELAGEM MATEMÁTICA DA PRODUÇÃO DE ETANOL POR VIA FERMENTATIVA

- 1) O problema de modelagem da produção de etanol por via fermentativa é muito complexo e é uma área de pesquisa ainda em evolução. O desenvolvimento das pesquisas leva à elaboração de modelos mais estruturados que integrem múltiplos substratos, inibição competitiva, efeito de fatores como dioxia, e outros componentes do caldo de fermentação: metabólitos, subprodutos, enzimas.
- 2) Os parâmetros cinéticos determinados para determinadas condições experimentais não são válidos para outras condições. Mudanças na cepa do micro-organismo, na safra do melaço usado como substrato, na composição do meio fermentativo, entre outras, modificam os valores dos parâmetros, que devem ser reestimados (ANDRADE, 2007).
- 3) Os resultados experimentais do equilíbrio líquido-vapor a baixas pressões (<400 mmHg) relatados por Atarasi (2001) mostram que:
 - a) Uma mistura preparada com sacarose, glicose, frutose, glicerol e etanol dissolvidos em meio de fermentação, ferve a uma temperatura menor do que a estimada pelo modelo para uma mistura equivalente sem meio de

fermentação, para a mesma pressão de equilíbrio (diferença de até 5 °C). Daí, decorre que outras substâncias contidas no meio fermentado, além de sacarose, glicose, frutose, glicerol e etanol modificam o equilíbrio líquido-vapor.

b) O efeito de CO₂ dissolvido sobre o equilíbrio é notório e também incrementa a pressão de equilíbrio até em 65 mmHg, em comparação ao caso binário etanol-água .

Estes resultados contribuem para explicar porque os resultados experimentais (ATALA, 2004, 2006), são diferentes dos obtidos por simulação para as condições de operação do Flash e sugerem que é preciso incluir outros fatores, como a condição de polaridade da maioria das substâncias encontradas no caldo de fermentação (salientada por ÁLVARES, 2008) e outros componentes nos modelos de equilíbrio, como: álcoois diversos, ácidos orgânicos, além de células de levedura que permanecem no meio; para uma melhor modelagem do equilíbrio líquido-vapor do caldo de fermentação na unidade de separação Flash.

Diante desta evidência alguns pesquisadores manifestam que existe carência de dados termodinâmicos importantes que precisam ser obtidos pelo trabalho experimental.

2.7. CONCLUSÃO

Neste capítulo, foram apresentados aspectos importantes da produção de bioetanol, relacionados com as tecnologias de primeira e segunda geração, e alternativas de integração ou acoplamento de processos para melhorar a produtividade e diminuir os custos de produção, com ênfase no processo de fermentação extrativa contínua com separador flash, que é aplicado no desenvolvimento deste trabalho.

A revisão bibliográfica mostra as complexidades do desenvolvimento de modelos para o processo fermentativo extrativo, com modelos não estruturados e

não segregados baseados nas cinéticas de: crescimento da biomassa, consumo de substrato e formação de produto, para o processo fermentativo; e com uso de modelos para cálculo do equilíbrio de fases, para a unidade de separação flash. Observando-se que, as maiores dificuldades se apresentam na reestimação dos parâmetros dos modelos por mudanças na temperatura ou em outras condições operacionais, e na previsão das condições operacionais reais no tanque flash.

2.8. REFERÊNCIAS BIBLIOGRÁFICAS

1. Ahmad, F., Jameel, A.T., Kamarudin, M.H. and Mel, M. (2011). Study of growth kinetic and modeling of ethanol production by *Saccharomyces cerevisiae*. *African Journal of Biotechnology*, v. 16(81), pp. 18842-18846.
2. Álvarez, V.H., Rivera, E.C., Costa, A.C., Maciel Filho, R., Maciel, M.R.W. and Aznar, M. (2008). Bioethanol Production Optimization: A Thermodynamic Analysis. *Applied Biochemistry and Biotechnology*, v. 148, pp. 141-149.
3. Alves, J.G.L.F. **Estudo da influência da temperatura na cinética de crescimento anaeróbico de *Saccharomyces cerevisiae***. Dissertação de Mestrado. Faculdade de Engenharia de Alimentos, Universidade Estadual de Campinas, Brasil, 1996.
4. Andrade, R.R.. **Procedimento para o Desenvolvimento de um Modelo Matemático Robusto para o Processo de Fermentação Alcoólica**. Dissertação de Mestrado. Faculdade de Engenharia Química, Universidade Estadual de Campinas. Brasil, 2007.
5. Andrietta, S.R. **Modelagem, simulação e controle de fermentação alcoólica contínua em escala industrial**. Tese de Doutorado. Faculdade de Engenharia de Alimentos, Universidade Estadual de Campinas. Brasil, 1994.
6. Atala, D. I.; Costa. A. C.; Maciel. and R; Mauger, F. (2001). Kinetic of ethanol fermentation with high biomass concentration considering the effect of temperature. *Applied Biochemistry and Biotechnology*, v. 91-93, pp. 353-365.
7. Atala, D.I.P. **Fermentação Alcoólica com Alta Densidade Celular: Modelagem Cinética, Convalidação de Parâmetros e Otimização do Processo**. Dissertação de Mestrado, Faculdade de Engenharia de Alimentos, Universidade Estadual de Campinas, Brasil, 2000.

8. Atala, D.I.P. **Montagem, Instrumentação, Controle e Desenvolvimento Experimental de um Processo Fermentativo extrativo de Produção de Etanol**. Tese de Doutorado. Faculdade de Engenharia de Alimentos, Universidade Estadual de Campinas, 2004.
9. Atala, D.I.P., & Maciel, R. (2006). **Experimental development, optimization and control of a Continuous extractive production process for the ethanol production**. Postdoctoral research. Process 151335/2005-2, FAPESP, Brazil.
10. Atarassi, M.M. **Modelagem, simulação e controle de um evaporador flash acoplado a uma unidade de fermentação alcoólica contínua**. Dissertação de Mestrado. Faculdade de Engenharia de Alimentos, Universidade Estadual de Campinas. Brasil, 2005.
11. Banco Nacional de Desenvolvimento Econômico e Social (BNDES). **Bioetanol de Caña de Azúcar Energia para el desarrollo sostenible**. BNDES-CGEE. Rio de Janeiro, Brasil. 2008. Disponível em: <http://www.bioetanoldecanaadeazucar.org/>. Acesso em: 15 setembro 2009.
12. Banerjee, R.; Pandey, A. (2002), Bio-industrial applications of sugarcane bagasse: a technological perspective. *International Sugar Journal*, v. 104, pp. 64-67.
13. Basso, T.O., **Melhoramento da Fermentação Alcoólica em *Saccharomyces cerevisiae* por Engenharia Evolutiva**. Tese de Doutorado, Universidade de São Paulo, São Paulo, Brasil, 2011.
14. Bastos, V.D., **Etanol, Alcoolquímica e Biorrefinarias**. BNDES Setorial, Rio de Janeiro, n. 25, pp. 5-38, 2007.
15. Biofuels Platform. ENERS Energy Concept. **Production of biofuels in the world in 2009**. Disponível em: <http://www.biofuels-platform.ch/en/infos/production.php?id=bio-ethanol>. Acesso em: 10 janeiro 2012.
16. Bon, E.P.S. (2007). **Ethanol production via enzymatic hydrolysis of sugarcane bagasse and straw**. *Food and Agricultural Organisation Symposium: The role of agricultural biotechnologies for production of bio-energy in energy in developing countries*. Inst. Química – Universidade Federal Rio de Janeiro. Disponível em: <http://www.fao.org/biotech/presentation/bon.pdf>. Acesso em: 15 outubro 2010.

17. Bouville, M.. (2008). **Fermentation kinetics including product and substrate inhibitions plus biomass death: a mathematical analysis.** Disponível em: <<http://arxiv.org/pdf/q-bio/0509014.pdf>>. Acesso em: 15 janeiro 2012.
18. Brasil, Ministério do Desenvolvimento, Indústria e Comércio Exterior (2012). **Álcool combustível.** Disponível em: <<http://www.mdic.gov.br/sitio/interna/interna.php?area=2&menu=999>>. Acesso em: 10 de janeiro de 2012.
19. Brigham, J. S., Adney, W. S., and Himmel, M. E. (1996), in Handbook on Bioethanol: Production and Utilization, Wyman, C., ed., Taylor and Francis, Washington, DC, pp. 119–141.
20. Byung-Hwan, U. (2007). **Optimization of Ethanol Production from Concentrated Substrate.** PhD Thesis. Auburn University, USA.
21. Cardona, C.A., L.F. Gutiérrez, and Ó.J. Sánchez. 2008. **Process integration: Base for energy saving.** In: D.M. Bergmann, Editor, *Energy efficiency research advances*, Nova Science Publishers, Hauppauge, NY, USA, pp. 173-212.
22. Carvalho, J.C.M. e Sato, S. **Engenharia Bioquímica.** In: Borzani, W.; Lima, U.A.; Aquaroni, E., 1975., v. 3, Cap. 10: Fermentação descontínua alimentada, pp. 205-211.
23. Chang, V. S., Burr, B., Holtzapple, M. T. (1997). Lime pretreatment of switchgrass. *Applied Biochemistry and Biotechnology*, v. 63-65, pp. 3-19.
24. Chang, V. S., Nagwani, M., Holtzapple, M. T. (1998). Lime pretreatment of crop residues bagasse and wheat straw. *Applied Biochemistry and Biotechnology*, v. 74, pp. 135–159.
25. Chang, V.S., Holtzapple, M. T. (2000). Fundamental factors affecting biomass enzymatic reactivity. *Applied Biochemistry and Biotechnology*, v. 86, pp. 5-37.
26. Chattaway, T., Goma,G., and Renaud, P.Y. (1988). Modelling ethanol and secondary inhibitions of ethanol fermentation in a multistage reactor. *Biotechnology and Bioengineering*, v. 32 (3), pp. 271-276.
27. Cho, C.W., Hwang, S.T. (1991). Continuous membrane fermentor separator for ethanol fermentation. *Journal of Membrane Science*, v. 57 (1), pp. 21-42.

28. Chovau, S., Gaykawad, S., Straathof, A.J.J., Van der Bruggen, B.(2011). Influence of fermentation by-products on the purification of ethanol from water using pervaporation. *Bioresource Technology*, v. 102. pp. 1669–1674.
29. Christen, P, Minier, M, and Renon, H.(1990). Ethanol extraction by supported liquid membrane during fermentation. *Biotechnology and Bioengineering*, v. 36, pp. 116-123.
30. Chum ,H. L, Johnson D. K, Black, S., Baker, J., Grohmann, K., Sarkanen, K. V. (1988). Organosolv pretreatment for enzymatic hydrolysis of poplars: 1. Enzymatic hydrolysis of cellulosic residues. *Biotechnology and Bioengineering*, v. 31, pp. 643-649.
31. Cohen, L.M. **Modelagem e simulação do processo de fermentação extrativa a vácuo com uma câmara de flash e separação do CO2 utilizando uma coluna de absorção**. Dissertação de Mestrado. Faculdade de Engenharia Química, Universidade Estadual de Campinas, Brasil, 2011.
32. Costa, A. C.; Atala, D. I. P.; Maciel filho, R.; Maugeri filho, F. (2001). Factorial design and simulation for the optimization and determination of control structure for an extractive alcoholic fermentation. *Process Biochemistry*, v. 37, pp. 125-137.
37. Cysewski, G.R., Wilke, C.R. (1977). Rapid ethanol fermentations using vacuum and cell recycle. *Biotechnology and Bioengineering*, v. 19 (8), pp. 1125–1143.
38. D'Amore, T. (1992).Improving Yeast Fermentation Performance. *Journal of the Institute of Brewing*, v. 98, pp. 375-382.
39. Dale B.E., Moreira M.J. (1982). Freeze-explosion technique for increasing cellulose hydrolysis. *Biotechnology and Bioengineering Symposium*, v.12, pp. 31- 43.
40. Dale, M.C., Lei, S. and Zhou, C. A **Laboratory and Pilot Plant scaled Continuous Stirred Reactor Separator for the Production of Ethanol from Sugars, Corn Grits/starch or Biomass streams**. In: Proceedings, Second Biomass Conference of the Americas: Energy, Environment, Agriculture, and Industry, 1995, Portland, OR. Disponível em: <<http://www.osti.gov/bridge/servlets/purl/109653-63T4GV/webviewable/109653.pdf>>. Acesso em: 20 de junho 2010.
41. Dale, M.C., Moelhman, M. and Zhou, Ch. Design of a Pilot Scale Continuous Stirred Ethanol Reactor Separator with Solvent Absortion and Extractive Distillation. CoFE 1995. Disponível em: <<http://bio-process.com/wp-content/uploads/2010/01/Pilot-Design.pdf> >. Acesso em: 5 de julho 2011.

42. Daugulis, A.J., Axford, D.B., Ciszec, B., and Malinowski, J.J. (1994). Continuous fermentation of high-strength glucose feeds to ethanol. *Biotechnology letters.*, v. 16 (6), pp. 637- 642.
43. Dias, M.O.S, et. al. (2011). Improving bioethanol production from sugarcane: evaluation of distillation, thermal integration and cogeneration systems, *Energy*, v. 36 (6), pp. 3691-3703.
44. Dias, M. O.S.. **Simulação do processo de produção de etanol a partir do açúcar e do bagaço, visando a integração do processo e a maximização da produção de energia e excedentes do bagaço.** Dissertação de Mestrado. Faculdade de Engenharia Química. Universidade Estadual de Campinas. Brasil, 2008.
45. Duarte, E.R. **Avaliação e desenvolvimento de algoritmos de controle aplicado a um processo extrativo de fermentação alcoólica contínua.** Tese Doutorado. Faculdade de Engenharia Química, Universidade Estadual de Campinas, 2007.
46. Dunn I.J., Heinzle E, Ingham J, Prenosil J.E. (2003). **Biological Reaction Engineering Dynamic Modeling Fundamentals with Simulation Examples.** Second, Completely Revised Edition. Weinheim: WILEY-VCH GmbH & Co.KGaA.
47. Ertola, R, Yantorno, O. and Mignone, C. (2008). **Microbiologia Industrial.** Organization of American States. Washington, D.C., USA.
Disponível em:
<<http://portal.oas.org/LinkClick.aspx?fileticket=whdxh3wv%2fWc%3d&tabid=581>>. Acesso em: enero 2012.
48. Fan L.T, Lee Y. H, Gharpuray, M.M. (1982). The nature of lignocellulosics and their pretreatments for enzymatic hydrolysis. *Advances in Biochemical Engineering*, v. 23, pp.157-187.
49. Ferreira, E. **Contribuição para o Estudo da Otimização da Fermentação Alcoólica operando em Batelada-alimentada.** Tese de Mestrado. Faculdade de Engenharia Química, Universidade Estadual de Campinas, Brasil, 2005.
50. Fredenslund A., Jones R.L., Prausnitz J.M.. (1975). Group-contribution estimation of activity coefficients in nonideal liquid mixtures; *AIChE Journal*, v. 21(6), pp. 1086-1099.
51. Fuentes L.L.G. **Determinação de dados cinéticos da deslignificação do bagaço de cana-de-açúcar e da hidrólise enzimática no pré-tratamento**

com hidróxido de cálcio. Dissertação de Mestrado. Faculdade de Engenharia Química, Universidade Estadual de Campinas, Brasil, 2009.

52. Galassi, G.R. **Estudo do processo fermentativo contínuo para produção de etanol utilizando células auto-imobilizadas em reatores tipo torres.** Dissertação de Mestrado. Faculdade de Engenharia Química, Universidade Estadual de Campinas, 2007.
53. Ghose, T.K. and Tyagi, IR.D. (1979) Rapid ethanol fermentation of cellulose hydrolisate. II. Product and substrate inhibition and optimization of fermentor design. *Biotechnology and Bioengineering*, v. 21 (8), pp. 1401-1420.
54. Godia-Casablanca, F., López-Santín, J.. **Cinética microbiana**, En: Godia-Casablanca, F., López-Santín, J.(Eds.) *Ingeniería Bioquímica*. Editorial Síntesis, Madrid, 1998.
55. Gollapalli, L. E., Dale, B. E., Rivers, D. M. (2002). Predicting digestibility of ammonia fiber explosion (AFEX) treated rice straw. *Applied Biochemistry and Biotechnology*, v. 98–100, pp. 23–35.
56. Gomes, E., Guez, M.A.U., Martin, N., Silva, R. (2007). Thermostable enzymes: sources, production and industrial applications. *Química Nova*, v. 30(1). pp. 136-145.
57. Groot, W. J., Kraayenbrink, M. R. Waldram, R. H. Lans. R. G. J. M. and Luyben. K. Ch. A. M. (1992). Ethanol production in an integrated process of fermentation and ethanol recovery by pervaporation. *Bioprocess and Biosystems Engineering*, v. 8 (3-4), pp. 99-111.
58. Guo, Z.P., Zhang, L., Ding, Z.Y. and Shi, G. Y.(2011). Minimization of glycerol synthesis in industrial ethanol yeast without influencing its fermentation performance. *Metabolic Engineering*, v. 13 (1), pp. 49–59.
59. Hendler, B. **Construção de um software de simulação e ajuste de curvas cinéticas para o processo de fermentação em batelada alimentada.** Dissertação de Mestrado. Faculdade de Engenharia Química, Universidade Estadual de Campinas, Brasil. 2011.
60. Henniges, O. and Zeddies, J., Economics of Bioethanol in the Asia-Pacific-Australia. In: F.O. Licht's. *World Ethanol and Biofuels*. Report, v. 3 (11), 2005.
61. Hsu, T. A. **Pretreatment of Biomass**. In: Wyman, C.E. (Ed.), *Handbook on Bioethanol, Production and Utilization*. Washington, DC.: Taylor & Francis, 1996.

62. International Energy Agency. Biofuel Production (IEA). **Energy Technology Essentials**. Reporte, Janeiro, 2007. Disponível em: <<http://www.iea.org/techno/essentials2.pdf>>. Acesso em: 15 janeiro 2012.
63. Jacobsen, S.E., and Wyman, C.E. (2000). Cellulose and Hemicellulose Hydrolysis Models for Application to Current and Novel Pretreatment Processes. *Applied Biochemistry and Biotechnology*, v. 84–86, pp.81-96.
64. Kaar W.E, Holtzapple M.T. (2000). Using lime pretreatment to facilitate the enzymatic hydrolysis of corn stover. *Biomass Bioenergy*. Vol .18, pp.189–199.
65. Kaseno, Miyazawa, I., Kokugan, T. (1998). Effect of product removal by a pervaporation on ethanol fermentation. *Journal of Fermentation and Bioengineering*, v. 86 (5), pp. 488-493.
66. Kim, S., Holtzapple, M. T. (2005). Lime pretreatment and enzymatic hydrolysis of corn stover. *Bioresource Technology*, v. 96 (18), pp.1994-2006.
67. Kim, S., Holtzapple, M. T. (2006). Delignification kinetics of corn stover in lime pretreatment. *Biosource Technology*, v. 97, pp. 778-785.
68. Kirk, T. K, Farrell R. L. (1987). Enzymatic combustion: The microbial degradation of lignin. *Annual Review of Microbiology*, v. 41, pp. 465-505.
69. Krishnan, M.S.,Ho, N.W.Y. and Tsao, G. (1999). Fermentation Kinetics of Ethanol Production from Glucose and Xylose by Recombinant *Saccharomyces* 1400 (pLNH33). *Applied Biochemistry and Biotechnology*, v.78 (1-3), pp. 373-388.
70. Larsen, B. L., Ramussen, P., Fredenslund, A. (1987). A modified group contribution model for prediction of phase equilibria and heats of mixing. *Industrial & Engineering Chemistry Research*, v. 26. pp. 2274- 2286.
71. Lee, J. (1997). Biological Conversion of Lignocellulosic Biomass to Ethanol. *Journal of Biotechnology*, v. 56(1), pp.1-24.
72. Lee, J.H, e Rogers, P. L. (1983).The fermentation kinetics of ethanol production by *Zymomonas mobilis*.*The Chemical Engineering Journal*, v. 27(2). pp. B31-B38.
73. Lessmann, W.F. **Estudo do processo descontínuo alimentado para a produção de amiloglicosidase por *Apergillus awarnori* NRRL3112 para concentrações de polissacarídeo de 40 e 80 g/L**. Dissertação de Mestrado, Escola Politécnica, Universidade de São Paulo, Brasil, 1993.

74. Levenspiel, O. (1980) Monod equation: a revised and a generalization to product inhibition situations. *Biotechnology and Bioengineering*, v. 22 (8), pp. 1671-1687.
75. Liu, C.Z., Wang, F., Ou-Yang, F. (2009). Ethanol fermentation in a magnetically fluidized bed reactor with immobilized *Saccharomyces cerevisiae* in magnetic particles. *Bioresource Technology*, v. 100 (2), pp. 878-882.
76. Liu, H.S., Hsien-Wen, H. (1990). Analysis of gas stripping during ethanol fermentation—I. In a continuous stirred tank reactor. *Chemical Engineering Science*, v. 45 (5), pp. 1289-1299.
77. Luedeking, R., Piret, E.L. (1959). A kinetic study of the lactic acid fermentation. Batch process at controlled pH. *Journal of Biochemical and Microbiological Technology and Engineering*, v. 1(4), pp. 393 - 412.
78. Luong, J.H.T. (1985) Kinetics of ethanol inhibition in alcohol fermentation. *Biotechnology and Bioengineering*, v. 27 (3), pp. 280-285.
79. Magazoni, F.C., Monteiro, J.B., Cardenil, J.M., Colle, S. (2010). Cooling of ethanol fermentation process using absorption chillers. *International Journal of Thermodynamics*, v. 13 (3), pp. 111-118.
80. Maiorella, B., Blanch, H.W. and Wilke, C.R. (1983). By-Product Inhibition Effects on Ethanolic Fermentation by *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering*, v. 25, pp. 103-121.
81. Maiorella, B., Wilke, Ch.R and Blanch, H.W. (1981). **Alcohol production and recovery**. In: Fletcher, A. (Ed.), *Advances in Biochemical Engineering*. Heidelberg: Springer Verlag, v. 20, pp. (43-92), 1981.
82. Maiorella, B.L., Blanch, H. W., Wilke, C. R. (1984). Economic evaluation of alternative ethanol fermentation processes. *Biotechnology and Bioengineering*, v. 26 (9), pp. 1003-1025.
83. Matsumura, M., Märkl, H. (1986). Elimination of ethanol inhibition by perstraction. *Biotechnology and Bioengineering*, v. 28 (4), pp. 534-541.
84. Meirelles, A.J.A. **Expansão da Produção de Bioetanol e Melhoria Tecnológica da Destilação Alcoólica**. In: Workshop “Produção de Etanol”, EEL/USP, 2006. Disponível em:
<http://www.apta.sp.gov.br/cana/anexos/PPaper_sessao_4_Antonio_Meirelles.pdf>.
Acesso em: 20 dezembro 2010.
85. Minier, M. e Goma, G. (1982). Ethanol production by extractive fermentation. *Biotechnology and Bioengineering*, v. 24 (7), pp. 1565 - 1579.

86. Morrison, D., Dyk, J.S., Pletschke, B.I. (2011). Effect of chemicals on XynA. *BioResources*, v. 6(3), pp. 3132-3141.
87. Mosier, N., Wyman, C. E., Dale, B., Elander, R., Lee, Y.Y., Holtzapple, M. T., Ladisch, M. (2005). Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource Technology*, v. 96 (6), pp. 673–686.
88. Mulder, M.V.H., Henrikman, J.O., Hegeman, H. and Smolders, C.A. (1983). Ethanol-water separation by pervaporation. *Journal of Membrane Science*, v. 16, pp. 269-284.
89. Neto, P. O. **Efeito de fatores inibidores na fermentação alcoólica**. In: FAPESP. Workshop do Projeto Diretrizes de Políticas Públicas para a Agroindústria Canavieira do Estado de São Paulo: Qualidade da Matéria Prima. Lorena, 2008. Disponível em:
<http://www.apta.sp.gov.br/cana/anexos/position_paper_painel3_pedro_oliva.pdf>.
Acesso em: 10 novembro 2010.
90. O'Brien, D.J. and Craig, Jr., (1996). Ethanol Production in a Continuous Fermentation/Membrane Pervaporation System. *Applied Microbiology and Biotechnology*, v. 44, pp. 669-704.
91. OECD/FAO (2011). OECD-FAO Agricultural Outlook 2011-2020, OECD publishing and FAO. Disponível em:
<http://www.oecd.org/pages/0,3417,en_36774715_36775671_1_1_1_1_1,00.html>
Acessed em: 5 janeiro 2012.
92. Paturau, J.M. (1989). **By- products of the Cane Sugar Industry**. 3rd Ed., Amsterdam, The Netherlands: Elsevier.
93. Pauletti, M.S., Santa Cruz, L., Mazza, G., Rozycki, S., Sabbag, N., Costa, S. (2004). Fabricación de Yogur con bacterias inmovilizadas. *Ciencia y tecnologia de los Alimentos*, v. 4 (3), pp. 190-196.
94. Peng, P., Shi, B. and Lan Y. (2010). A review of membrane materials for ethanol recovery by Pervaporation. *Separation Science and Technology*, v. 46 (2), pp. 234-246.
95. Peres, A. M., Macedo, E. A.(1997). Modified UNIFAC model for the calculation of thermodynamic properties of aqueous and non-aqueous solutions containing sugars. *Fluid Phase Equilibria*, v. 139. pp. 47-74.
96. Pham, C.B., Motoki, M., Matsumura, M., Kataoka, H. (1989). Simultaneous ethanol fermentation and stripping process coupled with rectification. *Journal of Fermentation and Bioengineering*, v. 68 (1), pp. 25-31.

97. Playne, M. J. (1984). Increased digestibility of bagasse by pretreatment with alkalis and steam explosion. *Biotechnology and Bioengineering*, v. 26, pp. 426-433.
98. Porto, L. M. **Modelagem de Processo Industrial de Fermentação Alcoólica Contínua com Reatores de Mistura Ligados em Série**. Tese de Doutorado. Faculdade de Engenharia Química, Universidade Estadual de Campinas, Brasil, 2005.
99. Praunitz, J.; Anderson, T.; Grens, E.; Eckert C.; Hsieh, R.; O'Connell, J. **Computer calculations for multicomponent vapor-liquid and liquid-liquid equilibria**. Prentice Hall Inc. New Jersey. 1980.
100. Prausnitz, J. M.; Lichtenthaler, R. M.; Azevedo, E. G. **Molecular thermodynamics of fluid phase equilibria**. 2a ed. Englewood Cliffs: Prentice-Hall, 1986.
101. Raal, J.D., Muhlbauer, A.L. **Phase equilibria: Measurement and Computation**. Washington, D.C.:Taylor & Francis, 1998.
102. Rabelo S, Maciel, R., Costa A. A. (2008). Comparison between lime and alkaline hydrogen peroxide pretreatments of sugarcane bagasse for ethanol production. *Applied Biochemistry and Biotechnology*, v. 148, pp. 45-48.
103. Rabelo, S.C. **Avaliação e otimização de pré-tratamentos e hidrólise enzimática do bagaço de cana-de-açúcar para a produção de etanol de segunda geração**. Tese de Doutorado. Faculdade de Engenharia Química, Universidade Estadual de Campinas, 2010.
104. Ramos, L. P., Breuil, C., Kushner, D. J., Saddler, J. N. (1992). Steam pretreatment conditions for effective enzymatic hydrolysis and recovery yields of Eucalyptus viminalis wood chips. *Holzforschung*, v. 46, pp.149-154.
105. Salcedo, F.O.M. **Aplicación de la Tecnología de Fluidos Supercríticos a la Purificación de Carbohidratos Prebióticos**. Tesis Doctoral, Universidade Autonoma de Madrid. España, 2009.
106. Rao, D.G. **Introduction to Biochemical Engineering**. Second Edition. New Delhi, India: Tata Mc Graw Hill, 2010.
107. Renon, H., Prausnitz, J. M.(1968) Local compositions in thermodynamic excess functions for liquid mixtures. *AIChE J.*, v. 14(1), pp. 135-144.
108. Saliba, E.O.S., Rodriguez, N.M., Morais, S.A.L., Piló-Veloso, D. (2001). Lignins - isolation methods and chemical characterization. *Ciência Rural*, v. 31 (5), pp. 917-928.

109. Sandler, I. S. **Chemical and engineering thermodynamics**. 2 ed. New York: John Wiley & Sons, 1989.
110. Scanes, K.T., Hohmann, S. and Prior, B.A.(1998). Glycerol Production by the Yeast *Saccharomyces cerevisiae* and its Relevance to Wine: A Review. *South African Journal for Enology and Viticulture*, v. 19 (1), pp. 17- 24.
111. Schmidell, W. and Facciotti, M. C. R.. **Biorreatores e Processos Fermentativos**. In: Schmidell, Willibaldo *et al.* (Coord.). *Biotecnologia Industrial: Engenharia Bioquímica*. São Paulo: Edgar Blücher, v. 2, pp. 179-192. 2001.
112. Segel, I. H. **Enzyme kinetics: behavior and analysis of rapid equilibrium and steady-state enzyme systems**. New York: Wiley-Interscience, 1975.
113. Sewalt, V.J.H., Glasser, W., Beauchemin, K.A. (1997). Lignin impact on fiber degradation III: reversal of inhibition of enzymatic hydrolysis by chemical modification of lignin and by additives. *Journal of Agricultural and Food Chemistry*, v. 45(5), pp. 1823-1828.
114. Silva, F.L.H. **Modelagem, simulação e controle de fermentação alcoólica contínua extrativa**. Campinas. Tese Doutorado. Faculdade de Engenharia de Alimentos, Universidade Estadual de Campinas, Brasil, 1998.
115. Siqueira, E. P. F. **Efeito da temperatura na cinética da fermentação alcoólica com alta densidade celular**. Dissertação de Mestrado, Faculdade de Engenharia de Alimentos, Universidade Estadual de Campinas, Brasil, 1997.
116. Smith, J. M. e Van Ness, H. C. **Introdução à termodinâmica da engenharia química**. Tradução de Horacio Macedo. 3 ed. Rio de Janeiro: Guanabara Koogan S. A., 1980.
117. Szengyel, Z., Zacchi, G., Varga, A., & Réczey, K. (2000). Cellulase production of *Trichoderma reesei* Rut C30 Using steam-pretreated spruce. Hydrolytic potential of cellulases on different substrates. *Applied Biochemistry and Biotechnology*, v. 84-86, pp. 679-691.
118. Tang, Y.Q., An, M. Z., Zhong, Y. L., Shigeru, M. Wu, X.L., Kida, K. (2010). Continuous ethanol fermentation from non-sulfuric acid-washed molasses using traditional stirred tank reactors and the flocculating yeast strain KF-7. *Journal of Bioscience and Bioengineering*, v. 109 (1), pp. 41-46.
119. Teymouri, F., Perez, L., Alizadeh H, Dale, B.E. (2005). Optimization of the ammonia fiber explosion (AFEX) treatment parameters for enzymatic

- hydrolysis of corn stover. *Bioresource technology*, v. 96 (18), pp. 2014-2018.
120. Tosetto, G.M. **Influência da matéria-prima no comportamento cinético de levedura na produção de etanol**. Dissertação de Mestrado. Faculdade de Engenharia Química, Universidade Estadual de Campinas, Brasil, 2002.
 121. Trejos, V.M., Alzate, J.F. and Garcia, M.A.G (2009). Descripción matemática y análisis de estabilidad de procesos fermentativos. *Dyna*. Vol 76 (158), pp.111-121.
 122. Valholme, R., Demedts, B., Morreel, K., Ralph, J., Boerjan, W.. (2010). Lignin biosynthesis and structure. *Plant Physiology*, v. 153 (3), pp. 895-905.
 123. Vidal P.F and Molinier J. (1988). Ozonolysis of lignin-improvement of *in vitro* digestibility of poplar sawdust. *Biomass*, v.16, pp.1-17.
 124. Vidal, J.. **Thermodynamics, applications in chemical engineering and the petroleum industry**. Paris, Francia: TECHNIP, 1997.
 125. Viegas, M.C. **Otimização de sistema de fermentação alcoólica contínua utilizando reactores tipo torre e leveduras com características floculantes**. Tese Doutorado. Faculdade de Engenharia Química, Universidade Estadual de Campinas, Brasil, 2003.
 126. Vieira, M.C.A.(2007). **Setor Sucroalcooleiro Brasileiro: Evolução e Perspectivas**. DEAGRO. Brasil. Disponível em:
<http://www.bndes.gov.br/SiteBNDES/export/sites/default/bndes_pt/Galerias/Arquivos/conhecimento/liv_perspectivas/07.pdf>. Acesso em: 12 dezembro 2010.
 127. Von Sivers, M. and Zacchi, G. (1995). A techno-economical comparison of three processes for the production of ethanol from pine. *Bioresource Technology*, v. 51, pp. 43–52.
 128. Walas, S. M. **Phase equilibria in chemical engineering**. Boston: Butterworth publishers, 1985.
 129. Walford, S.N. (2008). Sugar cane bagasse: How easy is it to measure its constituents?. *Proceedings of The South African Sugar Technologists' Association.*, v. 81, pp. 266 - 273.
 130. Watanabe, M. (2009). **Ethanol Production in Brazil: Bridging its Economic and. Environmental Aspects**. International Association for Energy Economics, p.45-48. Disponível em:
<<http://www.iaee.org/en/publications/newsletterdl.aspx?id=89>>. Acesso em: 15 dezembro 2011.

131. Worldwatch Institute. (2006). **Biofuels for Transportation. Global Potential and Implications for Sustainable Agriculture and Energy in the 21st.Century.** Report. Washington, D.C. Disponível em: <<http://www.worldwatch.org/system/files/EBF038.pdf>>. Acesso em: 15 julho 2011.
132. Wu, Y., Huang, W., Xiao, Z., and Zhong Y. (2004). Ethanol recovery from fermentation broth by pervaporation using a composite polydimethylsiloxane membrane. *Chinese Journal of Chemical Engineering*, v. 12 (4), pp. 586 – 589.
133. Zarpellon, F, & Andrietta, S.R. (1992). Fermentação contínua para produção de álcool. *Stab. Açúcar, Álcool e Subprodutos*, v. 10(4), pp. 23-28.
134. Žeravic, S. and Stevanovic, J.S. (2007). The Kinetic Models of the Bioprocess with free and Immobilized Cells. *CI&CEQ*, v. 13 (14), pp. 216-226.

CAPÍTULO 3. REDES NEURAIS ARTIFICIAIS

3.1. INTRODUÇÃO

As redes neurais artificiais são consideradas como uma das técnicas de inteligência artificial e foram desenvolvidas para emular os neurônios biológicos humanos, com a sua aplicação pode se generalizar comportamentos a partir de informação fornecida na forma de exemplos. As redes neurais artificiais têm demonstrado um alto desempenho na modelagem processos não lineares, e também para controle de processos, sendo a sua aplicação extensa em muitas áreas das engenharias de processos: craqueamento catalítico, processos de polimerização, petroquímica, desenvolvimento de novos materiais, detecção de falhas, processos siderúrgicos, processos de tratamento de efluentes industriais, etc.

A ampla difusão de aplicações foi devida às seguintes características atrativas (ASSIDJO *et al.*, 2006):

- As redes neurais artificiais têm habilidade para aproximar funções não lineares arbitrárias.
- Elas podem ser treinadas facilmente usando dados registrados do sistema em estudo.
- Elas são facilmente aplicáveis para sistemas multivariáveis.
- Elas não requerem especificação da relação estrutural entre dados de entrada e saída.

Na área de bioprocessos, as redes neurais foram aplicadas para a modelagem empírica (BARRERA-CORTES e BARUCH, 2000, PRAMANIK, 2004) e híbrida (MANTOVANELI, 2005, RIVERA *et al.* 2006, ASSIDJO *et al.*, 2009) de processos de fermentação com *Sacharomyces cerevisiae* em batelada ou contínuos para produção de etanol. Nos casos de modelagem híbrida, as redes neurais foram usadas para calcular parâmetros cinéticos: taxa de crescimento de biomassa μ_x , taxa de consumo de substrato μ_s , taxa de formação de produto μ_p , os valores são utilizados nas equações diferenciais que descrevem os

comportamentos dinâmicos da biomassa, substrato e produto nos modelos cinéticos clássicos.

Para a modelagem e controle de processos de fermentação alcoólica contínua com unidade de separação flash tem-se realizado relativamente poucas pesquisas com aplicação de redes de neurais, sendo que as pesquisas realizadas utilizaram simulação computacional a partir de dados fornecidos pelos modelos cinéticos clássicos da fermentação e dados de equilíbrio binário etanol-água para simular a separação do etanol a partir do caldo de fermentação no evaporador flash. Podem ser citadas as contribuições de Costa *et al.* (2001), Costa *et al.* (2002), Meleiro *et al.* (2005) e Duarte (2004 e 2007). Nesses estudos foi utilizado o modelo matemático desenvolvido por Costa *et al.* (2001) para o processo fermentativo proposto por Silva (1998), os parâmetros cinéticos foram obtidos experimentalmente por Atala (2000). Harada (2001) utilizou redes neurais para desenvolver um modelo híbrido neuronal do processo fermentativo com o modelo cinético clássico e parâmetros propostos por Silva (1998). O único trabalho em que se faz modelagem a partir dos dados experimentais para o processo de fermentação alcoólica contínua com unidade de separação flash é o de Rivera *et al.* (2011) que propôs softsensors para monitoramento das concentrações de etanol nesse processo, baseados em modelos empíricos de redes neurais.

A seguir é apresentada uma revisão dos fundamentos teóricos das redes neurais para a sua utilização na modelagem empírica e controle de processos no presente trabalho.

3.2. CONCEITO E CARACTERÍSTICAS

Segundo Haykin (2007) uma rede neural artificial é um processador massivamente paralelo constituído por unidades de processamento simples, que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para o uso. Ela se assemelha ao cérebro em dois aspectos.

- 1) O conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem.
- 2) Forças de conexão entre unidades de processamento de informação (neurônios) conhecidas como pesos sinápticos são utilizadas para armazenar o conhecimento adquirido.

O conceito salienta como características importantes da rede, sua estrutura, com unidades de processamento simples e interconectadas, e suas capacidades de: processamento paralelo da informação, aquisição de conhecimento por aprendizagem, armazenamento e disponibilização para uso futuro. Segundo Hinton (2001), a propriedade mais interessante das redes neurais é sua habilidade de aprender com exemplos através da adaptação dos pesos das conexões, uma explicação complementar dos processos e paradigmas de aprendizagem é apresentada no apêndice 2.A.

Além das capacidades mencionadas no conceito, a rede tem capacidade de generalizar a informação aprendida, entendendo por generalização o fato de que a rede neural pode produzir saídas adequadas para entradas que não estavam presentes durante o treinamento (aprendizagem).

Na literatura são reconhecidos os benefícios da utilização das redes neurais como resultado das capacidades acima mencionadas. Gomes e Ludemir (2008) manifestam que as redes neurais são ferramentas úteis para o reconhecimento de padrões ou classificação de dados mesmo na presença de ruídos ou dados incompletos.

Haykin (2007) menciona como propriedades relevantes que oferece o uso das redes neurais:

1. Não linearidade. Propriedade muito importante, particularmente se o mecanismo físico que gera o sinal de entrada é inerentemente não linear.

2. Mapeamento de entrada-saída. A aprendizagem supervisionada envolve a modificação dos pesos sinápticos de uma rede neural pela aplicação de um conjunto de exemplos da tarefa. Cada exemplo consiste de um sinal de entrada único e de uma resposta desejada correspondente. Assim, a rede aprende dos exemplos a construir um mapeamento de entrada-saída para o problema considerado.
3. Adaptabilidade. A rede pode ser facilmente retreinada para lidar com pequenas modificações nas condições operativas do meio ambiente. Além disso, quando é requerida a sua operação em um ambiente não estacionário, uma rede neural pode ser projetada para modificar seus pesos sinápticos em tempo real.
4. Resposta a evidências. No contexto de classificação de padrões, a rede pode ser utilizada para rejeitar padrões ambíguos, caso eles estejam presentes, e com isso melhorar o desempenho de classificação da rede.
5. Informação contextual. Cada neurônio da rede é afetado pela atividade de todos os outros neurônios na rede, conseqüentemente, a informação contextual é tratada naturalmente pela rede neural.
6. Tolerância a falhas. Uma rede neural, implementada na forma física (em hardware), tem o potencial de ser inerentemente tolerante a falhas, ou capaz de realizar computação robusta, no sentido de que seu desempenho se degrada suavemente sob condições de operação adversas.
7. Implementação em VLSI (*very-large-scale-integration*).

A natureza maciçamente paralela de uma rede neural a faz ser potencialmente rápida na computação de certas tarefas. Esta mesma característica torna uma rede neural adequada para implementação utilizando tecnologia de integração em escala muito ampla.
8. Uniformidade de análise e projeto.

Basicamente, as redes neurais desfrutam de universalidade como processadores de informação. A mesma notação é utilizada em todos os domínios envolvendo a aplicação de redes neurais.

9. Analogia neurobiológica. O cérebro é uma prova viva de que o processamento paralelo tolerante a falhas não é somente fisicamente possível, mas também rápido e poderoso.

Enquanto ao funcionamento das redes neurais, cada neurônio em uma rede processa os sinais que recebe e envia os sinais processados para outros neurônios. Em uma estrutura usual de rede neural artificial os neurônios estão interconectados e distribuídos em camadas. O objetivo de uma rede neural é procesar a informação de acordo com seu prévio treinamento com dados de entrada-saída. No procesamento de informação as redes neurais são extremadamente paralelas, suas numerosas operações são executadas simultaneamente (SIMÕES e SHAW, 2007).

Os modelos de redes neurais artificiais podem ser definidos tendo como referência as seguintes características: O conjunto de unidades computacionais, as funções de ativação das unidades computacionais, a topologia da rede, o processo de aprendizado da rede e a interação com o ambiente externo que fornece informação para a rede e/ou interage com ela.

As unidades computacionais ou neurônios têm como componentes básicos: um conjunto de sinapses, um somatório e uma função de ativação (HAYKIN, 2007).

- 1) **As sinapses ou elos de conexão** são as conexões entre neurônios, cada uma caracterizada por um peso sináptico. O peso sináptico w_{ji} caracteriza a conexão entre a saída do neurônio i e a entrada do neurônio j .
- 2) **O somador** ou combinador linear calcula a soma ponderada de cada sinal de entrada x_i multiplicada pelo seu peso sináptico w_{ji} .

O modelo de neurônio inclui também um parâmetro de ajuste para o resultado da combinação linear que é chamado de bias, o bias b_j é considerado equivalente a um peso sináptico com sinal de entrada igual a 1.

3) **A função de ativação** que recebe o resultado do somador e restringe a amplitude da saída de um neurônio a um intervalo permissível. Tipicamente, o intervalo normalizado da amplitude da saída de um neurônio é $[0,1]$ ou alternativamente $[-1,1]$.

A Figura 3.1 descreve um neurônio artificial, na mesma, $x_1, x_2, \dots, x_n$, são os sinais de entrada, $w_{j1}, w_{j2}, \dots, w_{jn}$ são os pesos sinápticos do neurônio j , s_j é a saída do combinador linear devido aos sinais de entrada; b_j é o bias; $f(s)$ é a função de ativação; y_j é o sinal de saída do neurônio j .

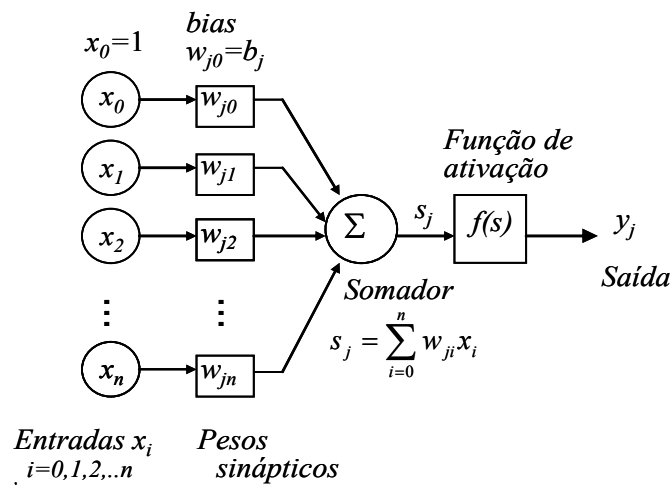


Figura 3.1. Modelo de um neurônio

3.2.1. TIPOS DE FUNÇÕES DE ATIVAÇÃO

A função de ativação transforma o sinal de entrada em um sinal de saída. A escolha da função de ativação é considerada, por muitos especialistas, tão importante quanto a arquitetura ou o algoritmo de aprendizagem da rede neural (GOMES e LUDERMIR, 2008). As funções de ativação mais utilizadas na construção de redes neurais artificiais são: função sigmoide, função linear, função de limiar, e a função gaussiana.

3.2.1.1. FUNÇÃO SIGMOIDE

A função sigmoide tem gráfico com forma de S, é definida como uma função estritamente crescente, contínua e diferenciável em todos os pontos, tem aproximação assintótica a seus valores de saturação e mostra um equilíbrio adequado entre comportamento linear e não linear (HAYKIN, 2007; MENNON *et al.* 2000). São alguns exemplos a função sigmoide “logística” e a tangente hiperbólica. Segundo Mennon *et al.* (1996), em poucos casos as curvas sigmóides podem ser descritas por fórmulas.

3.2.1.1.1. Função sigmoide logística.

É a função de ativação mais popular utilizada em redes neurais (MENNON *et al.* 2000). Pode ser definida como:

$$f(s) = \frac{1}{1 + \exp(-as)} \quad (3.1)$$

Onde: s é a entrada da rede para o neurônio, $f(s)$ é a saída do neurônio e a é o parâmetro de inclinação da função sigmoide. A saída do neurônio com função de ativação logística varia entre 0 e +1.

3.2.1.1.2. Função tangente hiperbólica

Pode ser definida como:

$$f(s) = \tanh(s) = \frac{1 - e^{-as}}{1 + e^{-as}} \quad (3.2)$$

Onde: $f(s)$ é a saída do neurônio que varia entre -1 e +1.

As funções de transferência sigmóides são usadas tradicionalmente em redes *feedforward* com aprendizagem *backpropagation*. Segundo Haykin (2007) se a função de ativação sigmoide incorporada no modelo do neurônio da rede é antissimétrica (i.e ex. tangente hiperbólica), a rede pode, em geral, aprender mais rápido (em termos do número de iterações necessárias para treinamento) do que quando ela é não-simétrica (i.e. função logística padrão).

3.2.1.2. FUNÇÃO LINEAR

É uma equação linear da forma:

$$f(s) = as \quad (3.3)$$

Onde a é um número real que define a saída linear para os valores de entrada.

A função linear também pode ser definida por segmentos dentro de um intervalo, nesse caso se conhecem como funções lineares definidas por partes (*piecewise linear functions*) e consistem de um número finito de segmentos lineares. As funções de limiar e de rampa são casos especiais dessas funções. As funções lineares definidas por partes são mais fáceis de calcular do que as funções não lineares, tais como as funções sigmóides, e têm sido usadas como aproximações das mesmas (MENNON *et al.*, 2000).

Um caso de função linear definida por partes é:

$$f(s) = \begin{cases} 1 & \text{se } s \geq +\frac{1}{2} \\ as & \text{se } +\frac{1}{2} < s < \frac{1}{2} \\ -1 & \text{se } s \leq -\frac{1}{2} \end{cases} \quad (3.4)$$

Onde $f(s)$ tem três segmentos de linha reta, o segmento intermediário tem coeficiente angular a .

3.2.1.3. FUNÇÃO DE LIMIAR

Para uma saída do neurônio $f(s)$ entre 0 e +1, a função de limiar é expressa como:

$$f(s) = \begin{cases} 1 & \text{se } s \geq 0 \\ 0 & \text{se } s < 0 \end{cases} \quad (3.5)$$

Neste caso, a saída de um neurônio assume o valor 1, se a entrada é não negativa, e 0, em caso contrário.

Quando se deseja que a saída do neurônio $f(s)$ esteja entre -1 e +1, a função de limiar pode ser expressa como:

$$f(s) = \begin{cases} 1 & \text{se } s > 0 \\ 0 & \text{se } s = 0 \\ -1 & \text{se } s < 0 \end{cases} \quad (3.6)$$

O modelo de neurônio de McCulloch e Pitts (1943) é uma unidade lógica tipo limiar (Threshold Logic Unit, TLU) com entradas binárias (0 ou 1) e uma saída binária, os pesos associados a cada entrada são +1 ou -1.

3.2.1.4. FUNÇÃO GAUSSIANA

A saída do neurônio é expressa como:

$$f(s) = e^{\left[\frac{-(x-C_i)^2}{\tau_i} \right]} \quad (3.7)$$

Onde C_i é um vetor que determina um centro associado e τ_i é um escalar que representa o “espalhamento” da função de ativação da unidade. Dessa maneira, a saída da unidade é máxima quando $x=C_i$ e decai suavemente na medida em que x se afasta do centro C_i .

A função de ativação gaussiana é utilizada em redes RBF (redes de funções de base radial).

Na literatura existem estudos que mostram aplicações específicas de outras funções de ativação para o aprendizado da rede neural, como descrito, a seguir.

Hornik (1991) mostraram que funções de ativação contínuas, limitadas e inconstantes permitem que as redes *feedforward* multicamadas com apenas uma única camada escondida sejam aproximadores universais de funções.

Leshno *et al.* (1993) mostraram utilidades das funções de ativação contínuas definidas por partes (piecewise linear functions) no desempenho das redes feedforward multicamadas na aproximação de funções.

Guarnieri *et al.* (1999) and Vecchi *et al.* (1998) relataram que em redes *feedforward* multicamadas, com o uso de uma função de ativação *spline* adaptativa baseada na *spline* cúbica de Catmull-Rom como função de ativação dos neurônios, é possível ter melhorias em termos de capacidade de generalização e de velocidade do aprendizado em ambas as tarefas de reconhecimento de padrões e processamento de dados.

Zhang (2008) realizou estudos para desenvolver modelos de redes neurais para simulação e previsão, na área de pesquisa de avaliação da mudança climática global. Ele usou funções $\cos(x)$ e $\sin(x)/x$, como funções de ativação dos neurônios, com bons resultados em uma rede multicamada com uma camada de entrada com unidades de entrada, duas camadas ocultas com unidades de processamento intermediárias, e uma camada de saída com unidades de saída.

Gomes e Ludermitz (2008a e 2008b) utilizaram uma rede *Perceptron* multicamada para aproximar funções, com as inversas das funções de ligação *complemento log-log* e *probit* como funções de ativação, e compararam seu uso como o uso de funções de ativação sigmóide logística. Segundo os autores, o uso das funções propostas é mais adequado em casos específicos de dados que seguem uma distribuição binomial e em casos de presença de observações extremas.

Sharma and Chandra (2010) estudaram a aplicação de redes neurais construtivas a problemas de regressão e verificaram o papel da função de ativação sigmóide logística com inclinação adaptativa em redes neurais *feedforward* construtivas para um melhor desempenho em generalização e menor tempo de treinamento.

3.2.2. TOPOLOGIA DAS REDES NEURAI ARTIFICIAIS (Arquiteturas de rede).

De acordo com a topologia, uma rede neural artificial pode ser classificada como:

3.2.2.1. REDE NEURAL ARTIFICIAL TIPO ALIMENTADA ADIANTE OU FEEDFORWARD

A rede neural está separada em camadas, onde todas as unidades enviam suas saídas apenas para as unidades situadas na próxima camada. Estas redes não têm realimentação local, não existem laços (NASCIMENTO e YONEYAMA, 2008).

A rede neural é dita *totalmente conectada* quando cada unidade de uma camada da rede está conectada a todas as unidades da camada adjacente seguinte. A rede é *parcialmente conectada* se algumas das conexões estiverem faltando na rede (HAYKIN, 2007).

Em uma rede neural em camadas, a mais simples é a que tem uma camada de nós de entrada e uma camada de neurônios de saída (nós computacionais). As redes com múltiplas camadas têm uma ou mais *camadas ocultas*. Por concisão, uma rede *feedforward* com i entradas, $h1$ neurônios na primeira camada oculta, $h2$ neurônios na segunda camada oculta e o neurônios na camada de saída é referida como uma rede $i: h1: h2: o$.

3.2.2.2. REDE NEURAL ARTIFICIAL TIPO RECORRENTE OU FEEDBACK

Uma rede recorrente ou *feedback* se distingue de uma rede neural *feedforward* por ter pelo menos um laço de realimentação local (KOLEN e KREMER, 2001). Segundo Nascimento e Yoneyama (2008), uma rede tipo *feedback* é em geral um sistema dinâmico não linear em que a estabilidade da rede é um tópico de grande importância. Mandic e Chambers (2001) apontam que as conexões recorrentes na rede neural podem torná-la mais instável e mais sensível ao ruído. Braga et al. (2000) manifestam que as redes recorrentes são mais apropriadas para resolver problemas que envolvem processamento temporal.

3.3. PERCEPTRONS DE MÚLTIPLAS CAMADAS

São redes de múltiplas camadas alimentadas adiante (*feedforward*), consistem de um conjunto de unidades sensoriais que constituem a *camada de*

entrada, e neurônios (unidades computacionais) distribuídos em uma ou mais *camadas ocultas* e uma *camada de saída*. Os sinais de entrada se propagam para frente através da rede camada por camada. Segundo Haykin (2007), essas redes têm sido aplicadas com sucesso para resolver diversos problemas difíceis através de seu treinamento de forma supervisionada com o algoritmo de retropropagação de erro (backpropagation). Hornik et. al (1989) afirmam que as redes *feedforward* de múltiplas camadas são aproximadores universais, porque têm a capacidade de aproximar qualquer função mensurável em qualquer grau desejado de precisão, e que qualquer falta de sucesso em aplicações deve surgir de aprendizagem inadequada, número insuficiente de unidades escondidas ou a falta de uma relação determinística entre a entrada e a saída desejada. A capacidade de aproximação de funções dessas redes é também demonstrada através da prova do Teorema de Cybenko (1989).

3.3.1. O ALGORITMO BACKPROPAGATION

Nesse algoritmo, os dados de entrada são repetidamente apresentados para a rede neural. A cada apresentação a saída da rede neural é comparada com a saída desejada e um valor de erro é calculado. O objetivo do algoritmo de *backpropagation* (retropropagação) é reduzir o erro pelo ajuste dos pesos sinápticos da rede, a cada iteração para que o resultado aproxime-se cada vez mais da saída desejada. O algoritmo fornece uma maneira de calcular o gradiente da função de erro na rede de forma eficiente usando a regra da cadeia de diferenciação. Uma apresentação do algoritmo é a seguinte:

Seja m o número de neurônios de saída. Suponha que $d_j(k)$ é a saída desejada a partir do neurônio de saída j , cuja saída real é $y_j(k)$. O sinal de erro e_j na saída do neurônio j , para a k -ésima iteração, é definido por:

$$e_j(k) = y_j(k) - d_j(k) \quad (3.12)$$

Definimos a soma dos quadrados do erro sobre todos os neurônios de saída para esta iteração k como:

$$E(k) = \frac{1}{2} \sum_{j=1}^m [y_j(k) - d_j(k)]^2 \quad (3.13)$$

e o erro total para todo o conjunto de N exemplos de treinamento como:

$$E_T = \sum_{k=1}^N E(k)$$

A variação do erro $E(k)$ devida à variação da saída é $y_j(k)$ do neurônio j na camada de saída é medida pela derivada parcial $\partial E(k) / \partial y_j(k)$, a qual é obtida diferenciando ambos os lados da Eq. (3.13):

$$\frac{\partial E(k)}{\partial y_j(k)} = y_j(k) - d_j(k) = e_j(k) \quad (3.14)$$

A entrada da rede para o neurônio j (o sinal aplicado à função de ativação do neurônio j) da camada de saída é:

$$s_j(k) = \sum_{i=1}^n w_{ji}(k) y_i(k) + b_j \quad (3.15)$$

Onde n é o número de entradas aplicadas ao neurônio j ; w_{ji} representa o peso sináptico conectando a saída do neurônio i à entrada do neurônio j ; y_i é o sinal funcional que aparece na saída do neurônio i da camada anterior na iteração k .

O bias b_j poder ser considerado como um peso sináptico w_{j0} que corresponde a uma entrada fixa $y_0 = +1$. Então a eq. (3.15) pode ser escrita:

$$s_j(k) = \sum_{i=0}^n w_{ji}(k) y_i(k) \quad (3.16)$$

A variação do erro $E(k)$ devido à variação da entrada $s_j(k)$ para o neurônio j na camada de saída é medida pela derivada parcial $\partial E(k) / \partial s_j(k)$, a qual é obtida diferenciando ambos os lados da Eq. (3.13):

$$\frac{\partial E(k)}{\partial s_j(k)} = \frac{\partial E(k)}{\partial y_j(k)} \frac{\partial y_j(k)}{\partial s_j(k)} \quad (3.17)$$

A derivada $\partial E(k) / \partial y_j(k)$ foi calculada na eq. (3.14) e a derivada $\partial y_j(k) / \partial s_j(k)$ é calculada utilizando a função de ativação associada com o neurônio j . A função de ativação transforma a entrada $s_j(k)$ do neurônio j na saída $y_j(k)$. Um exemplo de função de ativação é a sigmóide: $y_j(k) = 1 / (1 + \exp(-\lambda s_j(k)))$,

onde λ é o ganho de ativação $\lambda > 0$, $s_j(k)$ é calculado pela eq. (3.16). As funções de ativação descrevem relacionamentos entrada / saída tais que $\partial y_j(k) / \partial s_j(k)$ existe e é finita.

O neurônio j na camada de saída esta conectado ao neurônio i da camada anterior com o peso sináptico w_{ji} . Então, precisamos calcular a variação do erro $E(k)$ devido à variação no peso sináptico $w_{ji}(k)$ que é medida pela derivada parcial $\partial E(k) / \partial w_{ji}(k)$:

$$\frac{\partial E(k)}{\partial w_{ji}(k)} = \frac{\partial E(k)}{\partial s_j(k)} \frac{\partial s_j(k)}{\partial w_{ji}(k)} \quad (3.18)$$

A derivada $\partial E(k) / \partial s_j(k)$ foi calculada na eq. (3.17) e a derivada $\partial s_j(k) / \partial w_{ji}(k)$ é calculada utilizando a eq. (3.15).

$$\frac{\partial s_j(k)}{\partial w_{ji}(k)} = y_i(k)$$

A derivada parcial $\partial E(k) / \partial w_{ji}(k)$ pode ser obtida pela aplicação da regra da cadeia, utilizando a Eq. (3.17) em (3.18):

$$\frac{\partial E(k)}{\partial w_{ji}(k)} = \frac{\partial E(k)}{\partial y_j(k)} \frac{\partial y_j(k)}{\partial s_j(k)} \frac{\partial s_j(k)}{\partial w_{ji}(k)} \quad (3.19)$$

A eq. (3.19) permite o cálculo de $\partial E(k) / \partial w_{ji}(k)$ com os pesos sinápticos que conectam a entrada de cada neurônio j da camada de saída com a saída de cada neurônio i da camada anterior.

O erro $E(k)$ é afetado pelos pesos de conexão dos neurônios que estão localizados em outras camadas da rede diferentes da camada de saída.

Por exemplo, na camada l , seja $y_j^{(l)}(k)$ a saída do neurônio j , $s_j^{(l)}(k)$ a entrada da rede a esse neurônio e $w_{ji}^{(l)}(k)$ o peso sináptico que conecta a entrada do neurônio

j com a saída do neurônio i da camada anterior ($l - 1$). Para o neurônio j da camada l a entrada está relacionada com a saída pela equação:

$$y_j^{(l)}(k) = f_j^{(l)}(s_j^{(l)}(k)) \quad (3.20)$$

Além disso, $s_j^{(l)}(k)$ pode ser expressa como a soma ponderada das saídas dos neurônios da camada anterior ($l - 1$):

$$s_j^{(l)}(k) = \sum_{i=0}^{n^{(l-1)}} w_{ji}^{(l)}(k) y_i^{(l-1)}(k) \quad (3.21)$$

Pela aplicação da regra da cadeia é calculada a derivada $\partial E(k) / \partial w_{ji}^l(k)$:

$$\frac{\partial E(k)}{\partial w_{ji}^{(l)}(k)} = \frac{\partial E(k)}{\partial y_j^{(l)}(k)} \frac{\partial y_j^{(l)}(k)}{\partial s_j^{(l)}(k)} \frac{\partial s_j^{(l)}(k)}{\partial w_{ji}^{(l)}(k)} \quad (3.22)$$

A saída do neurônio j da camada l pode ser conectada a mais de um neurônio na camada acima. Somando sobre todas as conexões que emanam do neurônio j para a camada acima, temos:

$$\frac{\partial E(k)}{\partial y_j^{(l)}(k)} = \sum_j \frac{\partial E(k)}{\partial s_j^{l+1}} w_{ji}^{l+1} \quad (3.23)$$

Desde que na eq. (3.15)

$$\frac{\partial s_j^{l+1}}{\partial y_i^{(l)}} = w_{ji}^{l+1} \quad (3.24)$$

O algoritmo de retropropagação aplica uma correção $\Delta w_{ji}(k)$ ao peso sináptico $w_{ji}(k)$ que é proporcional à derivada parcial $\partial E(k) / \partial w_{ji}(k)$.

Os pesos sinápticos da rede são atualizados depois que todas as derivadas parciais $\partial E(k) / \partial w_{ji}(k)$ são calculadas para todas as conexões na rede, os pesos são atualizados de acordo com a regra delta generalizada:

$$w_{ji}^l(k+1) - w_{ji}^l(k) = -\eta \frac{\partial E(k)}{\partial w_{ji}^l(k)} \quad (3.25)$$

Onde η é o parâmetro da taxa de aprendizagem do algoritmo de retropropagação. O uso do sinal negativo na eq. (3.25) indica a descida do gradiente no espaço dos pesos.

Segundo Bose e Liang (1996) uma tentativa para melhorar a velocidade de convergência, enquanto se minimiza a possibilidade de oscilação consiste em adicionar um termo nomeado de *momentum* à formulação básica da diminuição do gradiente. Neste caso, o vetor dos pesos no tempo $(k+1)$ é relacionado aos vetores de pesos nos tempos (k) e $(k-1)$ por:

$$w_{ji}^l(k+1) = w_{ji}^l(k) - \left[\eta \frac{\partial E}{\partial w_{ji}^l(k)} + \beta \Delta w_{ji}^l(k-1) \right] \quad (3.26)$$

Onde β é o momentum, que determina o efeito das mudanças passadas na mudança atual de pesos.

Na aplicação do algoritmo de retropropagação, distinguem-se dois passos distintos no processo de computação do erro. O primeiro passo é conhecido como *passo para frente*, ou *propagação*, e o segundo como *passo para trás*, ou *retropropagação*.

A propagação começa na primeira camada oculta, se calculam as entradas $s_j^{(l)}(k)$, Eq. (3.21) e saídas $y_j^{(l)}(k)$, Eq. (3.20), dos neurônios da rede prosseguindo para frente através da rede, camada por camada. Se o neurônio está na primeira camada oculta, na Eq. (3.20): $y_i^{(l-1)}(k) = x_i(k)$, onde $x_i(k)$ é o i -ésimo elemento do vetor de entrada $\mathbf{x}(k)$. Na camada de saída se calcula o sinal de erro $e_j(k)$, Eq. (3.12).

A retropropagação é o processo em que se calculam as derivadas parciais $\partial E(k) / \partial w_{ji}^l(k)$ e se ajustam os pesos sinápticos de acordo com a regra delta

generalizada, Eq. (3.26), começando na camada de saída e logo camada por camada através da rede até modificar os pesos sinápticos que conectam as variáveis de entrada com os neurônios da primeira camada oculta.

3.4. A GENERALIZAÇÃO

O termo “generalização” é tomado emprestado da psicologia. A generalização é um fenômeno em psicologia onde os sujeitos condicionados respondem a estímulos semelhantes aos que foram condicionados a responder.

Para as redes neurais, depois que a rede tenha aprendido com sucesso um conjunto de N exemplares de treinamento, a capacidade de generalização da rede faz referência à precisão com que a rede prediz novos exemplares desconhecidos, mas obtidos com a mesma distribuição de probabilidade que os exemplares de treinamento (BOSE e LIANG, 1996).

No caso de problemas de aproximação de funções, uma rede neural treinada que realiza boas interpolações com dados de teste retirados da mesma população utilizada para obter os dados de treinamento, pode ser considerada como uma rede que tem boa capacidade de generalização.

O sobre-treinamento (*overtraining*) e o sobreajuste (*overfitting*) são dois fenômenos que afetam a capacidade de generalização das redes neurais, o conceito de *overtraining* é explicado por Bose e Liang (1996) como o fenômeno pelo qual, após certo número de ciclos (épocas) de treinamento, mais épocas de treinamento reduzirão ainda mais o erro de aprendizagem (apenas ligeiramente em muitos casos) no conjunto de treinamento, mas vão produzir erros cada vez maiores (geralmente em uma quantidade muito considerável) em exemplares de teste.

Nascimento e Yoneyama (2008) explicam o conceito de *overfitting*, considerando o treinamento de uma rede neural de tipo *feedforward* como uma versão estendida e multidimensional do problema de determinar os parâmetros de um polinômio para interpolar um conjunto de pontos de um espaço

unidimensional. Um excessivo número de parâmetros no polinômio (um número excessivo de pesos na rede neural) pode resultar em *overfitting*, ou seja, o polinômio passa exatamente por todos os pontos usados durante o treinamento sem descrever a verdadeira relação de entrada-saída que gerou os pontos. Segundo a explicação, o *overfitting* se produz pelo excessivo número de parâmetros que são utilizados pela função (polinomial) que representa uma população de pontos. No caso da rede neural, a situação ideal é determinar o menor número de neurônios nas camadas escondidas que possa produzir uma boa generalização. Demuth *et al.*, (2008) também afirmam que reduzir a rede até um tamanho pequeno o suficiente pode prevenir o *overfitting*, quanto maior é a rede mais complexas são as funções que a rede pode criar.

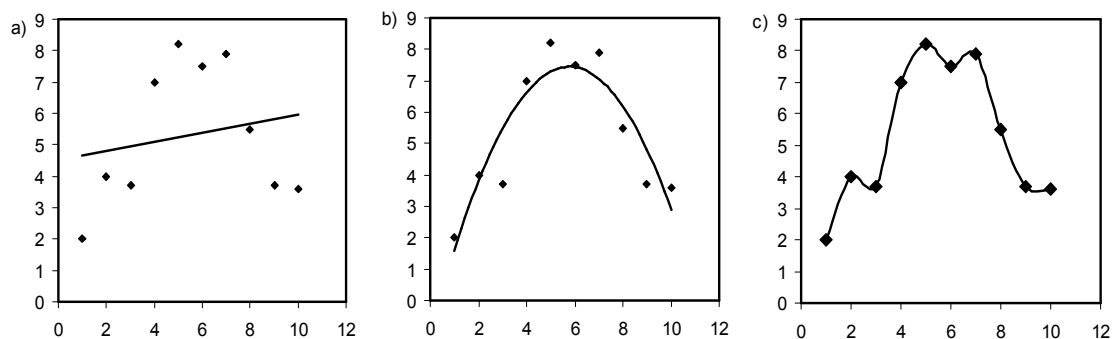


Figure 3.2. a) *underfitting*, b) *fitting* , c) *overfitting*

Segundo Haykin (2007), o sobreajuste (*overfitting*) se produz quando a rede é treinada em excesso (*overtraining*), como consequência ela perde a habilidade de generalizar entre padrões de entrada-saída similares, por exemplo, quando a rede neural aprende um número excessivo de padrões de entrada-saída, a rede acaba memorizando os dados de treinamento, então as saídas da rede mostram características presentes nos dados de treinamento, mas não na função subjacente que deve ser modelada.

O sobretreinamento como causa do sobreajuste é também mencionado por Tetko *et al.* (1995), quando afirma que o *overfitting* não tem nenhuma influência

sobre a habilidade de predição da rede neural quando o *overtraining* é evitado utilizando validação cruzada.

Do acima mencionado, temos que os fatores que ocasionam *overtraining* estão relacionados com: o número de exemplos de treinamento, o número de ciclos de treinamento e o tempo de treinamento; e os fatores causadores de *overfitting* estão relacionados com: o *overtraining*, o número de pesos sinápticos (número de parâmetros da rede) e um tamanho de rede maior do que o tamanho ótimo (sobredimensionamento da rede).

Na literatura, afirma-se que a generalização é influenciada por diversos fatores, como descritos, a seguir.

Bartlett (1997) afirma que para uma generalização válida a magnitude dos pesos é mais importante que o tamanho da rede neural. Weigend (1994), depois de analisar o desempenho das redes em função do seu tamanho afirma que as redes neurais grandes generalizam melhor do que as pequenas. Outros autores suportam a idéia de que as redes sobredimensionadas geralmente apresentam pouca capacidade de generalização, neste grupo estão: 1) os que propõem o uso de métodos de poda (*pruning*), cuja idéia é iniciar com uma arquitetura de dimensão elevada e ir retirando neurônios ou conexões até que se chegue a uma dimensão adequada e 2) os que propõem o uso de métodos construtivos, nos quais se inicia com uma arquitetura mínima de rede e se adicionam neurônios até que uma estrutura adequada seja encontrada.

Bose e Liang (1996) afirmam que, em geral, uma rede neural menor generaliza melhor do que uma maior. Mas menciona que quando a rede neural é muito pequena, ambos o erro de ajuste e o erro de interpolação podem ser grandes. Huss e Horne (1993) mencionam que, a generalização é influenciada por três fatores: 1) o tamanho do conjunto de treinamento e o quão representativo do

ambiente de interesse ele é, 2) a arquitetura da rede neural e 3) a complexidade física do problema em questão.

Narayan and Tagliarini (2005) manifestam que a habilidade de generalização de uma rede *perceptron* multicamada tem sido mostrada estar relacionada ao número e magnitudes dos pesos da rede. Segundo seus resultados, grandes magnitudes nos pesos potencialmente aumentam a propensão de uma rede para interpolar mal e quando são impostos limites nos pesos, o algoritmo *backpropagation* é capaz de descobrir redes com magnitudes de peso pequenas que exibem boa generalização.

Hermundstad, et al. (2011) estudaram a interdependência entre organização estrutural e desempenho funcional em redes neurais artificiais, e encontraram que arquiteturas de rede diferentes produzem superfícies de erro com características distintas, tais como a altura e largura dos mínimos locais, que por sua vez, determinam características de desempenho, como velocidade, precisão e adaptabilidade (capacidade de generalização), em tarefas de aproximação sequencial supervisionada de funções.

3.5. O PAPEL DAS UNIDADES ESCONDIDAS DOS *PERCEPTONS*

Os modelos de redes neurais artificiais são conhecidos também como modelos de caixa preta, no sentido que não se conhece bem como os neurônios operam, para conseguir que a rede proporcione saídas iguais ou aproximadas às saídas desejadas. No entanto, se conhece que os neurônios das camadas escondidas modificam a funcionalidade das redes neurais, por exemplo, sem camadas escondidas não é possível aprender relações entrada-saída não lineares, só se consegue resolver problemas linearmente separáveis (BRAGA *et al.*, 2000).

Nascimento e Yoneyama (2008) afirmam que o uso de camadas escondidas torna possível a recodificação dos padrões de entrada, e que cada camada escondida executa a recodificação do seu sinal de entrada. Com essa atividade, os neurônios das camadas escondidas permitem à rede extrair os atributos escondidos dos dados fornecidos nas entradas.

As redes neurais *feedforward* do tipo *perceptron* multicamadas são aplicadas para resolver problemas diferentes: aproximação de funções, classificação de padrões, controle de processos, filtragem de sinais, previsão não-linear de séries temporais, entre outros. Tendo em conta, as diferentes características das tarefas necessárias para resolver cada problema, seria errado concluir que uma camada de neurônios escondidos é suficiente para a rede resolver todos os tipos de problemas, é mais apropriado considerar que o número de camadas internas está em função ao problema que é objeto de aplicação de redes neurais.

Na literatura sobre o número de camadas a serem utilizadas para resolver problemas de aproximação de funções, podemos distinguir duas propostas, como mostrado em seguida.

1) Utilizar apenas uma camada de neurônios escondidos

Os autores (HASSOUN, 1995; BOSE e LIANG, 1996, entre outros) mencionam que é preciso o uso de uma função de ativação sigmoideal na camada oculta de neurônios em referência estrita ao teorema de aproximação universal de Cybenko (1989).

O teorema Cybenko pode ser formulado como:

Suponha que σ seja uma função sigmoideal contínua, sejam α_i e b_i números reais fixos, sejam w_{ij} elementos da matrix de pesos, onde $i=1,\dots,N$ e $j=1,\dots,M$. Então, dado $\varepsilon > 0$, existe uma soma $G(x_1,\dots,x_M)$ da forma:

$$G(x_1,\dots,x_M) = \sum_{i=1}^N \alpha_i \sigma\left(\sum_{j=1}^M w_{ij} x_j + b_i\right) \quad (3.27)$$

para a qual

$$|G(x_1,\dots,x_M) - f(x_1,\dots,x_M)| < \varepsilon \quad (3.28)$$

para todo x que se encontre no espaço de entrada.

Desde que a equação (3.27) representa a saída de um *perceptron* com uma camada oculta, o teorema afirma que um *perceptron* com uma única camada oculta e uma função de ativação sigmoideal contínua pode aproximar funções contínuas de M variáveis reais, com precisão arbitrária, sempre que não sejam colocadas restrições sobre o número de neurônios ou o tamanho dos pesos Cybenko (1989).

2) Utilizar duas ou mais camadas de neurônios escondidos

Os autores baseiam-se em limitações identificadas no teorema de Cibenko e em diferentes funções atribuídas às camadas de neurônios.

Com relação às limitações do teorema, para Cybenko (1989) não foram resolvidas questões como: quantos neurônios são necessários para produzir uma aproximação de uma qualidade dada?, que propriedades da função que está sendo aproximada desempenham um papel importante na determinação do número de neurônios?

Haykin (2007) assinala que embora o teorema afirme que um *perceptron* com uma única camada oculta pode aproximar com precisão arbitrária qualquer função contínua de múltiplas variáveis, o teorema não diz que a única camada

oculta é ótima no sentido do tempo de aprendizagem, facilidade de implementação, ou qualidade de generalização.

De fato, se olharmos para as equações (3.27) e (3.28) temos que o teorema de Cibenko estabelece que para um valor $\varepsilon > 0$ ($\varepsilon \in \mathbb{R}$), a magnitude da diferença $d = |G(x_1, \dots, x_M) - f(x_1, \dots, x_M)|$ entre a saída $G(x_1, \dots, x_M)$ de um perceptron com uma camada oculta e o valor da função $f(x_1, \dots, x_M)$ que está sendo aproximada; se encontra no intervalo $0 < |G(x_1, \dots, x_M) - f(x_1, \dots, x_M)| < \varepsilon$, com as seguintes conclusões:

- 1) Quanto mais próxima de 0 é a diferença, mais sobreajustada é a rede, portanto com menor capacidade de generalizar.
- 2) Existe mais de um perceptron com uma camada de neurônios oculta, cuja saída está no intervalo $] 0, \varepsilon [$, incluindo aquele com a melhor capacidade de generalização.
- 3) Quanto maior é o valor de ε , menor é o ajuste, resultando em uma pior representação da função implícita que relaciona entradas e saídas, e portanto, em uma pior capacidade de generalização.
- 4) O teorema não afirma que, em matéria de aproximação de funções, a generalização que pode ser obtida com um *perceptron* que tem uma camada oculta é melhor do que a que pode ser obtida com *perceptrons* que têm duas ou mais camadas ocultas.

Por outro lado, Haykin (2007) também baseia a sua proposta fazendo referência às características de funcionamento dos neurônios no processo de aproximar funções, quando diz que os neurônios da única camada oculta tendem a interagir entre si globalmente, e que em situações complexas esta interação torna difícil melhorar a aproximação em um ponto, sem piorá-la em algum outro, manifestando que, nesse caso, com duas camadas ocultas o processo de aproximação se torna mais gerenciável, pelo que pode se proceder tendo em conta que as camadas de neurônios escondidos cumprem diferentes funções: 1)

os neurônios da primeira camada oculta extraem as características locais que existem nas entradas, e 2) as características globais são extraídas na segunda camada oculta, onde os neurônios combinam as saídas dos neurônios da primeira camada. Haykin (2007) propõe um processo de aproximação de dois estágios, abordagem que considera similar em filosofia à técnica *spline* de ajuste de curvas, na qual ocorre um processo de aproximação polinomial por partes.

Yegnanarayana (2006) baseia a sua proposta em um ponto de vista conceitual e considera que:

- a) O problema de determinar a relação funcional entre os pontos de um espaço M dimensional de variáveis de entrada e os pontos de um espaço N dimensional de variáveis de saída, é chamado de problema de mapeamento de padrões.
- b) Uma rede neural com pelo menos duas camadas ocultas pode realizar qualquer tarefa de classificação de padrões complexa. Esta rede também pode realizar uma tarefa de mapeamento de padrões.
- c) Portanto, pelo menos duas camadas ocultas são necessárias em uma rede neural para capturar a relação funcional implícita entre pares de padrões de entrada-saída dados em problemas de aproximação de funções.

Em problemas de classificação de padrões, o critério teórico mais encontrado na literatura é utilizar *perceptrons* de mais de uma camada de neurônios ocultos para aproveitar o efeito de funções específicas das diferentes camadas de neurônios no processo de distribuição das entradas nas diferentes classes de acordo com os exemplos de treinamento (BOSE e LIANG, 1996; YEGNANARAYANA, 2006; HAYKIN, 2007). Na aplicação prática, os pesquisadores utilizam uma ou mais camadas de neurônios ocultos, em problemas de classificação de padrões.

3.6. APLICAÇÕES DAS REDES NEURAIIS: PONTOS FORTES E FRACOS

As redes neurais artificiais têm numerosas aplicações, entre elas:

Associação de padrões, classificação de padrões, aproximação de funções, controle de sistemas dinâmicos (manipuladores mecânicos, processos industriais, controle de veículos), processamento de sinais (cancelamento de ruídos, reconhecimento de voz, e outros), sistemas de decisão (médicos, financeiros, e outros), compressão de imagens e outras diversas possibilidades que são descritas na literatura especializada, por exemplo, recuperação da informação (FERNEDA, 2006), estimação da temperatura interna de transformadores de distribuição imersos em óleo (FREITAS *et al*, 2002), etc.

Simões e Shaw (2007) apontam que seu ponto forte de aplicação está em: reconhecimento de padrões, reconhecimento de caracteres e formas, estimação de funções não lineares, controle de processos e previsões financeiras. No caso de controle de processos, na ausência de modelos matemáticos de processos, as redes neurais podem utilizar um histórico de dados coletados dos sinais dos sensores e medidores utilizados na planta de processo, para se construir modelos preditivos, modelos que podem prever as reações do processo a novas condições. Neste trabalho, são utilizados este tipo de modelos em estratégias de controle preditivo.

Entretanto, mencionam vários pontos fracos, entre eles salientamos os seguintes:

- “O tempo de treinamento das redes neurais artificiais não é previsível, podendo em alguns casos ser muito longo. O retreinamento “*on line*” em geral requerido para um sistema de autossintonia só pode ser aplicado para sistemas com dinâmica muito lenta”.

Em geral, o tempo de treinamento depende de vários fatores: a natureza dos dados, a natureza do problema de modelagem, a quantidade de padrões entrada/saída, a complexidade do modelo e o método de treinamento; portanto o tempo de treinamento pode ser muito variado. No entanto, o conhecimento desses fatores facilita fazer previsões qualitativas do tempo de treinamento.

No que se refere ao retreinamento “*on line*”, é necessário dados do processo obtidos em tempo real, e para a sua aplicação a necessidade do retreinamento “*on line*” para um sistema de dinâmica muito lenta, tendo em conta que as redes neurais treinadas “*off line*” por ter em seu vetor de treinamento um conjunto de dados de várias perturbações e maior número de dados consegue representar bem o perfil do processo, segundo Duarte (2004) essas redes tem uma maior “memória” do processo; e que a cada retreinamento “*on line*” se modificam os parâmetros da rede neural para uma utilidade mais específica com possíveis limitações nas condições do retreinamento (número de dados, número de iterações, critério de tolerância) e perda do histórico do processo.

- “Uma vez que não se pode olhar para o sentido dos pesos e conexões de uma rede neural, torna-se impossível de se interpretar as causas de um comportamento particular”.

Essa preocupação é importante porque na literatura é mencionado que o modelo de redes neurais artificiais é um modelo de caixa preta. No entanto, a informação que permite o ajuste entre entradas e saídas do modelo está contida nos pesos e bias da rede neural. Neste trabalho, é desenvolvido um método que contribui para a interpretação da informação causal contida nos pesos e bias do modelo de redes neurais e para a explicação dos comportamentos das variáveis.

3.7. CONCLUSÃO

Neste capítulo, a fundamentação teórica tem como objetivo construir um marco conceitual que oriente as tarefas de utilização das redes neurais na modelagem empírica e controle de processos nos seguintes capítulos do presente trabalho. Apresentam-se características das redes neurais, os tipos de rede: alimentada adiante ou feedforward e recorrente ou feedback; revisam-se regras e paradigmas de aprendizagem, a capacidade de generalização das redes, o algoritmo de treinamento por *backpropagation*, o papel dos neurônios escondidos, e os pontos fortes e fracos das redes neurais.

A propriedade mais interessante das redes neurais é sua capacidade de processamento paralelo de informação para aprender e generalizar com base em exemplos, e através da adaptação dos pesos das conexões. Os neurônios das camadas ocultas têm um papel importante no processamento de informação, determinando a funcionalidade e desempenho das redes neurais em diversas tarefas.

3.9. REFERÊNCIAS BIBLIOGRÁFICAS

1. Ackley, D. H., Hinton, G. E., and Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science*, 9, pp.147-169.
2. Atala, D.I.P. **Fermentação Alcoólica com Alta Densidade Celular: Modelagem Cinética, Convalidação de Parâmetros e Otimização do Processo**. Dissertação de Mestrado, Universidade Estadual de Campinas, Brasil, 2000.
3. Assidjo, E., Yao B., Amame, D., Ado, G., Azzaro-Pantel, C. and Davin, A. (2006). Industrial Brewery Modelling by Using Neural Network. *Journal of Applied Sciences*, v. 6 (8), pp.1858-1862.
4. Assidjo, N.E. , Akaki, D., Yao, K.B., Eboi, T.Y. (2009). A Hybrid Neural Approach for Batch Fermentation Simulation. *Australian Journal of basic and Applied Sciences*, v. 3 (4), pp. 3930-3936.
5. Barrera-Cortes, J. and Baruch, I. (2000). **Recurrent Neural Network Model of a Fed-Batch Saccharomyces Cerevisiae Fermentation Process**. Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN'00), v. 4, pp. 589-594.
6. Bartlett, P.L.(1997). For valid generalization, the size of the weights is more important than the size of the network. *Advances in Neural Information Processing Systems*, v. 9, pp. 134-140.
7. Bose, N.K. and Liang, P. (1996). **Neural network fundamentals with graphs, algorithms, and applications**. Mcgraw-Hill, Inc. USA.
8. Braga, A. P., Carvalho A.C.P.L.F.; Ludemir, T. B. **Redes Neurais Artificiais: Teoria e Aplicações**. Rio de Janeiro, Brasil: LTC - Livros Técnicos e Científicos Editora S.A., 2000.

9. Brown, T.H., Kairiss, E.W., and Keenan, C.L. (1990) . Hebbian synapses: biophysical mechanisms and algorithms, *Annual Review of Neuroscience*, v. 13, pp. 475-511.
10. Costa, A. C., Atala, D. I. P., Maciel Filho, R. and Maugeri Filho, F. (2001). Factorial Design and Simulation for the Optimization and Determination of Control Structures for an Extractive Alcoholic Fermentation. *Process Biochemistry*, v. 37(2), pp. 125-137.
11. Costa, A. C., Meleiro, L.A.C., Maciel Filho, R.(2002). Non-linear predictive control of an extractive alcoholic fermentation process. *Process Biochemistry*, v. 38 (5), pp. 743-750.
12. Cybenko, G. (1989). Approximation by Superpositions of a Sigmoidal Function. *Mathematics of Control, Signals and Systems*, v.2, pp. 303-314.
13. Demuth, H., Beale, M., Hagan, M. Neural Network Toolbox™ 6 User's Guide. The MathWorks, 2008. Disponível em:
< <http://www.varpa.org/Docencia/Files/nnet.pdf>>. Acesso em: 3 dezembro 2009.
14. Duarte, E.R. **Estratégia de controle não linear baseada em redes neurais artificiais com aprendizagem “On line”**. Dissertação de Mestrado. Universidade Estadual de Campinas. Brasil, 2004.
15. Duarte, E.R. **Avaliação e desenvolvimento de algoritmos de controle aplicado a um processo extrativo de fermentação alcoólica contínua**. Tese de Doutorado. Universidade Estadual de Campinas. Brasil, 2007.
16. Ferneda, E. Redes Neurais e sua aplicação em sistemas de recuperação de informação. *Ci. Inf., Brasília*, v. 35 (1), pp. 25-30, 2006.
17. Freitas, A.A.C; Silva, I.N.; Souza, A.N. (2002). Aplicação de redes neurais na estimação da temperatura interna de transformadores de distribuição imersos em óleo. *Sba. Controle & Automação*, v. 13. (3). Disponível em:
<http://www.scielo.br/scielo.php?pid=S0103-17592002000300005&script=sci_arttext>. Acesso em: 20 janeiro 2009.
18. Galland C. and Hinton G.E. (1989) **Deterministic Boltzmann Learning in Networks with Asymmetric Connectivity**. University of Toronto. Department of Computer Science Technical Report. CRG-TR-89-6.
19. Gomes, G.S.S., e Ludermir, T.B. (2008). Redes neurais artificiais com funções de ativação complemento log-log e probit para aproximar funções na presença de observações extremas. Learning and Nonlinear Models. *Revista da Sociedade Brasileira de Redes Neurais (SBRN)*, v. 6, n. 2, pp. 142-153.

20. Guarnieri, S., Piazza, F., and Uncini, A. (1999). Multilayer Feedforward Networks with Adaptive Spline Activation Function. *IEEE Transactions on Neural Networks*, v. 10 (3). pp. 672-683.
21. Harada, L.H.P. (2001). **Modelagem híbrido neuronal aplicada a processos fermentativos**. Dissertação de Mestrado. Universidade Estadual de Campinas. Brasil, 2001.
22. Hassoun, M. **Fundamentals of artificial neural networks**. Massachusetts Institute of Technology, USA, 1995.
23. Haykin, S. (2007). **Redes neurais**. Princípios e prática. 2ª Edição. ARTMED EDITORA S.A., São Paulo, Brasil.
24. Hebb, D.O. **The Organization of Behavior: a neuropsychological theory**. New York: Jhon Wiley & Sons, 1949.
25. Hermundstad, A.M., Brown, K.S., Bassett, D.S. and Carlson, J.M. (2011). Learning, Memory, and the Role of Neural Network Architecture. *PLoS Comput Biol.*, v. 7(6): e1002063. doi:10.1371/journal.pcbi.1002063.
26. Hinton, G. **Supervised learning in supervised neural networks**. In: Robert, A.W. and Keil, F.C. (Eds). *The MIT Encyclopedia of the Cognitive Sciences*. 2001. MIT Press. Cambridge, Massachusetts, London, England. pp. 814-816.
27. Hornik, K., Sinchcombe, M., and White, M.(1989). Multilayer Feedforward are Universal Approximators. *Neural Networks*, v. 2. pp. 359-366.
28. Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, v. 4(2), pp. 251–257.
29. Hush, D. and Horne, B. (1993). Progress in supervised neural networks. *IEEE Signal Processing Magazine*, v. 10(1), pp. 8-39.
30. Kolen, J.F. and Kremer, S. (2001). **Dynamical Recurrent Networks**. In: J.f. Kolen and S. C. Kremer (Eds). *A field guide to dynamical recurrent neural network*. Institute of Electrical and Electronics Engineers, Inc. IEE. Press .New York, NY. USA. 2001
31. Leshno, M., Lin, V.Y., Pinkus, A., Schocken, S.,(1993). Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, v. 6 (6), pp. 861-867
32. McCulloch, W.S. and Pitts, W. (1943). A Logical Calculus of the ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, v. 5, pp.115-133.

33. Mandic, D.P. and Chambers, J.A. **Recurrent Neural Networks for Prediction**. Chichester, England: Jhon Wiley & Sons, Inc., 2001.
34. Mantovaneli, I.C.C. **Modelagem híbrido neuronal de um processo de fermentação alcoólica**. Dissertação de Mestrado. Universidade Estadual de Campinas. Brasil, 2005.
35. Meleiro, L. A. C, Costa, A.C., Maciel Filho, R. (2005). Non-Linear Multivariable Predictive Control of an Alcoholic Fermentation Process Using Functional Link Networks. *Brazilian Archives of Biology and Technology*, v.48, pp. 7-18.
36. Menon, A., Mehrotra, K., Mohan, C. K., and Ranka, S. (1996). Characterization of a class of sigmoid functions with applications to neural networks. *Neural Networks*, 9(5), pp. 819-835.
37. Menon, A., Mehrotra, K., Mohan, C. K., and Ranka, S. **Elements of artificial neural networks**. Cambridge: The MIT Press, 2000.
38. Narayan, S. and Tagliarini, G. An analysis of underfitting in MLP networks. *Proceedings of the 2005 IEEE International Joint Conference on Neural Networks (IJCNN)*, Montreal, Canada, v. 2, pp. 984-988.
39. Nascimento C.L. e Yoneyama, T. (2008). **Inteligência artificial em controle e automação**. Blücher-FAPESP. São Paulo, Brasil.
40. Pramanik, K., 2004. Use of artificial neural networks for prediction of cell mass and ethanol concentration in batch fermentation using *saccharomyces cerevisiae* yeast. *Institution of Engineers (India) Journal-CH*.Vol. 85, pp. 31-35.
41. Rivera, E.C., Mantovaneli, I., Costa, A.C. and M.R. Filho, (2006). Hybrid modeling for continuous production of bioethanol. *Computer Aided Chemical Engineering*. , v. 21, pp. 613-618.
42. Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986b) **Learning internal representations by error propagation**. In: Rumelhart, D. E. and McClelland, J. L., (Eds.). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Volume 1: Foundations, MIT Press, Cambridge, MA. pp. 318-362.
43. Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986a). Learning representations by back propagating errors. *Nature*, v. 323, pp. 533-536.
44. Sharma, S.K., Chandra, P. (2010). Constructive neural networks: a review. *Intenational Journal of Engineering Science and Technology*. 2(12), pp. 7847-7855.

45. Silva, F.L.H. **Modelagem, simulação e controle de fermentação alcoólica contínua extrativa**. Campinas. Tese Doutorado. Faculdade de Engenharia de Alimentos, Universidade Estadual de Campinas, Brasil, 1998.
46. Simões, M.G. e Shaw, I.S. **Controle e modelagem fuzzy**. 2ª edição. São Paulo, Brasil: Blücher - FAPESP, 2007.
47. Tetko, I.V., Livingstone, D.J., Luik, A.I. (1995). Neural Network Studies.1. Comparison of overfitting and overtraining. *Journal of Chemical Information and Computer Science*, v. 35(5), pp. 826-833.
48. Vecchi, L., Piazza, F. and Uncini, A. (1998). Learning and approximation capabilities of adaptive spline activation function neural networks. *Neural Networks*, v. 11(2), pp. 259–270.
49. Weigend, A. (1994). **On overfitting and the effective number of hidden units**. In Proceedings of the 1993 Connectionist Models Summer School. Michael Mozer, M., Smolensky, P., Touretsky, D., Elman, J., Weigend, A. (editors). pp. 335-342. Edit. Lawrence Erlbaum Associates, Inc., Broadway, Hillsdale, New Jersey. USA.
50. Yegnanarayana, B. **Artificial Neural Networks**. New Delhi, India: Prentice Hall of India, 2006.
51. Zhang, M. (2008). **A data simulation system using CSINC polynomial higher order neural networks**.
Disponível em: < <http://en.scientificcommons.org/43474694>>.
Acesso em: 15 outubro 2011.

CAPÍTULO 4. MODELAGEM E IDENTIFICAÇÃO DE PROCESSOS

4.1. INTRODUÇÃO

Os modelos são úteis para compreender, explicar e prever determinados aspectos da realidade, expressam o avanço da ciência e da tecnologia, e fazem possível o seu maior desenvolvimento. Modelar é como aproximar um círculo utilizando polígonos, com apenas três pontos de referência o modelo proposto será um triângulo, e o erro será significativo, com mais pontos de referência, um hexágono, com ainda mais pontos de referência, um decágono, o erro do modelo será menor, mas o modelo será sempre aproximado. A hipótese de trabalho foi que o aspecto da realidade sob estudo é um círculo, a pesquisa posterior vai esclarecer se é parte de um cilindro, um cone ou uma esfera.

Um modelo de processo é uma abstração matemática de um processo real (SEBORG *et al.*, 2004), sendo uma ferramenta útil para um melhor entendimento dos fenômenos envolvidos, assim como para o desenvolvimento, otimização e controle do processo em estudo. Os modelos matemáticos de processos podem ser classificados genericamente como teóricos, empíricos ou empírico-fenomenológicos.

Os modelos teóricos de processos são modelos lineares ou não lineares, denominados também de mecanísticos, fenomenológicos ou de princípios básicos, porque são modelos construídos utilizando as leis físicas, os princípios de conservação (balanços de massa, energia e/ou momento) que governam o comportamento do sistema real. A maior vantagem é que podem ser utilizados para prever em uma ampla faixa de condições. No entanto, o desenvolvimento deste tipo de modelos é usualmente demorado e exige esforço especialmente para processos complexos. Com pouco conhecimento do processo não é possível construir seu modelo fenomenológico.

Modelos empíricos são aqueles que não estão baseados em pressupostos teóricos fenomenológicos, sua construção é feita utilizando procedimentos de identificação de processos, que são descritos na próxima seção.

Modelos empírico-fenomenológicos, são aqueles que utilizam componentes teóricos e componentes empíricos (equações empíricas) que descrevem frações desconhecidas do modelo as quais com os pressupostos teóricos não se consegue descrever.

4.2. IDENTIFICAÇÃO DE PROCESSOS

A identificação dos processos envolve a construção de modelos de processos com dados de entrada/saída obtidos experimentalmente, sem recorrer a quaisquer pressupostos teóricos relativos à natureza fundamental e às propriedades do sistema. Esta atividade é também chamada de identificação de sistemas (LJUNG, 1996).

Ogunnaike e Ray (1994) assinalam que a identificação de processos trata os processos como “caixa preta”, não se requer um conhecimento preciso e completo do processo, somente que os dados de saída sejam obtidos em resposta a mudanças nas entradas. Os modelos obtidos são particularmente úteis para processos pouco entendidos ou processos complexos, se precisam de métodos especiais para obter modelos não lineares.

O processo de construir um modelo empírico por identificação de processos consiste nas seguintes etapas:

1. Definição do problema
2. Coleta e pré-tratamento dos dados
3. Seleção da estrutura do modelo
4. Estimação dos parâmetros da estrutura selecionada
5. Validação do modelo

4.2.1. DEFINIÇÃO DO PROBLEMA

Nesta etapa definem-se os objetivos e impõem-se limites da abordagem, delimitando a complexidade do modelo.

a. Definição dos objetivos do modelo

Antes de iniciar as tarefas da modelagem é necessário definir os propósitos para os quais o modelo de processo vai ser construído.

b. Delimitar a complexidade do modelo

Algumas considerações básicas para delimitar a complexidade do modelo são:

- 1) Diferentes modelos são possíveis para um dado processo e os modelos podem ser mais úteis para uma aplicação que para outra. O valor fundamental de qualquer modelo é a sua utilidade, de forma que é necessário eleger o modelo mais útil para a aplicação desejada.
- 2) O modelo não pode representar todos os aspectos macroscópicos e microscópicos de um processo, de forma que é necessário salientar aspectos de interesse em função dos objetivos do modelo.
- 3) Os aspectos de interesse são expressões ainda gerais, portanto é preciso operacionalizá-los identificando e definindo as variáveis de entrada e saída do modelo.
- 4) O modelo empírico é desenvolvido diretamente a partir de dados experimentais, sendo necessário verificar se a informação experimental disponível é suficiente para construir o modelo com a complexidade desejada ou se são necessários mais dados.

4.2.2. COLETA E PRÉ-TRATAMENTO DOS DADOS

O desenvolvimento atual nas técnicas digitais de instrumentação, o aumento na capacidade de armazenamento e a melhoria no desempenho dos recursos computacionais atuais, faz possível o monitoramento contínuo em tempo

real do comportamento do processo, onde centenas ou milhares de variáveis de processo são simultaneamente medidas e registradas em volumosos meios de armazenamento (PRATA, 2009b). A informação é facilmente acessível e útil para o desenvolvimento de modelos, otimização e controle da operação das plantas industriais. Entretanto, como os dados são obtidos através de instrumentos físicos que apresentam um nível de precisão e estão sujeitos a diversas variabilidades na operação do processo, estão naturalmente corrompidos por erros. Portanto, para garantir que os dados forneçam informações precisas e retratem o processo de forma fidedigna, é necessário utilizar procedimentos de pré-tratamento de dados. No entanto, estes procedimentos devem ser utilizados adequadamente para não remover informação útil da variação de interesse.

Segundo Cinar *et al.*, (2005), os *outliers* (erros grosseiros anômalos) em dados e o ruído do sinal podem ter uma forte influência sobre os valores dos parâmetros e a estrutura dos modelos desenvolvidos. Consequentemente, a detecção de *outliers* e o pré-tratamento de dados usando técnicas de reconciliação e suavização de dados são críticos para o desenvolvimento de modelos precisos.

Entre os procedimentos de pré-tratamento de dados temos os seguintes:

a) Redução de ruído nos sinais

O ruído é um sinal não desejável que varia muito rapidamente (França, 2007), o nível de ruído nos sinais monitorados se transforma em variação dos dados não relacionada com respeito à saída desejada. Desse modo, os dados, além de informação do processo, podem incluir ruídos procedentes das medições.

O propósito da redução de ruído do sinal é separar o ruído da informação do processo. Segundo Cinar *et al.*, (2005) para esse propósito podem ser utilizadas diversas técnicas, tais como:

- 1) Filtração de ruído utilizando modelos de série de tempo, entre eles filtros que incluem termos com média móvel MA (*moving average*), e filtros chamados de autoregressivos com média móvel ARMA (*autoregressive moving average*). O

resultado gráfico é a suavização (*smoothing*) das variações não lineares, identificando a variação representativa que contém a informação de interesse, com o que se consegue eliminar o ruído aleatório (*denoising*).

- 2) Redução de ruído utilizando Análise de Componentes Principais -PCA (*Principal Component Analysis*), neste caso, analisam-se as relações entre todas as variáveis medidas utilizando todo o conjunto de dados. Os componentes principais são um novo conjunto de coordenadas que são ortogonais umas às outras, os primeiros componentes principais (as primeiras coordenadas) recolhem os dados com informação do processo significativa, os outros componentes principais (as outras coordenadas) contêm a maior parte do ruído da medição aleatório. Neste caso, reconstruindo a informação das medições utilizando só os componentes principais com informação do processo, o ruído das medições pode ser separado.
- 3) Decomposição de sinais, para isso podem ser utilizadas as diversas técnicas explicadas na literatura de processamento de sinais (ORFANIDIS, 1995; HAYES, 1996; REIS *et al.*, 2009).

b) *Detecção de erros grosseiros ou outliers.*

Segundo Prata (2009b), os erros grosseiros são aqueles originários de eventos não aleatórios, tendo pouca ou nenhuma ligação com o valor medido. Podem estar relacionados com mau funcionamento de instrumentos ou falhas no processo (por exemplo, vazamentos). Os *outliers* são uma classe de erros grosseiros, valores espúrios que correspondem a algum comportamento anormal, resultante de distúrbios não medidos.

Na apresentação gráfica dos dados, os *outliers* são pontos isolados que podem ser detectados facilmente pela inspeção visual. Bagajewicz (2000) assinala que as tarefas centrais são: identificar a existência, localização e o tipo de erro grosseiro; e determinar o tamanho do erro grosseiro. Após identificar os erros

grosseiros, duas respostas são desejadas: eliminar a medição que contém o erro grosseiro, ou, corrigir a medição ou o modelo.

Segundo (FREITAS, 2009), os dados operacionais podem ser filtrados empregando um filtro de ordem zero que consiste basicamente em substituir toda medida considerada inválida (contraditória ou nula) pela medida anterior. Prata (2009a) usou um filtro de ordem zero para obter valores de frações de polímero extraíveis em xileno quando os valores reais não estavam disponíveis porque os valores determinados *off line* em laboratório foram obtidos com bastante tempo de atraso, em um estudo de reconciliação de dados e estimação de parâmetros em tempo real em polimerizações de polipropileno.

c) Reconciliação empírica de dados

Assumindo-se como ponto inicial para a modelagem empírica a não preexistência de um modelo teórico do processo. O objetivo da reconciliação empírica de dados é converter os dados contaminados do processo em informação consistente sobre o comportamento do processo, utilizando como fundamentos para resolver contradições, o conhecimento prático da operação da planta, dados históricos de funcionamento consignados em relatórios e a experiência dos especialistas no processo. Sendo a primeira etapa da reconciliação de dados, a detecção, identificação e eliminação dos *outliers*.

d) Caracterização estatística básica dos dados.

Realizada a seleção e o pré-tratamento dos dados, podem-se realizar os cálculos de médias, de desvio padrão, da variância e da correlação dos dados obtidos. A análise da associação entre variáveis pode ser realizada por correlação cruzada. Para se calcular o índice de correlação, normalmente utiliza-se o coeficiente de correlação linear r de Pearson, no entanto, um baixo valor desse índice não elimina a hipótese de que as variáveis podem estar correlacionadas de forma não linear. A correlação é dita ser não linear se a taxa de variação entre duas variáveis relacionadas não é constante. Outro fator determinante é a

variabilidade. Uma baixa variabilidade pode gerar uma baixa correlação, se comparada a uma série com alta variabilidade.

e) Caracterização dinâmica

Neste caso, são empregadas as ferramentas estatísticas para caracterização dinâmica dos dados operacionais. O objetivo é a identificação de períodos de operação em estado estacionário ou transiente. Para tal propósito são realizados cálculos de médias e de variâncias móveis e de seus respectivos intervalos de confiança, e também cálculos de espectros de variância, que consistem no cálculo de variâncias ao longo de todo um horizonte de amostragens, utilizando a estratégia de janelas móveis (FREITAS, 2009).

4.2.3. SELEÇÃO DA ESTRUTURA DO MODELO

A etapa crucial da identificação de sistemas não lineares é a seleção da estrutura dos modelos. Este passo implica postular modelos, os quais são considerados como possíveis candidatos e selecionar o que se considera o mais adequado para os propósitos desejados. Os modelos podem ser: lineares, não lineares, de estado estacionário, de resposta dinâmica, de tempo contínuo, de tempo discreto, de funções de transferência, em espaço de estados, SISO (*Single Input – Single Output*), SIMO (*Single Input – Multiple Output*), MISO (*Multiple Input – Single Output*), MIMO (*Multiple Input – Multiple Output*) ou outros. Considerações mais amplas podem ser encontradas no apêndice 2.B.

4.2.4. ESTIMAÇÃO DOS PARÂMETROS DA ESTRUTURA SELECIONADA

A estimação dos parâmetros desconhecidos consiste em encontrar os valores dos parâmetros que permitam o melhor ajuste do modelo aos dados experimentais.

Modelos não lineares sobreparametrizados conseguem prever satisfatoriamente pontos experimentais utilizados para o processo de estimação de parâmetros, no entanto tendem a apresentar erros de predição quando o

modelo se utiliza para prever respostas em outras regiões experimentais. Nestos casos, o erro paramétrico se transfere para o erro de predição e o modelo apresenta regimes dinâmicos não apresentados pelo sistema original (AGUIRRE e BILLINGS, 1995; RODRIGUES, 1996). Considerações mais amplas podem ser encontradas no apêndice 3.A.

4.2.5. VALIDAÇÃO DO MODELO

Para a validação do modelo usualmente se utilizam dados adicionais do processo, porque um bom teste para conhecer a qualidade do modelo é comparar as saídas do modelo com os valores dos dados experimentais de um conjunto de dados que não foi usado para ajustar os parâmetros do modelo.

4.3. MODELOS NÃO LINEARES PARA IDENTIFICAÇÃO DE PROCESOS

Esse tipo de modelo é obtido utilizando medições de entradas e saídas do sistema real e procedimentos de identificação de sistemas não lineares, os modelos resultantes são chamados de modelos empíricos ou modelos de caixa-preta.

Na literatura algumas representações utilizadas para controle e modelagem não linear são:

- a) Modelos polinomiais NARMAX
- b) Modelos de Volterra
- c) Modelos de Wiener- Hamerstein
- d) “Wavelets”
- e) Modelos empíricos construídos com técnicas de inteligência artificial
 - e1 Modelos com Redes Neurais Artificiais
 - e2. Modelos FUZZY
 - e3. Modelos NEUROFUZZY

No presente trabalho, foi proposto o desenvolvimento de modelos empíricos construídos com técnicas de inteligência artificial, especificamente, Redes Neurais Artificiais, a explicação a seguir, segue esta orientação, no entanto uma

abordagem mais detalhada dos outros modelos não lineares utilizados para identificação de procesos se encontra no apêndice 3.B.

4.3.5. MODELOS EMPÍRICOS COM TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL

4.3.5.1. MODELOS EMPÍRICOS COM REDES NEURAIS ARTIFICIAIS

A ideia básica em modelagem empírica de processos utilizando redes neurais consiste em utilizar uma estrutura de rede para identificar as relações existentes entre variáveis independentes e dependentes de um processo. Para que a rede identifique as relações é necessário treiná-la com as informações dessas variáveis. Assim, o uso de uma rede neural para a modelagem de um processo dispensa a necessidade de informação sobre sua fenomenologia.

As redes neurais são uma poderosa ferramenta para modelagem de sistemas não lineares devido à sua propriedade de aproximação universal (NASCIMENTO e YONEYAMA, 2008). As redes neurais mais utilizadas em identificação de sistemas não - lineares são as do tipo *feedforward* (CLAUMANN, 2003; RANKOVIC e NIKOLIC, 2008).

O *perceptron* multicamada é uma rede neural *feedforward* cuja forma compacta para o caso MISO, é dada pela equação:

$$F(x, w) = \psi \left(\sum_k w_{ok} \phi \left(\sum_j w_{kj} \phi \left(\dots f \left(\sum_i w_{li} x_i \right) \right) \right) \right) \quad (4.16)$$

Onde $f(\cdot)$, $\phi(\cdot)$, $\phi(\cdot)$, $\psi(\cdot)$; são funções de ativação, w são vetores de pesos sinápticos de conexões entre neurônios de camadas próximas, e x_i é o i -ésimo elemento do vetor de entrada x .

O esquema de funções não lineares desse aproximador universal, descrito pela equação (4.16), não é usual na teoria clássica da aproximação de funções (HAYKIN, 2007).

A modelagem de processos baseada em redes neurais como identificação de sistemas pode ser descrita da seguinte forma. Um processo pode ser considerado como um sistema de múltiplas entradas e múltiplas saídas (MIMO, multiple input-múltiple output) com vetor de entrada x e vetor de saída y . Considere que o sistema tem um mapeamento de entrada-saída não linear descrito pela relação funcional:

$$y = f(x). \quad (4.17)$$

Onde os valores das entradas e as saídas são conhecidos e a função f é desconhecida.

A partir dos valores de entradas e saídas se pode formar um conjunto de exemplos rotulados:

$$E = \{(x_i, y_i)\}_{i=1}^N \quad (4.18)$$

O conjunto de exemplos pode ser utilizado para treinar uma rede neural como um modelo do sistema. Assim y_i é a saída desejada para o vetor de entrada x_i . Suponha que $\hat{y}_i$ represente a saída da rede em resposta ao vetor de entrada x_i :

$$\hat{y}_i = f_{NN}(x_i, w_i). \quad (4.19)$$

A diferença entre y_i e $\hat{y}_i$ fornece um sinal de erro ε_i , que é utilizado para ajustar os parâmetros livres da rede w_i , de forma a minimizar a diferença entre a saída da rede $\hat{y}_i$ e a saída do sistema desconhecido y_i , calculada sobre o conjunto de treinamento inteiro.

A rede neural identifica o sistema desconhecido quando:

$$\|f_{NN}(x, w_i) - f(x)\| < \varepsilon \text{ para todo } x. \quad (4.20)$$

Onde o erro aproximativo ε é um número positivo suficientemente pequeno para a tarefa. A Figura 4.1 mostra o processo de identificação de sistemas utilizando redes neurais (MORENO-ARMENDARIZ, et al., 2001; GIL e PÁEZ, 2007).

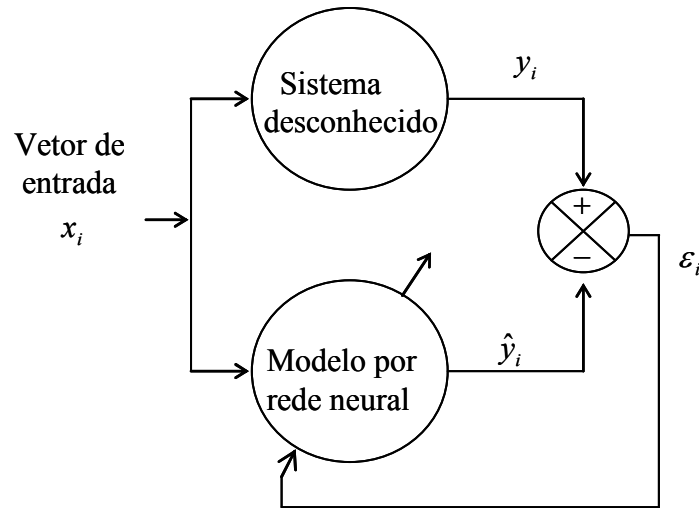


Figura 4.1. Identificação de sistemas utilizando redes neurais.

Para o desenvolvimento de um modelo baseado em redes neurais são importantes as seguintes etapas:

1. Definição do problema;
2. Coleta dos dados de treinamento e de teste;
3. Pré e pós-processamento dos dados;
4. Definição da Topologia da rede;
5. Treinamento;
6. Teste e validação.

4.3.5.1.1. Definição do problema

Antes de iniciar as tarefas da modelagem é preciso fixar os propósitos a conseguir com o modelo de rede neural. Algumas perguntas que deveriam ser respondidas nesta etapa são as seguintes:

- 1) Para que você pretende utilizar o modelo?

- 2) Quão simples ou complexo deveria ser o modelo?
- 3) Que aspectos do processo considera-se que são os mais relevantes, que deveriam estar no modelo?
- 4) Quais são as variáveis de entrada e saída do modelo?
- 5) São disponíveis dados empíricos sobre as variáveis consideradas no modelo?
- 6) Como vai ser testado o desempenho do modelo?
- 7) De quanto tempo se dispõe para construir o modelo?

4.3.5.1.2. Coleta dos dados de treinamento e de teste

Após o pré-processamento dos dados e após verificar se ainda existem inconsistências nos dados operacionais devido a medições incorretas, mau funcionamento do equipamento de processo ou outros casos, o total de dados selecionados deve ser distribuído em grupos que serão utilizados para treinamento e avaliação do desempenho do modelo.

Na literatura são mencionados diversos procedimentos de distribuição de dados, úteis para aplicar a técnica de validação cruzada, que são mais amplamente explicados no item 4.3.5.1.6.

a) *Divisão dos dados em dois grupos.*

Um conjunto de treinamento, é utilizado para treinar a rede e um conjunto de teste, é usado para validar o modelo.

b) *Divisão dos dados em três conjuntos:* treinamento, validação e teste (HAYKIN, 2007).

c) *Divisão de dados exaustiva* (*exhaustive data splitting*; ARLOT, 2010).

Dividem-se os dados utilizando todas ou uma parte das $\binom{n}{n_v}$ combinações que podem ser obtidas quando o total de dados se divide em dois grupos.

Onde: n são os dados disponíveis (pares entrada-saída), n_v é o conjunto de pares de entrada-saída utilizado para a validação do modelo e $n_t = n - n_v$ é o

conjunto de pares entrada-saída utilizado para treinar a rede e determinar seus parâmetros.

d) Divisão de dados parcial (*partial data splitting*; ARLOT, 2010)

Considerando que $\binom{n}{n_v}$ conjuntos de treinamento podem ser computacionalmente intratáveis, vários esquemas de divisão parcial de dados foram propostas como alternativas.

4.3.5.1.3. Pré e Pós-processamento de dados

A fim de eliminar a interferência das diferentes magnitudes das variáveis no processo de aprendizagem da rede, todos os dados são pré-processados (normalizados), de maneira que os valores das entradas e as saídas se encontrem no mesmo intervalo. Após a validação do modelo, os valores são pós-processados para retorná-los às grandezas originais.

Os procedimentos de pre e pós-processamento dos dados são diversos e os autores devem usá-los de acordo com o problema em questão, por exemplo, o intervalo de 0 a 1 para entradas e saídas pode ser usado por compatibilidade com a função de ativação sigmóide. Outros autores consideram que como os valores de 0 e 1 são valores infinitos, para a função sigmoide é recomendável diminuir o intervalo de valores para o intervalo [0,1 ; 0,9] ou ainda a [0,2 ; 0,8], para facilitar a convergência durante o treinamento da rede.

A expressão linear geral para adimensionalizar ou normalizar as entradas de um vetor x , em um mesmo intervalo dado por $[L_{inf}, L_{sup}]$ é:

$$x_n(i) = \left[\frac{x_{máx} - x(i)}{x_{máx} - x_{\min}} \right] L_{\inf} + \left[\frac{x(i) - x_{\min}}{x_{máx} - x_{\min}} \right] L_{\sup} \quad (4.21)$$

A expressão linear geral para retornar o vetor de saídas normalizadas y_n do modelo, após validação, às grandezas originais é:

$$y(i) = \left[\frac{L_{máx} - y_n(i)}{L_{máx} - L_{inf}} \right] y_{\min} + \left[\frac{y_n(i) - L_{inf}}{L_{máx} - L_{inf}} \right] y_{máx} \quad (4.22)$$

A adimensionalização logarítmica é obtida utilizando as seguintes equações:

$$y_n(i) = \frac{\log_{10}(y(i) - a)}{b} \quad (4.23)$$

Onde:

$$a = \frac{L_{\sup} [\log_{10}(y_{\min})] - L_{\inf} [\log_{10}(y_{máx})]}{L_{\sup} - L_{\inf}} \quad (4.24)$$

$$b = \frac{\log_{10}(y_{máx} / y_{\min})}{L_{\sup} - L_{\inf}} \quad (4.25)$$

A adimensionalização logarítmica se torna melhor que a linear quando no conjunto de dados de entrada o maior valor é muito superior ao menor valor (SILVA, 1997).

4.3.5.1.4. Definição da Topologia da rede

Uma das questões não resolvidas na literatura de redes neurais é qual topologia deveria ser usada para um problema dado. A seleção da topologia requer que seja escolhido o número apropriado de neurônios distribuídos em camadas e as suas conexões (ANDERS e KORN, 1996). No problema de aproximação de funções uma regra prática para obter uma boa generalização é usar a rede neural menor que possa ajustar os dados (REED, 1993), pois uma rede neural com tamanho mínimo é menos provável que aprenda também o ruído dos dados de treinamento (em problemas reais), e pode assim generalizar melhor com novos dados não considerados no treinamento da rede.

Durante o processo de determinação do melhor modelo de rede se propõem diversas topologias de rede, e, então, utilizando diversos procedimentos

e critérios, é selecionada a topologia ótima para o modelo de rede neural representativo do processo.

A topologia do modelo de rede neural pode ser obtida por aplicação dos seguintes métodos:

1. Parada antecipada.
2. Validação cruzada.
3. Procedimentos estatísticos
4. Construção de redes
5. Poda de redes (*pruning*)
6. Uso de algoritmos evolutivos

Os métodos 1 e 2 são explicados mais para frente no item 4.3.5.1.6, referido a validação e teste do modelo baseado em redes neurais, e um detalhamento dos métodos 3, 4, 5 e 6 pode ser encontrado no apêndice 3.C.

4.3.5.1.5. Treinamento da rede neural

No processo de aprendizagem, um mapeamento de entrada-saída é codificado em pesos sinápticos e limiares de um *perceptron* multicamadas, e a partir dessa perspectiva, o processo de aprendizagem na etapa de treinamento equivale a uma escolha dos parâmetros da rede para este conjunto de dados (HAYKIN, 2007).

Na etapa de treinamento experimenta-se com: algoritmos de treinamento, taxas de aprendizado, momentum, tempo de treinamento; para determinar os melhores parâmetros da rede.

As etapas: definição da topologia, treinamento da rede neural e validação do modelo de rede neural, estão estreitamente interligados, um excesso de treinamento para uma topologia determinada conduz a uma pobre capacidade do modelo para generalizar com dados não considerados no treinamento. A tendência normal é que o erro de treinamento diminua quando o número de neurônios ocultos é aumentado ou quando mais iterações de treinamento são

usadas. Em ambos os casos, pode resultar o problema de sobretraining ou *overtraining*, pelo qual o ajuste com dados de treinamento torna-se quase perfeito, mas o erro de generalização torna-se pior.

As redes neurais mais utilizadas em identificação de sistemas não lineares são as do tipo *feedforward* e grande parte desse sucesso pode ser atribuído ao algoritmo de treinamento supervisionado e iterativo conhecido por retropropagação do erro (*backpropagation*), relatado por Rumelhart, Hinton and Williams em 1986 (CLAUMANN, 2003; RANKOVIC e NIKOLIC, 2008).

4.3.5.1.5.1. Métodos para melhorar a velocidade de aprendizagem do algoritmo de retropropagação

No contexto das redes multicamada *feedforward* na literatura, algumas das desvantagens atribuídas ao algoritmo de *backpropagation* são:

- 1) O algoritmo tem uma restrição computacional, a computação realizada pelo neurônio é influenciada apenas por aqueles neurônios que estão em contato físico com ele (HAYKIN, 2007).
- 2) Uma das maiores desvantagens do algoritmo de *backpropagation* é a sua lenta velocidade de convergência (CHO e KIM, 1993).
- 3) Em muitas aplicações, o número de interconexões ou pesos em uma rede neural é tão grande que o tempo de aprendizagem para o algoritmo de retropropagação convencional pode se tornar excessivamente longo (JOHANSSON, *et al.*, 1991).

Uma variedade de métodos foi aplicada às redes neurais para melhorar a velocidade de aprendizagem do algoritmo de *backpropagation*. Lahmiri (2011) identifica duas categorias principais. A primeira categoria usa técnicas heurísticas desenvolvidas com base em uma análise do desempenho do método da descida mais íngreme (*steepest descent*) padrão, nesta categoria estão variedades do

algoritmo do gradiente descendente com taxa de aprendizagem adaptativa e/ou momentum. A segunda categoria usa técnicas de otimização numéricas padrão, nesta categoria estão incluídos os seguintes métodos:

- Os métodos de gradiente conjugado
- O método de Levenberg-Marquardt
- Métodos quase-Newton

Dentre esses métodos no presente trabalho, foi utilizado o de Levenberg-Marquardt, o qual é explicado no seguinte item, uma descrição mais detalhada dos outros métodos pode ser encontrada no apêndice 3.D.

4.3.5.1.5.1.1. O algoritmo Levenberg- Marquardt para o treinamento da rede

Utilizando-se o algoritmo de retropropagação do erro com algumas modificações é possível utilizar um algoritmo de minimização de mínimos quadrados para o treinamento de redes neurais, denominado de Levenberg-Marquardt (HAGAN e MENHAJ, 1994).

Quando a função que é minimizada tem a forma da Equação 4.67, o problema de otimização é chamado de minimização de mínimos quadrados não linear (*nonlinear least squares minimization*):

$$I = f(w) = \frac{1}{2} \sum_{i=1}^N e_i^2(w) \quad (4.67)$$

O índice de desempenho I pode ser definido como a soma do quadrado do erro e_i , associado ao i -ésimo padrão de treinamento da rede.

A propriedade distintiva dos problemas de mínimos quadrados é que, dada a matriz Jacobiana, J , a Hessiana, H , é:

$$H \approx \nabla^2 f(w) = J(w)^T J(w) \quad (4.68)$$

A regra de atualização dos pesos do algoritmo de Levenverg-Marquardt pode ser explicada como resultado de uma modificação feita no método de Newton da seguinte forma:

Um modo intuitivo simples para encontrar o mínimo de uma função é modificar os parâmetros utilizando o gradiente da função:

$$w_{i+1} = w_i - \lambda \nabla f \quad (4.69)$$

A atualização dos parâmetros depende do gradiente da função, o gradiente simples tem vários problemas de convergência (RANGANATHAN, 2004). O uso do método de Newton para resolver a equação $\nabla f(w) = 0$ permite utilizar informação da curvatura, bem como do gradiente. Expandindo-se o gradiente de f usando uma série de Taylor em torno de w_0 tem-se:

$$\nabla f(w) = \nabla f(w_0) + (w - w_0)^T \nabla^2 f(w_0) + \dots \quad (4.70)$$

Assumindo que f é quadrático em torno de w_0 e resolvendo para o mínimo de w , tem-se a regra de atualização para o método de Newton:

$$w_{i+1} = w_i - (\nabla^2 f(w_i))^{-1} \nabla f(w_i) \quad (4.71)$$

Onde w_0 foi substituído por w_i e w por w_{i+1} .

O método de Newton usa a suposição de que f é quadrático, nesse caso a Hessiana H não precisa ser avaliada exatamente, pode ser utilizada uma aproximação, Equação 4.68. Então, o método converge rapidamente, mas a velocidade de convergência é sensível ao ponto inicial (RANGANATHAN, 2004).

Levenberg propôs um algoritmo baseado nessa observação, considerando a Hessiana e a matriz identidade na regra de atualização:

$$w_{i+1} = w_i - (H + \lambda I)^{-1} \nabla f(w_i) \quad (4.72)$$

Onde H , é a matriz Hessiana avaliada em w_i .

A desvantagem é que, quando o valor de λ é grande, a matriz Hessiana calculada não é utilizada. Uma vez que, a Hessiana é proporcional à curvatura, é possível aproveitar melhor a sua contribuição substituindo a matriz identidade com a diagonal da matriz Hessiana, esse discernimento crucial foi fornecido por Marquardt, resultando na regra de atualização de Levenberg-Marquardt :

$$w_{i+1} = w_i - (H + \lambda \text{diag}[H])^{-1} \nabla f(w_i) \quad (4.73)$$

Desta forma, é obtido um movimento grande na direção com baixa curvatura e um movimento maior na direção com alta curvatura, e o problema clássico de “erro do vale” não ocorre mais.

O algoritmo de Levenberg-Marquardt parece ser o método mais rápido para o treinamento de redes neurais de tamanho moderado (até algumas centenas de parâmetros livres na rede) (FIORIN *et al.*, 2011).

4.3.5.1.5.2. A regularização bayesiana

O objetivo de treinar uma rede neural para identificação de processos é conseguir que o seu desempenho com dados novos seja tão bom como o mostrado com dados de treinamento. Quando esse desempenho desejado é alcançado dizemos que a rede generaliza bem e, portanto, pode representar apropriadamente a verdadeira função subjacente que relaciona as entradas e saídas de um processo.

A regularização Bayesiana é uma técnica proposta por MacKay (1992), a qual para melhorar a generalização restringe o tamanho dos pesos da rede. A ideia é que com pesos de tamanhos pequenos se consegue uma maior suavidade nas respostas da rede, e se pode representar melhor a verdadeira função subjacente, a qual tem um grau de suavidade no seu comportamento (FORESEE e HAGAN, 1997). Como resultado da aplicação desta técnica, se obtém uma rede com pesos de tamanho pequeno, com suavidade na resposta, e com muito boa

aproximação entre as respostas da rede e os valores alvo utilizados, não precisando por tanto do uso de conjuntos de teste.

Na técnica de regularização bayesiana a função objetivo é:

$$F = \alpha E_W + \beta E_D \quad (4.74)$$

Onde:

α e β são os parâmetros da função objetivo.

E_D é a função objetivo típica utilizada para treinar redes do tipo MLP sem regularização, que é a soma dos erros quadráticos dada pela equação:

$$E_D = \frac{1}{n} \sum_{i=1}^n (t_i - \hat{y}_i)^2 \quad (4.75)$$

$\hat{y}_i$ representa a resposta da rede neural.

E_W é a soma dos quadrados dos pesos e *bias* dada por:

$$E_W = \frac{1}{np} \sum_{i=1}^{np} w_i^2 \quad (4.76)$$

np é o número de parâmetros da rede (pesos e *bias*).

MacKay (1992) define as três funções básicas de probabilidade seguintes:

1) A função de verosimilhança (*likelihood*), que é a probabilidade de que as saídas alvo, t^n , ocorram no conjunto de dados D, dados os pesos W, entradas x^n , de acordo com a distribuição de probabilidade:

$$P(D|x^n, W, \beta, M) = \frac{\exp[-\beta E_D]}{Z_D(\beta)} \quad (4.77)$$

2) A probabilidade prévia (*prior probability*) é dada para pesos sinápticos de rede w alternativos de acordo com:

$$P(w|\alpha, M) = \frac{\exp[-\alpha E_w]}{Z_w(\alpha)} \quad (4.78)$$

3) A probabilidade posterior (*posterior probability*) dos pesos sinápticos w é:

$$P(w|D, \alpha, \beta, M) = \frac{\exp[-\alpha E_w - \beta E_D]}{Z_F(\alpha, \beta)} = \frac{\exp[-F]}{Z_F(\alpha, \beta)} \quad (4.79)$$

Onde:

$$Z_D = (2\pi / \beta)^{N/2} \text{ e } Z_w = (2\pi / \alpha)^{N/2} \quad (4.80)$$

Na abordagem Bayesiana, os pesos ótimos deveriam maximizar a probabilidade posterior $P(w | D, \alpha, \beta, M)$. Maximizar a probabilidade posterior é equivalente a minimizar a função objetivo regularizada: $F = \alpha E_w + \beta E_D$

Os parâmetros α e β são determinados pela regra de Bayes:

$$P(\alpha, \beta | D, M) = \frac{P(D | \alpha, \beta, M) P(\alpha, \beta | M)}{P(D | M)} \quad (4.81)$$

Nesta equação, a quantidade de interesse para atribuir probabilidades para (α, β) é $P(D | \alpha, \beta, M)$ e se pode encontrar uma expressão para esta função de probabilidade utilizando a regra de Bayes:

$$P(w | D, \alpha, \beta, M) = \frac{P(D | w, \beta, M) P(w | \alpha, M)}{P(D | \alpha, \beta, M)} \quad (4.82)$$

Resolvendo para $P(D | \alpha, \beta, M)$ e substituindo as probabilidades conhecidas obtém-se:

$$P(D | \alpha, \beta, M) = \frac{\left[\frac{1}{Z_D(\beta)} \exp(-\beta E_D) \right] \left[\frac{1}{Z_w(\alpha)} \exp(-\alpha E_w) \right]}{\frac{1}{Z_D(\beta)} \exp(-\beta E_D - \alpha E_w)} \quad (4.83)$$

$$P(D | \alpha, \beta, M) = \frac{Z_F(\alpha, \beta)}{Z_w(\alpha) Z_D(\beta)} \quad (4.84)$$

Nesta equação não se conhece $Z_F(\alpha, \beta)$, que é estimada supondo que F tem um mínimo simples como função de w , em w_{MP} , e assumindo que a aproximação a F no mínimo é quadrática. Z_F é aproximada então por (MACKAY 1992):

$$Z_F \approx e^{-F(w_{MP})} (2\pi)^{k/2} (\det(H_{MP})^{-1})^{1/2} \quad (4.85)$$

Onde k é a dimensão do vetor de pesos w , $H = \beta \nabla^2 E_D + \alpha \nabla^2 E_w$ é a matriz Hessiana da função objetivo.

Colocando estes resultados na Equação (4.84), obtem-se os valores ótimos para α e β no ponto mínimo.

As equações resultantes dadas por Foresee e Hagan (1997) são:

$$\alpha^{MP} = \frac{\gamma}{2E_w(w_{MP})} \quad (4.86)$$

$$\beta^{MP} = \frac{n - \gamma}{2E_D(w_{MP})} \quad (4.87)$$

Onde $\gamma = N - 2\alpha_{MP} \text{tr}(H_{MP})^{-1}$ é chamado o número efetivo de parâmetros e é uma medida de quantos parâmetros na rede são efetivamente utilizados para diminuir a função de erro. N é o número total de parâmetros na rede neural.

A otimização dos parâmetros α e β requer o cálculo da Hessiana, para facilitar a sua computação Foresee e Hagan (1997) propõem o uso da aproximação de Gaus-Newton para a matriz Hessiana assinalando que a sobrecarga computacional adicional é mínima quando utilizada a regularização Bayesiana com o algoritmo de otimização de Levenberg-Marquardt para localizar os pesos ótimos.

Segundo Foresee and Hagan (1997), após treinamento devem ser feitas as seguintes verificações:

- 1) Se $\gamma \approx N$, então a rede pode não ser suficientemente grande para representar adequadamente a função verdadeira.

- 2) Seja γ_1 o número de parâmetros efetivamente utilizados da rede neural R_1 com número de neurónios n_1^o (da qual se duvida que seja suficientemente grande) e γ_2 o número de parâmetros efetivamente utilizados de uma rede neural R_2 com número de neurónios n_2^o maior do que n_1^o :
- a) Se $\gamma_1 = \gamma_2$, então n_1^o é suficientemente grande.
 - b) Se $\gamma_1 \approx \gamma_2$, $E_{D1} \approx E_{D2}$, $E_{w1} \approx E_{w2}$, então n_1^o é suficientemente grande (E_{D1} e E_{w1} são valores da função objetivo para a rede R_1 e E_{D2} e E_{w2} são valores da função objetivo para a rede R_2).

Dias *et al.* (2003) fizeram um estudo comparativo de resultados entre o uso das técnicas de regularização Bayesiana e *Early Stopping* para evitar o problema do *overtraining*, e reportam que a técnica de regularização teve um melhor desempenho do que a técnica *Early Stopping*, proporcionando modelos de rede mais consistentes. No entanto afirmam que os modelos obtidos com regularização são mais complexos do que os obtidos com *Early stopping*.

Segundo Demuth and Beale (2002), no treinamento de redes neurais para aproximação de funções, a regularização Bayesiana fornece à rede melhor desempenho em generalização do que *Early stopping*, porque usa todos os dados, não precisa de um conjunto de dados de teste separado do conjunto de dados de treinamento. A vantagem é mais observável quando a quantidade de dados é pequena.

4.3.5.1.6. Validação e teste do modelo baseado em redes neurais

4.3.5.1.6.1. A validação cruzada (cross validation)

Treinar um modelo com um conjunto de dados e testar o seu desempenho estatístico com os mesmos dados gera um resultado excessivamente otimista. A razão para isto é que no treinamento, os parâmetros do modelo são otimizados para refletir as peculiaridades do conjunto de dados utilizados.

A validação cruzada é uma ferramenta padrão em estatística que fornece procedimentos para obter uma boa estimativa do desempenho estatístico do modelo, testando a sua saída com um conjunto de dados novos não utilizados no treinamento (HAYKIN, 2007).

4.3.5.1.6.2. Variantes da validação cruzada

A) Divisão dos dados em dois grupos

O *hold out method* é o tipo mais simples de validação cruzada, o conjunto de dados é dividido em um subconjunto de treinamento e um subconjunto de validação que é utilizado para testar o desempenho do modelo (ARLOT, 2010). O cômputo é simples, no entanto os resultados podem ser significativamente diferentes dependendo de como foi feita a divisão.

B) Divisão dos dados em três conjuntos

Segundo Haykin (2007) para a validação cruzada, primeiro os dados são aleatoriamente divididos em conjunto de treinamento e conjunto de teste (*test set*), o conjunto de dados de treinamento é ainda dividido em dois subconjuntos: um subconjunto de estimação usado para selecionar o modelo e um subconjunto de validação usado para validar o modelo com um conjunto de dados diferente do usado para estimar os parâmetros. No entanto, o *overfitting* ainda pode estar presente no melhor modelo validado. Para evitar o *overfitting*, a capacidade de generalizar do modelo selecionado é medida com o conjunto de teste, o qual é diferente do subconjunto de validação.

C) Divisão de dados exaustiva (*exhaustive data splitting*)

C1) *Leave-one-out*

Para um conjunto de n integrantes se realizam n experimentos, em cada um deles o conjunto de validação tem um par entrada-saída e o conjunto de treinamento com $n-1$ pares entrada-saída, que se usam como exemplos de treinamento. O número de conjuntos de treinamento é n (ARLOT, 2010).

C2) *Leave-p-out*

Realizam-se vários experimentos, o conjunto de n pares entrada-saída é dividido em dois grupos em cada experimento: Conjunto de validação com n_v pares entrada-saída e conjunto de treinamento com $n - n_v$ pares entrada-saída que se usam como exemplos de treinamento (ARLOT, 2010).

O número de conjuntos de treinamento é: $T = \binom{n}{n_v}$ (4.88)

C3) *Divisão de dados parcial (partial data splitting)*

Considerando que $\binom{n}{n_v}$ conjuntos de treinamento podem ser computacionalmente intratáveis, esquemas de divisão parcial de dados foram propostas como alternativas.

C4) *Validação cruzada V-fold.*

Foi introduzida por Geisser (1975) como alternativa ao computacionalmente caro *Leave-one-out*.

Dividem-se os n pares entrada-saída em V subconjuntos de aproximadamente igual cardinalidade n/V , cada subconjunto sucessivamente desempenha o papel de subconjunto de validação cruzada. O conjunto de treinamento tem $n - n/V$ pares de entrada-saída.

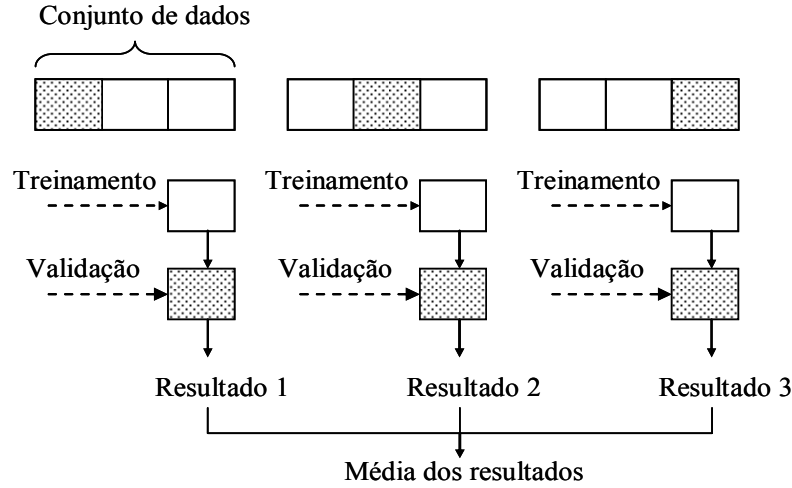


Figura 4.2. Método de V-fold para Validação cruzada, com V=3.

C5) Validação cruzada incompleta balanceada (BICV).

Foi proposta por Shao (1993), neste método se aplicam projetos de blocos incompletos balanceados (BIB) para distribuir o total de dados n em grupos de treinamento e validação. O BICV pode ser visto como uma alternativa ao método de validação cruzada V-fold quando o tamanho do conjunto de treinamento n_t é pequeno (ARLOT, 2010). Neste caso, o conjunto de validação tem n_v pares de entrada-saída:

$$n_v = n - n_t \quad (4.89)$$

Para a aplicação do desenho BIB se utilizam os parâmetros:

$a = n$ pares entrada-saída.

$k = n_v$ pares entrada-saída por bloque.

Então, o número de blocos é:

$$b = \binom{n}{n_v} \quad (4.90)$$

Atribui-se uma diferente combinação de pares entrada-saída para cada bloco.

O número de repetições para cada par entrada-saída é:

$$r = \binom{n-1}{n_v-1} \quad (4.91)$$

O número de vezes que cada par entrada-saída ocorre no mesmo bloco é:

$$\lambda = \binom{n-2}{n_v-2} = \frac{r(k-1)}{a-1} \quad (4.92)$$

O parâmetro λ deve ser um número inteiro (MONTGOMERY, 2001).

No BIB, dois pares entrada-saída ocorrem juntos o mesmo número de vezes em relação a os outros pares.

I) *Aprendizagem-teste repetido (Repeated learning-testing)*

Foi introduzida por Breiman *et al.* (1984) . Neste método os n pares entrada-saída são divididos repetidas vezes aleatoriamente em um conjunto de treinamento n_t de tamanho $n(1-p)$ e um conjunto de teste n_v de tamanho np , donde $0 < p < 1$, $n = n_t + n_v$, tipicamente $n_t \geq n_v$ (BURMAN, 1989). Se $p=1/3$ cada conjunto de teste é de tamanho $n/3$ e cada conjunto de treinamento é de tamanho $2n/3$. Para cada divisão são obtidas estimativas baseadas nos dados do conjunto de treinamento que depois são testadas com os dados do conjunto de teste.

4.3.5.1.6.3. A validação cruzada para seleção de modelos de redes neurais

A validação cruzada é uma importante ferramenta para seleção de modelos, a escolha do melhor modelo de rede neural é baseada na sua capacidade de generalização com dados novos contidos no conjunto utilizado para testar o desempenho do modelo.

Segundo Raghavarao e Padgett (2005), a seleção do melhor modelo está baseada no cálculo da média ao quadrado dos erros de previsão:

$$\bar{E} = \frac{(Y - \hat{Y})'(Y - \hat{Y})}{n_v} \quad (4.93)$$

Onde: Y é um vetor das n_v respostas que correspondem ao conjunto de validação. $\hat{Y}$ é o vetor de respostas estimadas para o conjunto de validação baseadas no modelo obtido com os n_t pares entrada-saída utilizados para treinar a rede e determinar seus parâmetros.

$\bar{E}$ é calculado para todas ou uma parte das $\binom{n}{n_v}$ combinações que podem ser obtidas quando o total de dados se divide em dois grupos. O modelo selecionado possui o menor valor na soma das médias ao quadrado dos erros de previsão resultantes.

Quando se utiliza a distribuição de dados em três conjuntos segundo o esquema sugerido por Haykin (2007) :

$$n = (n_e + n_v)_t + n_{tt} \quad (4.94)$$

Onde : n_{tt} é o número de pares entrada-saída do conjunto de teste.

$(n_e + n_v)_t$ é o número de pares entrada-saída do conjunto de treinamento divididos em n_e pares entrada-saída para o conjunto de estimação e n_v pares entrada saída para o conjunto de validação.

Na Equação 4.93, Y é um vetor das n_{tt} respostas que correspondem ao conjunto de teste.

4.3.5.1.6.4. A validação cruzada e parada antecipada (*Early stopping*)

A parada antecipada é uma técnica utilizada para evitar o sobre-treinamento (*overtraining*) na qual se determina o ponto de parada do treinamento por observação do erro produzido em predições com integrantes do grupo de validação cruzada.

O procedimento mais simples para aplicar o *Early stopping* é dividir o conjunto de dados em dois subconjuntos, um deles para treinamento e o outro para validação cruzada. O subconjunto de treinamento é usado para modificar os

pesos, e o subconjunto de validação é usado para estimar a capacidade de generalização do modelo (REED, 1993). Quando o treinamento progride, o erro deve estar diminuindo tanto no conjunto de treinamento como no conjunto de validação. Com maior treinamento o aumento do erro no subconjunto de validação em lugar da diminuição é uma indicação da existencia de *overtraining*, portanto, o ponto onde é observado o erro mínimo é o ponto de parada do treinamento. Segundo Haykin (2007), o que a rede aprende além do ponto de erro mínimo é essencialmente ruído contido no conjunto de treinamento.

No entanto, a situação real é muito mais complexa. As curvas de generalização reais obtidas com o subconjunto de validação quase sempre têm mais de um mínimo local. Prechelt (1998a) reporta um caso com 16 mínimos locais, e afirma que é impossível, em geral, dizer a partir do início da curva se o mínimo global já foi visto ou não, ou seja, se um aumento no erro de generalização indica *overfitting* real ou é apenas intermitente.

Por tanto, a aplicação do *Early stopping* nos casos de mínimos múltiplos leva a um gasto de tempo de treinamento significativo para garantir bons resultados em relação à capacidade de generalização do modelo. Nestes casos com o uso de só dois subconjuntos: treinamento e validação; existe o risco de que ainda o *overfitting* possa estar presente no modelo validado, se o treinamento foi parado em um mínimo local e não no mínimo global.

A divisão dos dados em três subconjuntos: treinamento, validação e teste; para aplicar *Early stopping* é outra possibilidade mencionada na literatura. Prechelt (1998b) utilizou o seguinte procedimento para medir o desempenho da rede:

- 1) Divisão do conjunto de dados em duas partes distintas: dados de treinamento e dados de teste.
- 2) Os dados de treinamento foram subdivididos em um conjunto de exemplos de treinamento utilizados para ajustar os pesos da rede e um conjunto de exemplos de validação usados para estimar o desempenho da rede durante o treinamento, conforme exigido pelos critérios de parada antecipada.

- 3) Os dados de teste foram usados para estimar o desempenho da rede após a conclusão do treinamento.

Segundo Kearns (1996) com dois grupos: treinamento e validação cruzada, uma porcentagem de aproximadamente 10% do conjunto total de dados é adequada para o subconjunto de validação cruzada. Haykin (2005) utiliza como referência estes resultados e indica que uma escolha apropriada é distribuir 20% do conjunto de treinamento para o subconjunto de validação. Nesse caso, o conjunto de treinamento divide-se em subconjunto de estimação e subconjunto de validação, e utiliza-se outro conjunto de teste diferente desse de validação para medir a capacidade de generalização do modelo.

Bose and Liang (1996) assinalam que 10% do conjunto de treinamento pode ser utilizado como subconjunto de validação.

Yahya, *et al* (2010) utilizaram *Early Stopping* com três subconjuntos: subconjunto de treinamento (60% dos dados), subconjunto de validação (20% dos dados) e subconjunto de teste (20% dos dados).

4.4. CONCLUSÕES

Neste capítulo são apresentadas as etapas necessárias para construir modelos empíricos por identificação de processos, estas etapas se iniciam com a definição do problema de identificação de processos, onde é importante ter um entendimento preciso dos objetivos e da complexidade do modelo a construir. Essa etapa direciona as seguintes etapas que são: a coleta e preparação dos dados, a seleção da estrutura do modelo, a estimação dos parâmetros da estrutura selecionada, e a validação do modelo.

Na seção referente aos modelos não lineares que podem ser úteis para resolver o problema de identificação de processos é apresentada uma ampla variedade de alternativas, com maior detalhe nos modelos de redes neurais do tipo *Feedforward* pela sua utilidade no desenvolvimento do presente trabalho.

Foram abordados os métodos para melhorar a velocidade de aprendizagem e os procedimentos para validação e teste dos modelos baseado em redes neurais. A informação do estado da arte apresentada neste capítulo e complementada nos apêndices 3.A.,3.B.,3.C. e 3.D., permite fazer uma boa escolha de procedimentos para modelagem empírica segundo a necessidade dos usuários. No entanto, existem questões para serem resolvidas com experiência na aplicação dos procedimentos, relacionadas com o número de neurônios, o número de épocas de treinamento e os melhores critérios para avaliar o desempenho dos modelos.

4.5. REFERÊNCIAS BIBLIOGRÁFICAS

1. Admuthe, L., Apte, S. and Amuthe, S. (2009). Topology and parameter optimization of ANN using genetic algorithm for application of textiles. IEEE international Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS 2009). Rende (Cosenza), Italy.
2. Aghbashlo, M., Kianmehr, M.H., Nazghelichi, T., and Rafiee, S. (2011). Optimization of an Artificial Neural Network Topology for Predicting Drying Kinetics of Carrot Cubes Using Combined Response Surface and Genetic Algorithm. *Drying Technology*, v. 29 (7), pp. 770-779.
3. Aguirre, L.A. and Billings, S.A. (1995). Dynamical effects of overparametrization in nonlinear models. *Physica D.*, v. 80, pp. 26-40.
4. Aguirre, L.A., Rodrigues, G.G. and Jácome, C.R.F. (1998). Identificação de Sistemas não lineares utilizando modelos NARMAX polinomiais-Uma revisão novos resultados. *SBA Controle & Automação*, v. 9 (2), pp. 90-106
5. AL Seyab, R.K.S. (2006). **Nonlinear Model Predictive Control Using Automatic Differentiation**. PhD Thesis, CRANFIELD University, UK.
6. Anders, U. and Korn, O. (1999). Model selection in neural networks. *Neural Networks*, v. 12(2), pp. 309-323.
7. Angelov, P.P. and Filev, D.P. (2004). An Approach to Online Identification of Takagi-Sugeno Fuzzy Models. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 34 (1), pp. 484-498.

8. Arifovic, J. and Gençay, R. (2001). Using genetic algorithms to select architecture of a feedforward artificial neural network. *Physica A: Statistical Mechanics and its Applications*, v. 289 (3-4), pp. 574-594.
9. Arlot, S. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, v. 4, pp. 40-79.
10. Attik, M, Bougrain, L. and Alexandre, F. Optimal brain surgeon variants for feature selection, in *International Joint Conference on Neural Networks (IJCNN)*, 2004.
11. Attik, M., Bougrain, L. and Alexandre, F. (2005). Neural Network Topology Optimization. In W.Duch, J. Kacprzyk, E. Oja, S. Zadrozny (Eds.), *Artificial Neural Networks Formal Models and their Applications- ICANN 2005, LNCS 3697*, pp. 53-58. Springer-Verlag, Berlin, Germany, 2005.
12. Babaie-Kafaki, S., Ghandari, R. and Mahdavi-Amiri, N. (2010). Two new conjugate gradient methods based on modified secant equations. *Journal of Computational and Applied Mathematics*, v. 234 (2010), pp. 1374-1386.
13. Babuška, R. and Verbruggen, H. (2003). Neuro-fuzzy methods for nonlinear system identification. *Annual Reviews in Control*, v. 27 (2003), pp. 73–85.
14. Bagajewicz, M. J. (2000). A Brief Review of Recent Developments in data Reconciliation and gross Error Detection/Estimation”, *Latin American Applied Research*, v. 30, pp. 335-342.
15. Beale, E.M. (1972). A derivation of conjugate-gradients. *Numerical Methods for Non-linear Optimization*. F.A. Lootsma, Ed. Academic Press, pp. 39-43.
16. Billings, S.A., Chen, S. and Korenberg, M.J. (1989). Identification of MIMO nonlinear systems using a forward-regression orthogonal estimator. *International Journal of Control*, v. 49(6), pp.2157-2189.
17. Bose, N.K. and Liang, P. (1996). **Neural network fundamentals with graphs, algorithms, and applications**. Mcgraw-Hill, Inc. USA.
18. Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). **Classification and regression trees**. Wadsworth Statistics/Probability Series. Wadsworth Advanced Books and Software, Belmont, CA.
19. Burman, P. (1989). A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika*, 76 (3), pp. 503-514.

20. Cervantes, A.L., Agamenoni, O.E. and Figueroa, J.L. (2003). Robust Identification of PWL-Wiener Models: Use in Model Predictive Control. *Latin American Applied Research*, v.33 (4). pp.1- 8
21. Chiu, S.L. (1997). Extracting Fuzzy Rules from Data for Function Approximation and Pattern Classification. In: D. Dubois, H. Prade and R.Yager (Eds.). *Fuzzy Information Engineering: A Guided Tour of Applications*. Jhon Wiley & Sons.
22. Cho, S.B. and Kim, J.H. (1993). Rapid Backpropagation Learning Algorithms. *Circuits Systems Signal Process*, v. 12 (2), pp. 155-175.
23. Cinar, A., Parulekar, S.J., Undey, C., BiroL, G. (2005) **Batch Fermentation, Modeling, Monitoring and Control**. Markel Dekker: New York, NY, USA.
24. Claumann, C.A. (2003). **Desenvolvimento e aplicações de redes neurais wavelets e da teoria de regularização na modelagem de processos**. PhD Thesis. Universidade de Santa Catarina, Florianópolis, Brasil.
25. Coca, D. and Billings, S.A. (1997). Continuous-time system identification for linear and nonlinear systems using wavelet decompositions. *International Journal of Bifurcation and Chaos*, v. 7 (1), pp. 87-96.
26. Conforth, M. and Meng, Y. (2008). IEEE Swarm Intelligence Symposium St. Louis MO, USA.
27. Crama, P. and Schoukens, J. (2002). Hammerstein - Wiener System Estimator Initialization. *Proceedings of ISMA2002. International Conference on Noise and Vibration Engineering*, Leuven, Belgium, v. III, pp. 1169-1176.
28. Dai, Y. H. and Liao, L. Z.(2001). New conjugacy conditions and related nonlinear conjugate gradient methods, *Applied Mathematics & Optimization*, v. 43, pp. 87-101.
29. Dai, Y.H.,Liao, L.Z. and Li D.(2004). On restart procedures form the conjugate gradient method. *Numerical algorithms*, v. 35, pp. 249-260.
30. Demuth, H. and Beale, M. 2002. **Neural Network Toolbox for Use with. MATLAB**. Natick, MA, USA: The MathWorks,Inc.
31. Dennis, J.E. and Schnabel, R.B. **Numerical methods for unconstrained optimization and nonlinear equations**. Prentice Hall, Englewood Cliffs, NJ, 1983.

32. Dias, F.M., Antunes, A. and Mota, A.M. (2003). Regularization versus Early Stopping: a case study with a real system. 2nd IFAC Conference Control Systems Design (CSD'03), Bratislava, República Eslovaca.
33. Dorigo, M., Maniezzo, V., and Coloni, A. Ant System: optimization by a colony of cooperating agents, *IEEE Transactions on Systems, Man and Cybernetics - Part B*, v. 26 (1), pp. 29-41, 1996.
34. Efe., Ö. and Kaynak, O. (2000). A Comparative Study of Soft Computing Methodologies in Identification of Robotic manipulators. *Robotics and Autonomous Systems*, v. 30 (3), pp. 221-230,
35. Fahlman, S. E.; Lebiere, C. (1990). The cascade-correlation learning architecture. In: D. S. Touretzky (Ed.). *Advances in NIPS*, v. 2, pp. 524–532, Morgan Kaufmann, San Mateo, CA, USA.
36. Fiorin, D. V., Martins, F.R., Schuch, N. J. and Pereira, E. B. (2011). Forecast of solar energy resource by using neural network methods. *Rev. Bras. Ensino Fís.*, v. 33 (1), pp. 1-36.
37. Fletcher, R. and Reeves, C. M. (1964). Function minimization by conjugate gradients, *Computer Journal*, v. 7, pp. 149–154.
38. Fogel, D.B. (1995) *Evolutionary computation: Toward a New Philosophy of machine Intelligence*. IEEE Press. Piscataway, NJ.
39. Foresee, F.D. and Hagan, M. T. (1997). Gauss - Newton Approximation to Bayesian Learning. *Proceedings of the 1997 International Joint Conference on Neural Networks*, v. 3, 1930–1935.
40. França, F. A. (2007). *Instrumentação e Medidas: grandezas mecânicas*, UNICAMP. SP. Brasil. Disponível em:
< http://www.fem.unicamp.br/~instmed/Instrumentacao_Medidas_Grandezas_Mecanicas.pdf >
Acesso em: 12 outubro, 2011.
41. Freitas, K.P. **Ferramentas estatísticas para monitoramento de modelos inferenciais**. Dissertação de Mestrado UFRJ, Brasil, 2009.
42. Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association*, v. 70, pp. 320–328.
43. Ghomsheh, V.S., Shoorehdeli, M. and Teshnehlab, M. (2007). Training ANFIS structure with modified PSO algorithm. *Proceedings of the 16th Mediterranean Conference on Control & Automation - 2007*, Athens, Greece.

44. Gil, R.V. & Páez, D.G.(2007). Identificación de sistemas dinámicos utilizando redes neuronales RBF. *Revista Iberoamericana de Automática e Informática Industrial*, v. 4 (2), pp. 32-42.
45. Gilbert, J. C. and Nocedal, J. (1992). Global convergence properties of conjugate gradient methods for optimization, *SIAM Journal of Optimization*, v. 2(1), pp. 21–42. Rapport de Recherche nro. 1268. June 1990. INRIA, Project PROMATH.
46. Grippo, L. and Lucidi, S. (1997). A Globally Convergent Version of the Polak-Riviere Conjugate Gradient Method. *Mathematical Programming*, v. 8, pp. 375-391.
47. Guillaume, S. (2001). Designing fuzzy inference systems from data: An interpretability oriented review, *IEEE Transactions on Fuzzy Systems*, v. 9, pp. 426-443.
48. Haber, R. and Unbehauen, H. (1990). Structure Identification of Nonlinear Dynamic Systems- A Survey on Input/Output Approaches. *Automatica*, v. 26, 651-677.
49. Hagan, M. T. and M. B. Menhaj (1994). Training feedforward networks with the marquardt algorithm. *IEEE Transactions on Neural Networks* 5(6), pp. 989–993.
50. Hager, W. and Zhang, H. (2006). CG_Descent, a Conjugate Gradient Method with Guaranteed Descent. *ACM Transactions on Mathematical Software*, v. 32 (1), pp. 113-137.
51. Hagiwara, M. (1994). A simple and effective method for removal of hidden units and weights, *Neurocomputing*, v. 6 (2), pp. 207–218.
52. Hassibi, B., Stork, D. G., and Wolf, G. J. (1993). Optimal brain surgeon and general network pruning, *IEEE ICNN'93*, v. 1, pp. 293–299.
53. Hayes, M. **Statistical Digital Signal Processing and Modeling**: New York, John Wiley and Sons, 1996.
54. Haykin, S. (2005). **Neural Networks**. A Comprehensive Foundation. 2ª Edição. Delhi, India: Pearson Prentice Hall.
55. Haykin, S. (2007). **Redes neurais**. Princípios e prática. 2ª Edição. São Paulo, Brasil: ARTMED EDITORA S.A.

56. Hestenes, M. R. and Stiefel, E. (1952). Methods for conjugate gradients for solving linear systems, *Journal of Research of the National Bureau of Standards*, v. 49 (6), pp. 409–436.
57. Höpner, F., and Klawonn, F. (2004). Learning Fuzzy Systems – An Objective Function-Approach. *Mathware & Soft Computing Journal*, v. 11 (5), pp. 143-162.
58. Igrejas, G.P.P. (2008). **Identificação Difusa e Controlo Preditivo do Processo de Fermentação Alcoólica**. Tese de Mestrado, Universidade do Minho, Portugal.
59. Jager, R. **Fuzzy Logic in Control**. Doctor Thesis. Delft University of Technology. Delft, Netherlands. 1995.
60. Jang, J.S. R., and Mizutani, E. (1996). Levenberg-Marquardt Method for ANFIS Learning, NAFIPS, 1996. Biennial Conference of the North American Fuzzy Information Processing Society, pp. 87 –91.
61. Jang, J.S.R. (1993). ANFIS: adaptive-network-based fuzzy inference system: *IEEE Transactions on Systems, Man and Cybernetics*, 23(3): 665-685.
62. Jang, J.S.R. (1993). ANFIS: adaptive-network-based fuzzy inference system: *IEEE Transactions on Systems, Man, and Cybernetics*, v. 23 (3), pp. 665-685.
63. Jang, J.S.R., SUN, C.T. and MIZUTANI, E. **Neuro-Fuzzy and Soft Computing**. Prentice Hall, Englewood Cliffs, 1997.
64. Jin, Y. (2000) Fuzzy Modeling of High-Dimensional Systems: Complexity reduction and Interpretability Improvement, *IEEE Transactions on Fuzzy Systems*, v. 8(2), pp. 212-221.
65. Johansson, E.M., Dowla, F.U., and Goodman, D.M. (1991). Backpropagation learning for multilayer feed-forward neural networks using the conjugate gradient method. *International Journal of Neural Systems*, v. 2 (4), pp. 291-301.
66. Juditski, A., Zhang, Q., Delyon, B., Glorennec, P.Y., and Benveniste, A. (1994). *Wavelets in identification*. INRIA, RENNES Cedex, France.
67. Kearns, M. (1996). A Bound on the Error of Cross Validation Using the Approximation and Estimation Rates, with Consequences for Information Training-Test Split. In: D.S. Touretzky, M.C. Mozer and M.E. Hasselmo (Eds.) *Neural Processing 8*, Morgan Kaufmann, pp. 183 - 189.

68. Kennedy, J., and Eberhart, R. "Particle Swarm Optimization, " in: *Proc.IEEE Intl. Conf. on Neural Networks (Perth, Australia)*, IEEE Service Center, Piscataway, NJ, IV:1942-1948, 1995.
69. Kennedy, P. (2003). *A guide to econometrics*. The MIT Press. Cambridge, MA.USA.
70. Kosko, B. (1994). Fuzzy Systems as Universal Approximators. *IEEE Transactions on Computers* , v. 43 (11), pp. 1329- 1333.
71. Lahmiri, S. (2011). A Comparative Study of Backpropagation Algorithms in Financial Prediction. *International Journal of Computer Science, Engineering and Applications (IJCSEA)*, v.1 (4). DOI: 10.5121/ijcsea.2011.1402.
72. Lecun, Y., Denker, J., Solla, S., Howard, R. E., and Jackel, L. D., Optimal brain damage. In *Advances in NIPS*, edited by D. S. Touretzky, v. 2, Morgan Kauffman, San Mateo, CA, 1990.
73. LI, D.H. and Fukushima, M. (2001). A modified BFGS method and its global convergence in no convex minimization, *Journal of Computational and Applied Mathematics*, v.129 (2001), pp. 15–35.
74. Liu, D., and Xu, G. (2011). A Perry descent conjugate gradient method with restricted spectrum. Disponível em:
< http://www.optimization-online.org/DB_FILE/2011/03/2958.pdf >.
Acesso em: 10 dezembro 2011.
75. Liu, G.P. (2001). **Nonlinear identification and control**. A neural network approach. London, UK: Springer-Verlag Ltd.
76. Liu, J. and Wang, S. (2011). Modified nonlinear conjugate gradient method with sufficient descent condition for unconstrained optimization. *Journal of Inequalities and Applications*, 2011:57, pp.1-12.
77. Livieris, I.E., and Pintelas, P. (2012). An advanced conjugate gradient training algorithm based on a modified secant equation. *International Scholarly Research Network, ISRN Artificial Intelligence*, v. 2012. doi: 10.5402/2012/486361.
78. Ljung L. (1996). System identification. In W. S. Levine (Ed.). *The Control Handbook*, CRC Press Inc., 1996, pp. 1033-1054.
79. Lukšan, L., Matonoha, C. and Vlček, J. (2008). Computational experience with modified conjugate gradient methods for unconstrained optimization.

Technical report No. 1038. Academy of Sciences of the Czech Republic. Prague. Disponível em:
 < http://www.optimization-online.org/DB_FILE/2009/02/2240.pdf >
 Acesso em: 12 dezembro 2011.

80. Luo, Q., Yang, W. and Yi, D. (2008). Kernel shapes of fuzzy sets in fuzzy systems for function approximation. *Information Sciences: an International Journal*, v. 178 (3), pp. 836-857.
81. Mackay, D.J.C. (1992). A Practical Bayesian Framework for Backpropagation Networks. *Neural Computation*, v. 4(3), pp. 448-472.
82. Mamdani, E.H. (1974). Application of fuzzy algorithms for control of simple dynamic plant, *IEEE Proceedings Control & Science*, v. 121(12), pp. 1585-1588.
83. Mänle, M. (2001). Parameter optimization for Takagi-Sugeno Fuzzy Models-Lessons Learnt. *IEEE International Conference on Systems, Man, and Cybernetics*, v. 1, pp. 111-116.
84. Mascioli, F.M., Varazi, G.M. and Martinelli, G. (1997). Constructive Algorithm for Neuro-Fuzzy Networks”, Proceedings of the Sixth IEEE International Conference on Fuzzy Systems, 1997, Barcelona-Spain, pp. 459 –464.
85. Mizumoto, M. (1991). Min-Max-Gravity method versus Product-Sum-Gravity method for fuzzy controls, Proc.4th IFSA World Congress, Brussels/Belgium, v. Eng., 1991, pp.127-130.
86. Mizumoto, M. (1995). Improvements of Fuzzy Control Methods. In: Hua Harry Li; Gupta, Madan M. (Eds.). Fuzzy logic and intelligent systems. *International Series in Intelligent Technologies*, v. 3, pp. 1-16. Springer Netherlands.
87. Montgomery, D. C. **Design and Analysis of Experiments**. 5° Ed., John Wiley & Sons, New York. 2001.
88. Moreno Armendariz, M.A.; Poznyak, A.S. & Yu-Liu, W. (2001). Control adaptable indirecto usando redes neuronales dinámicas. *Revista del Centro de Investigación*. Universidad La Salle, México, v. 04(16), pp. 13-32.
89. Mozer, M.C. and Smolensky, P. (1989). Skeletonization: A Technique for Trimming the Fat From a Network Via Relevance Assessment. In: D.S. Touretzky (Ed.). *Advances in Neural Information Processing*, v. 1, pp. 107-116, San Mateo, CA, 1989. Morgan Kaufmann.

90. Narushima, Y., and Yabe, H.(2011). Conjugate gradient methods based on secant conditions that generate descent search directions for unconstrained optimization. Disponível em:
< http://www.optimization-online.org/DB_FILE/2011/09/3184.pdf >
Acesso em: 14 dezembro 2011.
91. Nascimento C.L. e Yoneyama, T. (2008). **Inteligência artificial em controle e automação**. Blücher-FAPESP. São Paulo, Brasil.
92. Nozaki, K., Ishibuchi, H. and Tanaka, H. (1997). A simple but powerful heuristic method for generating fuzzy rules from numerical data, *Fuzzy sets and Systems*. Vol .86, pp. 251-270.
93. Ogunnaike, B.A., Ray, W.H. (1994). **Process Dynamic Modeling and Control**. New York NY, USA : Oxford University Press. Oxford.
94. Orfanidis, S.J. Introduction to Signal Processing. Upper Saddle River, NJ, Prentice Hall, Inc., 1995.
95. Passino, K. M. and Yurkovich, S. *Fuzzy control*. Menlo Park: Addison Wesley Longman, 1998.
96. Patwardhan, R.S., Lakshminarayanan and SHAH, S.L. (1997). Nonlinear Model Predictive Control using PLS based Hammerstein and Wiener Models. AICHE Annual Meeting, LA. USA, pp. 1-15. Disponível em:
< http://www.ualberta.ca/dept/chemeng/control/publication/reports/slshah/aiche97_1.pdf >
Acesso em: 20 novembro 2011.
97. Pedrycz, W. and Gomide, F. **An introduction to Fuzzy Sets: Analysis and Design**. Cambridge, MA: MIT Press, 1998.
98. Perry, A. (1978). A modified conjugate gradient algorithm. *Operations Search*, v. 26 (6), pp.1073-1078.
99. Polak, E., and Ribière, G. (1969). Note sur la convergence de methods de directions conjuguées, *Revue Francais d'Informatique et de Recherche Operationnelle*, v. 16, pp. 35–43.

Powell, M.J.D. (1977). Restart procedures for the conjugate gradient method. *Mathematical Programming*, v. 12 , pp. 241-254.
101. Powell, M.J.D. (1984). Nonconvex minimization calculations and the conjugate gradient method. In *Lecture Notes in Mathematics 1066*, pp. 122-141, Berlin, 1984. Springer-Verlag.
102. Prata, D.M. (2009b). **Reconciliação robusta de dados para monitoramento em tempo real**. PhD Tese. COPPE/UFRJ. Brasil.

103. Prata, D.M., *et al.* (2009a). Non linear dynamic data reconciliation and parameter estimation through particle swarm optimization: Application for an industrial polypropylene reactor. *Chemical Engineering Science*, v. 64 (18), pp. 3953-3967.
104. Prechelt, L. (1998a). Automatic early stopping using cross validation: quantifying the Criteria. *Neural Networks*, v. 11, Issue 4, June 1998, pp. 761-767.
105. Prechelt, L. Early Stopping - but when?. In: G. Orr, K.R. Müller (Eds.). *Neural Networks: Tricks of the Trade, Lecture Notes in Computer Science*, v. 1524, Chapter 2, pp. 55-69, Springer-Verlag, 1998b.
106. Ragg, T., Braun, H. and Landsberg, H. (1996). A Comparative Study of Neural Network Optimization Techniques. *13th International Conference on Machine Learning (ICML96), Workshop on evolutionary computing and machine learning*, Bari, Italy.
107. Raghavarao, D. and Padgett, L. (2005). **Bloks designs, analysis, combinatorics and applications**. Singapore: World Scientific Publishing.
108. Ranganathan, A. (2004). The Levenberg-Marquardt Algorithm. Disponível em: <www.cc.gatech.edu/people/home/ananth/lmtut.pdf>. Acesso em: 18 novembro 2011.
109. Rankovic, V.M. and Nikolic, I.Z. (2008). Identification of Nonlinear Models with Feedforward Neural Network and Digital Recurrent Network. *FME Transactions*, v. 36 (2), pp. 87-92.
110. Reed, R. (1993). Pruning Algorithms - A Survey. *IEEE Transactions on Neural Networks*, v. 4 (5), pp. 740-747.
111. Reis, M.S., Saraiva, P.M. and Bakshi, B.R. (2009). Denoising and Signal-to-Noise Ratio Enhancement: Wavelet Transform and Fourier Transform. *Comprehensive Chemometrics — Chemical and Biochemical Data Analysis*, v.2, pp. 25-55.
112. Reitermanova, Z. (2008). Feedforward Neural Networks-Architecture Optimization and Knowledge Extraction. *WDS'08 Proceedings of Contributed Papers, Part I*, pp. 159-164.
113. Rodrigues, G.G. **Identificação de Sistemas Dinâmicos Não-Lineares Utilizando Modelos NARMAX Polinomiais** – Aplicação a Sistemas Reais. Dissertação de Mestrado. Universidade Federal de Minas Gerais. 1996.

114. Ross, T. J. **Fuzzy Logic with Engineering Applications**, 2nd ed. West Sussex, England: John Wiley & Sons, Ltd., 2004.
115. Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*. Vol 323, pp. 533–536.
116. Saha P., Krishnans. H., RAO V. S. and Patwardhan S. C., Modelling and predictive control of MIMO nonlinear systems using Wiener–Laguerre models, *Chemical Engineering Communications*, Taylor and Francis Group, v. 191, pp. 1083–1119.
117. Seborg, D.E., Edgar, T.F. and Mellichamp, D.A. **Process Dynamics and Control**. Hoboken, New Jersey, USA: Jhon Wiley & Sons, Inc., 2004.
118. Sentoni G. B., Agamennoni, O., Dessages, A., and Romagnoli, J. (1996). Approximate models for nonlinear process control, *AIChE Journal*, v. 42(8), pp. 2240–2250.
119. Shao, J. (1993). Linear model selection by cross-validation. *J. Amer. Statist. Assoc.*, 88(422): 486–494.
120. Silva, J.E.L. **Simulação e controle preditivo linear (com modelo de convolução) e não linear (com modelo baseado em redes neurais artificiais) de colunas recheadas de absorção com reação química**. Dissertação de Mestrado, Universidade Estadual de Campinas. 1997.
121. Simões, M.G. e Shaw, I.S. **Controle e modelagem fuzzy**. 2^a edição. São Paulo, Brasil: Blücher-FAPESP, 2007.
122. Stahlberger, A. and Riedmiller, M. Fast network pruning and feature extraction by using the unit-OBS algorithm. In: M.C. Mozer,. M.I. Jordan, and T. Petsche. (Eds.). *Advances in Neural Information Processing Systems*, v. 9, p.655. The MIT Press (1997).
123. Tahmasebi, P. and Hezarkhani, A. (2010). Comparison of Optimized Neural Network with Fuzzy Logic for Ore Grade Estimation. *Australian Journal of Basic & Applied Sciences*, v. 4(5): 764-772.
124. Takagi, T., and Sugeno, M. (1985), Fuzzy Identification of Systems and Its applications to Modeling and Control, *IEEE Transactions on Systems, Man, and Cybernetics*, v. 15(1), pp. 116-132.
125. Wei, H.L., Billings, S.A., Zhao, Y.F. and Guo, L.Z. (2010). An adaptative wavelet network for spatio-temporal system identification. *Neural Networks*, v. 23 (10), 1286-1299.

126. Wu, C., and Chen, G. (2010). New type of conjugate gradient algorithms for unconstrained optimization problems. *Journal of Systems Engineering and Electronics*, v. 21(6), pp. 1000-1007.
127. Yabe, H., and Takano, M. (2004). Global convergence properties of nonlinear conjugate gradient methods with modified secant condition, *Comput. Optim. Appl.*, v. 28, pp. 203 - 225.
128. Yager, R.R. and Filev, D.P. (1994). Generation of Fuzzy Rules by Mountain Clustering, *Journal of Intelligent & Fuzzy Systems*, v. 2, pp. 209-219.
129. Yahya, M.N., Otsuru, T., Tomiku, R. and Okozono, T. (2010) Investigation the capability of neural network in predicting reverberation time on classroom. *International Journal of Sustainable Construction Engineering & Technology*, v. 1 (1). pp. 1-13.
130. Yuan, G. (2009). Modified nonlinear conjugate gradient methods with sufficient descent property for large-scale optimization problems, *Optimization Letters*, v. 3 (1), pp. 11–21.
131. Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, v. 8 (3), pp. 338-353.
132. Zeng, X.J. and Singh, M.G. (1995). Approximation theory of fuzzy systems-MIMO case. *IEEE Transactions on Fuzzy Systems*, v. 3 (2), pp. 219-235.
133. Zhang, L. and Zhou, W. (2008). Two descent hybrid conjugate gradient methods for optimization, *Journal of Computational and Applied Mathematics*, v. 216 (1), pp. 251–264.
134. Zhang, L., Zhou, W., and LI, D. (2007). Some descent three-term conjugate gradient methods and their global convergence, *Optimization Methods and Software*, v. 22 (4), pp. 697–711.
135. Zhang, Q., and Benveniste, A. (1992). Wavelet Networks. *IEEE transactions on Neural Networks*, v. 3 (6), pp. 889-898.

CAPÍTULO 5. CONTROLE DE PROCESSOS BASEADO EM MODELO

5.1. INTRODUÇÃO

O controle de processos permite satisfazer várias necessidades que se apresentam na operação industrial dos processos, tais como: rejeitar as diferentes perturbações, que normalmente ocorrem no decorrer da operação do processo, garantir a qualidade do produto, melhorar a segurança operacional, reduzir o custo do produto através da manutenção de um desempenho ótimo do processo e reduzir o impacto ambiental (FISCHER,1991). Portanto, para atingir estes propósitos, é importante garantir a qualidade do projeto e desempenho dos sistemas de controle.

Seborg *et al.* (2004) consideram duas abordagens gerais para o projeto dos sistemas de controle. 1) Na abordagem tradicional, a estratégia de controle e o *hardware* do sistema de controle são selecionados com base em conhecimento do processo, experiência e intuição. 2) Na abordagem baseada em modelo, o primeiro passo é desenvolver o modelo do processo, o qual pode ser usado em métodos de projeto de controladores com base em modelos, ou pode ser incorporado na lei de controle, ou pode ser usado para avaliar estratégias de controle e determinar valores preliminares dos parâmetros dos controladores através de simulações no computador. Segundo estes autores, para processos complexos um modelo dinâmico de processo deveria ser desenvolvido, de modo que o sistema de controle possa ser apropriadamente projetado.

Na abordagem baseada em modelo, uma tecnologia muito importante pela sua aplicação na indústria de processos é o controle preditivo com modelo (MPC, do inglês *Model Predictive Control*), o qual é tratado a seguir.

5.2. CONTROLE PREDITIVO COM MODELO (MPC)

O Controle Preditivo com Modelo (MPC) não designa uma estratégia de controle específico, mas uma gama muito ampla de métodos de controle que fazem um uso explícito de um modelo de processo para obter o sinal de controle,

minimizando uma função objetivo (CAMACHO e BORDONS, 2004). Segundo Ongunnaïke e Ray (1994), o modelo de processo é útil para duas tarefas centrais:

- 1) Predição explícita do comportamento futuro da planta.
- 2) Cálculo de ação de controle corretiva apropriada requerida para a condução da saída predita o mais próximo possível do valor-alvo desejado.

A metodologia MPC inclui as técnicas: DMC, do inglês *Dynamic Matrix Control* (CUTLER e RAMAKER, 1980), MAC do inglês *Model Algorithmic Control* (RICHALET *et al.*, 1978), GPC do inglês *Generalized Predictive Control* (CLARKE *et al.*, 1987a,b), IMC do inglês *Internal Model Control* (GARCIA e MORARI, 1982 e 1985 a y b), QDMC do inglês *Quadratic Dynamic Matrix Control* (GARCIA, 1984; GARCIA e MORSHEDI, 1984 e 1986). SMPC do inglês *Simplified Model Predictive Control* (ARULANLAN e DESHPANDE, 1987), entre outras.

MPC é uma metodologia de controle avançado que tem tido grande impacto na prática industrial, com aplicações principalmente em refinarias de petróleo, plantas petroquímicas, plantas químicas e indústrias de papel e celulose. Nessas indústrias, o MPC tornou-se o método de escolha para problemas difíceis de controle multivariável que incluem restrições de desigualdade (ONGUNNAIKE e RAY, 1994; QUIN e BADGWELL, 2003). Entretanto, relativamente poucas aplicações têm sido reportadas em outras indústrias de processo, mesmo que MPC seja uma abordagem bastante geral e não limitada a um tipo de indústria (KWONG, 2005).

A Figura 5.1 mostra aspectos básicos do MPC. A saída atual y (variável controlada), a saída predita $\hat{y}$, o *set point* ou valor alvo w , a variável manipulada u , o horizonte de controle, H_c e o horizonte de predição, H_p . No instante de amostragem k , o controlador recebe informação sobre a saída atual do processo y_k , e calcula uma sequência de ações de controle futuras ($u_{k+1}, u_{k+2}, \dots, u_{k+H_c}$) para que a saída predita $\hat{y}$, se movimente ao *set point* w da melhor maneira possível. A primeira ação de controle é implementada e, no próximo instante de amostragem se repetem os cálculos de controle com a nova informação da saída do processo.

O horizonte de controle H_c é movido um intervalo adiante (horizonte móvel). Um modelo do processo é utilizado para o cálculo das saídas preditas $\hat{y}$ sobre um horizonte de predição H_p e um procedimento de otimização baseado na minimização de um índice de performance ou função objetivo é utilizado para o cálculo das próximas H_c ações de controle.

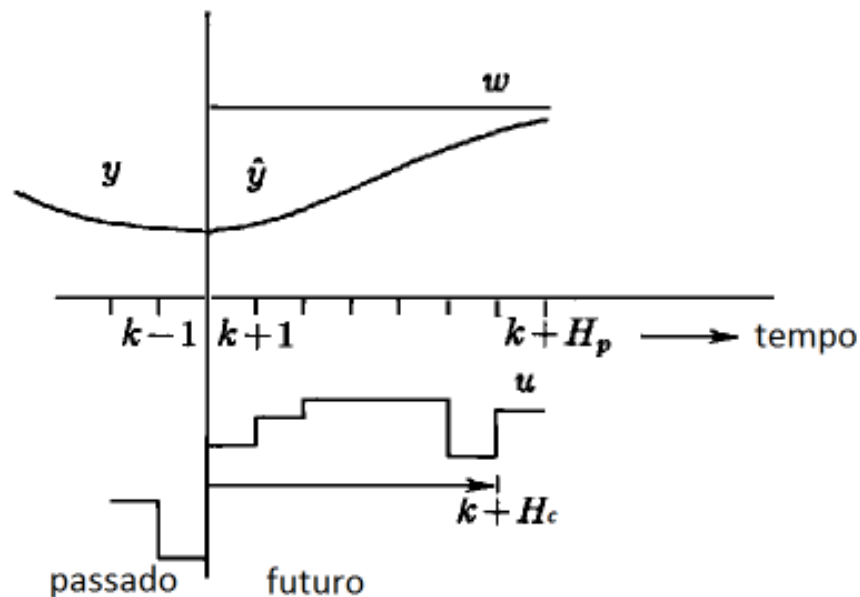


Figura 5.1. O conceito de controle preditivo (Soertemboek, 1990).

5.2.1. CARACTERÍSTICAS DOS CONTROLADORES PREDITIVOS

Segundo Soertemboek (1990) qualquer controlador preditivo tem quatro características:

- 1) Um modelo de predição do processo a ser controlado.
- 2) A trajetória de referência para a saída do processo.
- 3) O critério ou função objetivo que é minimizado, a fim de obter a sequência ótima da saída do controlador ao longo do horizonte de predição.
- 4) O procedimento de minimização em si mesmo.

5.2.1.1. MODELOS DE PREDIÇÃO

Segundo Ongunnaïke e Ray (1994), algum modelo apropriado é utilizado para a predição da saída do processo ao longo de um horizonte de predição pré-determinado na ausência de ações de controle adicionais. Para modelos discretos no tempo, isto significa prever a variável controlada y , i tempos de amostragem: $\hat{y}(k+1), \hat{y}(k+2), \dots, \hat{y}(k+i)$; tendo o tempo atual k como a origem da predição e baseado na ação de controle atual e em ações de controle passadas: $u(k), u(k-1), \dots, u(k-j)$.

Para Seborg (2004), o sucesso do MPC depende da precisão do modelo de processo, o modelo de processo pode ser um modelo físico (fenomenológico) ou um modelo empírico, linear ou não linear. Porém, tradicionalmente as aplicações industriais de MPC têm sido baseadas em modelos empíricos lineares discretos no tempo, na forma de equações de diferenças ou modelos de resposta ao degrau ou de resposta ao impulso.

No caso de modelos empíricos lineares, Ongunnaïke e Ray (1994) afirmam que os três mais conhecidos modelos para MPC convencional são as versões lineares e discretas de:

a) Modelos de convolução finita

Estes tipos de modelos têm duas formas inteiramente equivalentes:

a1) O modelo de resposta ao impulso

$$y_k = \sum_{i=0}^k h_i u_{k-i} \quad (5.1)$$

Onde h_i são os coeficientes de resposta ao impulso do processo. A saída do processo nos instantes futuros de amostragem é predita usando os coeficientes de resposta ao impulso h_i e entradas passadas u_{k-i} .

a2) O modelo de resposta ao degrau

$$y_k = \sum_{i=0}^k a_i \Delta u_{k-i} \quad (5.2)$$

Onde a_i são os coeficientes de resposta ao degrau do processo

$$\Delta u(k) = u(k) - u(k-1) \quad (5.3)$$

Os coeficientes de resposta ao impulso e de resposta ao degrau estão relacionados segundo as seguintes equações

$$h(i) = a(i) - a(i-1) \quad (5.4)$$

$$a_i = \sum_{j=1}^i h_j, \quad (5.5)$$

Para os sistemas causais reais $h(0)$ e $a(0)$ são iguais a zero.

b) Modelos de espaço de estado discretos

Para o caso linear estes modelos são também conhecidos como modelos ARMAX (do inglês *AutoRegressive, Moving Average with eXogenous inputs*), e podem ser representados pela equação:

$$y(k) = \sum_{i=0}^k \alpha_i y(k-i) + \sum_{i=0}^k \beta_i u(k-i-m) \quad (5.6)$$

Os parâmetros α_i e β_i e m são obtidos utilizando procedimentos de identificação de procesos, $\alpha_0 = \beta_0 = 0$.

c) Modelos de função de transferência discretos

Estes são modelos obtidos a partir de:

$$y(z) = \frac{z^{-m} B(z^{-1})}{A(z^{-1})} u(z) \quad (5.7)$$

Onde $A(z^{-1})$ e $B(z^{-1})$ são polinômios na variável transformada z . A equação inclui também um retraço associado de m tempos de amostragem.

No caso de modelos empíricos não lineares podem ser utilizados para MPC os modelos apresentados no capítulo anterior:

a) Modelos polinomiais NARMAX

b) Modelos de Volterra

- c) Modelos de Wiener- Hamerstein
- d) “Wavelets”
- e) Modelos construídos utilizando técnicas de inteligência artificial
- e1) Modelos com Redes Neurais Artificiais
- e2) Modelos FUZZY
- e3) Modelos Neuro-Fuzzy

5.2.1.2. A TRAJETORIA DE REFERÊNCIA PARA A SAÍDA DO PROCESSO

A trajetória de referência é a saída de processo desejada, pode ser uma sequência arbitrária de pontos (SOERTEMBOEK,1990) ou pode ser especificada por uma equação (KWONG, 2005; COSTA *et al*, 2002; MELEIRO *et al*, 2005), por exemplo:

$$y_{k+i} = \alpha y_k + (1 - \alpha)sp \quad (5.8)$$

Onde:

y_{k+i} são os valores desejados da variável controlada; $i=1,2,...,Hp$, $0 \leq \alpha < 1$, sp é o set point atual.

Para $\alpha = 0$, o valor desejado da variável controlada é igual ao set point atual:

$$y_{k+i} = sp \quad (5.9)$$

5.2.1.3. FUNÇÕES OBJETIVO

A escolha da função objetivo é importante porque em controladores preditivos, a minimização da função objetivo produz a lei de controle preditivo.

Segundo Soertemboek (1990), as funções objetivo podem ser: funções objetivo de passo simples e funções objetivo multipasso.

a) Funções objetivo de passo simples.

Uma função deste tipo utiliza só uma predição e pode estar baseada no erro calculado entre o valor da previsão feita pelo modelo e o setpoint, a equação é:

$$J = [\hat{y}(k+1) - sp_k]^2 \quad (5.10)$$

Onde $\hat{y}(k+1)$ é o valor da previsão, e sp_k é o valor do setpoint para o tempo $t = k$.

A função objetivo pode ser modificada tendo em conta incrementos nos valores da saída do controlador $\Delta u(k)$, a equação resultante é:

$$J = [\hat{y}(k+1) - sp_k]^2 + \lambda \Delta u(k)^2 \quad (5.11)$$

Onde λ é um o fator de ponderação utilizado para penalizar as mudanças na saída do controlador.

O segundo caso melhora o desempenho do controlador, que no primeiro caso pode ser oscilante.

b) Funções objetivo multipasso

Se o horizonte de previsão é maior do que um, uma função multipasso é necessária, então a função objetivo é:

$$J = \sum_{i=1}^{Hp} [\hat{y}(k+i) - sp_k]^2 + \lambda \Delta u(k+i-1)^2 \quad (5.12)$$

Onde H_p é o horizonte de previsão.

Uma restrição adicional é frequentemente tomada em conta: os incrementos no valor da saída do controlador são iguais a zero, $\Delta u(k+i+1) = 0$, no intervalo $[H_c, H_p]$, onde H_c é o horizonte de controle ($H_c < H_p$).

O conceito de horizonte de controle é utilizado nos controladores DMC e QDMC, para o caso SISO a função objetivo é (COSTA *et al.*, 2002):

$$J = \sum_{i=1}^{Hp} [\hat{y}_{k+i}^c - sp_{k+i}]^2 + \lambda \sum_{i=1}^{Hc} \Delta u(k+i-1)^2 \quad (5.13)$$

Onde: $\hat{y}_{k+i}^c$ é a previsão do modelo corrigida, sp são valores de setpoint, λ é o fator de supressão que penaliza as acções de controle (saídas do controlador), Δu são incrementos na variável manipulada, H_p é o horizonte de previsão, e H_c é o horizonte de controle.

Para o caso MIMO a função objetivo é (MELEIRO *et al.*, 2005):

$$J = \sum_{j=1}^{Ncv} \sum_{i=1}^{Hp} [\hat{y}_{j,k+i}^c - sp_{j,k+i}]^2 + \sum_{j=1}^{Nmv} \lambda \sum_{i=1}^{Hc} \Delta u_{j,k+i-1})^2 \quad (5.14)$$

Onde: $\hat{y}_{j,k+i}^c$ é a previsão do modelo corregida, Ncv é o número de variáveis controladas, Nmv é o número de variáveis manipuladas. O significado dos demais termos é similar que para o caso SISO.

Nos casos de aplicação do controlador QDMC, as funções objetivo são utilizadas com um conjunto de restrições que estão detalhadas nos trabalhos de: Costa *et al.*, (2002) e Meleiro *et al.*, (2005).

A maior desvantagem de usar fatores de ponderação é a sua dificuldade para estabelecer os seus valores com antecendência. A seguinte função não faz uso de fatores de ponderação:

$$J = [P\hat{y}(k+1) - Rsp_k]^2 + Qu(k)^2 \quad (5.15)$$

Onde P, Q e R são polinômios. Segundo Soertemboek (1990), embora resultem mais parâmetros para sintonizar, é fácil de obter um comportamento de malha fechada desejado. Esta função objetivo é usada no controlador de variância mínima generalizada (GMV).

5.2.1.4. PROCEDIMENTOS DE MINIMIZAÇÃO

Se a função objetivo é quadrática, o modelo de predição é linear e não existem restrições para a saída do controlador, existe uma solução analítica. Por exemplo, Luyben (1990) mostra que no caso do controlador DMC clássico, os melhores valores para as saídas do controlador (ações de controle) que fazem com que a variável controlada coincida com a trajetória de referência é obtido com o procedimento de mínimos quadrados. Para o caso SISO, utiliza a função objetivo J dada por:

$$J = \sum_{i=1}^{Hp} \left[x_i - \sum_{k=1}^{Hc} a_{ik} m_k \right]^2 + f^2 \sum_{i=1}^{NC} [m_k]^2 \quad (5.16)$$

Onde: $m = (\Delta u)^{new}$, é o vetor que fornece os Hc valores futuros de Δu (valores de mudanças futuras na variável manipulada) que minimizam J .

A função J é otimizada com relação ao vetor m de modo que:

$$\frac{\partial J}{\partial m} = 0 \quad (5.17)$$

O resultado da derivação escrito na forma matricial é:

$$[A^t A + f^2 I] m = A^t x \quad (5.18)$$

Onde I é a matriz identidade, A é denominada de matriz dinâmica, e x é o vetor de valores medidos da variável controlada. A solução para o vetor m é:

$$m = [A^t A + f^2 I]^{-1} A^t x \quad (5.19)$$

Se houver restrições, o critério deve ser minimizado usando um método de otimização não linear. Segundo Soertemboek (1990), a minimização da função objetivo tendo em conta as restrições, resulta em melhor desempenho do sistema, devido ao fato de que o efeito das restrições na saída prevista do processo é tido em conta.

O problema de otimização da função objetivo quadrática J com as restrições lineares constitui um problema de programação quadrática. Existem vários métodos para resolver esse problema, os quais podem ser bastante complexos e muito demorados. Em controladores preditivos, a otimização tem sido realizada utilizando métodos de programação quadrática (QP). A formulação geral do problema é:

$$\min_x \phi(x) = \frac{1}{2} x^t Q x + c^t x \quad (5.20)$$

Sujeito a:

$$Ax \geq b \quad (5.21)$$

$$x \geq 0 \quad (5.22)$$

A solução é encontrada através da obtenção da função Lagrangiana associada ao problema e baseia-se na aplicação das condições de Karush-Kuhn-Tucker (KKT) necessárias (HIMMELBLAU, 2001; TAHA, 2004).

O método de programação quadrática é utilizado no controlador QDMC (GARCIA e MORSHEDI, 1984 e 1986).

Alternativamente aos métodos convencionais, outros métodos considerados não convencionais como: Algoritmos genéticos e Simulated Annealing podem resolver o problema de otimização, no entanto, por serem muito demorados dificultam a sua utilização nos algoritmos iterativos dos sistemas de controle de processos.

5.3. CONTROLE DE MATRIZ DINÂMICA (DMC)

O DMC foi desenvolvido por Cutler and Ramaker (1980). Segundo Quin and Badwell (2003) é um dos modelos com grande impacto no controle de processos industriais que serviu para definir o paradigma industrial do MPC, suas principais características são:

- 1) Modelo de convolução linear para a planta;
- 2) Objetivo de desempenho quadrático sobre um horizonte de previsão finito.
- 3) Comportamento futuro da saída da planta especificado tentando seguir o (setpoint), ponto de ajuste tão perto quanto possível.
- 4) Entradas ótimas calculadas como solução para um problema de mínimos quadrados.

5.3.1. MODELO DE CONVOLUÇÃO SISO

O controle por matriz dinâmica (DMC) utiliza modelos de resposta ao degrau chamados de modelos de convolução. A Figura 5.2 mostra a resposta de um sistema em malha aberta frente a uma perturbação degrau em uma variável manipulada. Os valores da variável de resposta (variável controlada) frente ao degrau unitário são expressos por: $a_1, a_2, \dots, a_T$, usando um período de amostragem T_s . O produto $N \cdot T_s$ pode ser considerado como tempo de resposta do sistema em malha aberta, tempo necessário para que a resposta em malha aberta atinga 99% do valor do estado estacionário, onde N é chamado de Horizonte do modelo.

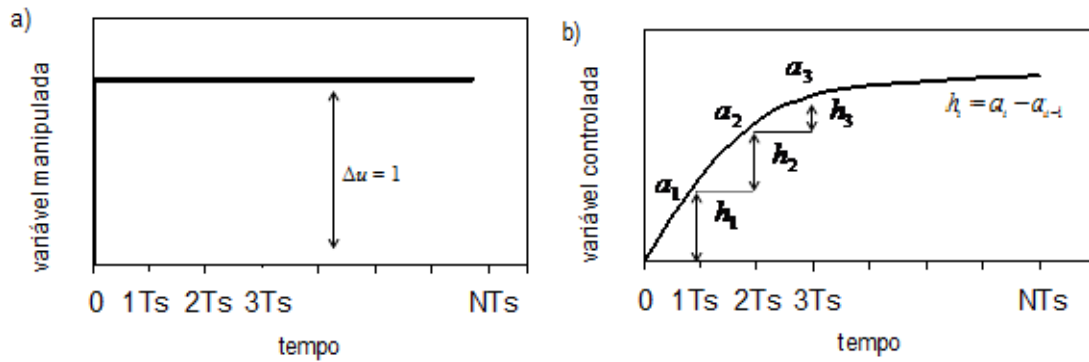


Figura 5.2. Resposta de um sistema em malha aberta frente a uma perturbação degrau em uma variável manipulada.

Para o desenvolvimento a seguir define-se $\hat{c}_k$, como o valor predito da variável controlada, e u_k o valor da variável manipulada no k -ésimo instante de amostragem. Define-se também c_k como o valor atual da variável controlada, então, $\hat{c}_k = c_k$, se não houver erros de modelagem nem perturbações no processo. Ambas as variáveis manipulada e controlada são expressas como variáveis desvio. Denotando-se $\Delta u_k = u_k - u_{k-1}$.

A Figura 5.3, mostra a resposta dinâmica do sistema quando são aplicadas três perturbações degrau na variável manipulada Δu_1 , Δu_2 e Δu_3 , nos tempos de amostragem $k = 0$, $k = Ts$ e $k = 2Ts$.

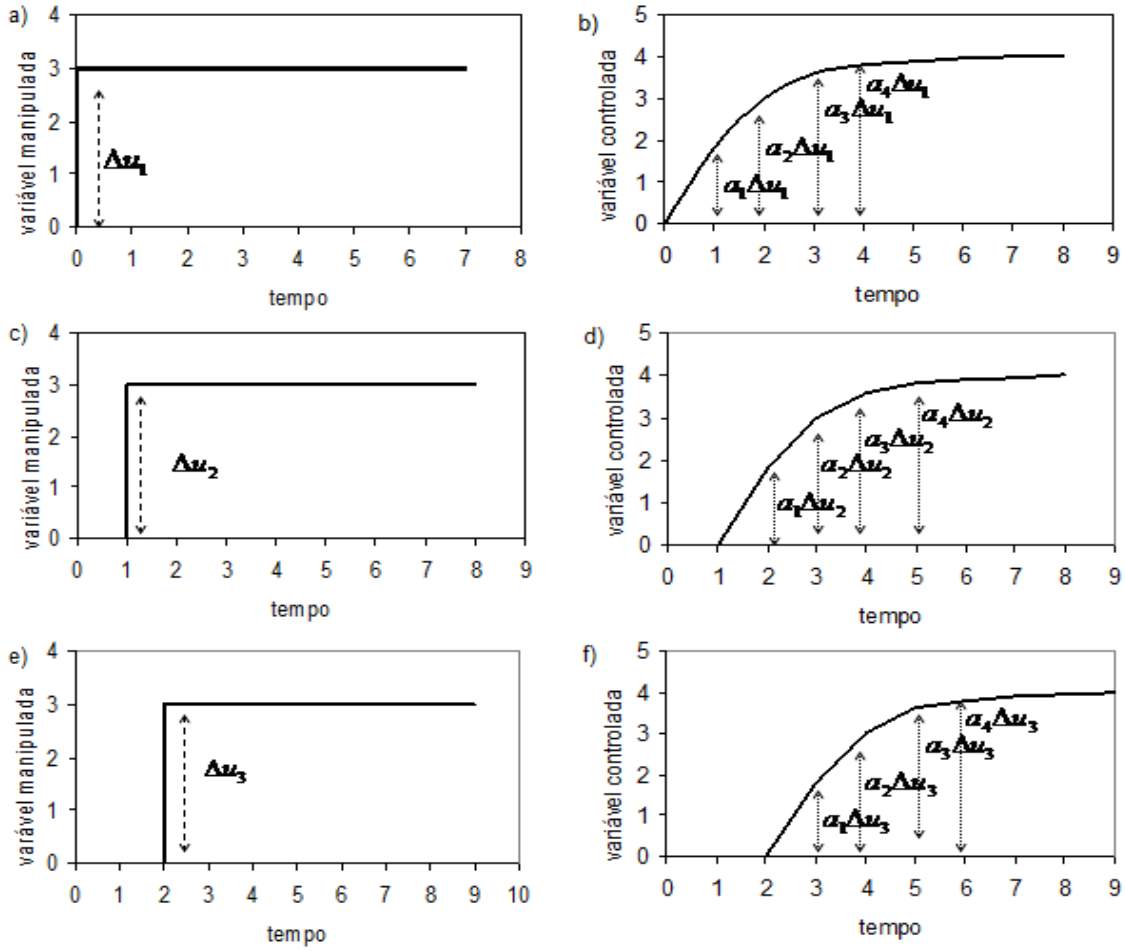


Figura 5.3. Resposta dinâmica do sistema quando são aplicadas três perturbações degrau na variável manipulada.

Em sistemas lineares, segundo o princípio de superposição, a resposta a qualquer combinação linear de seqüências de entradas é simplesmente uma combinação linear de seqüências de saídas (KWONG, 2005), então, da Figura 5.3, para a seqüência de três perturbações degrau na variável manipulada Δu_1 , Δu_2 e Δu_3 , o valor da resposta (variável controlada) $\hat{c}_k$, em cada tempo de amostragem T_s é uma combinação linear que pode ser calculada por:

$$\hat{c}_1 = a_1 \Delta u_1$$

$$\hat{c}_2 = a_2 \Delta u_1 + a_1 \Delta u_2$$

$$\hat{c}_3 = a_3 \Delta u_1 + a_2 \Delta u_2 + a_1 \Delta u_3$$

$$\hat{c}_4 = a_4 \Delta u_1 + a_3 \Delta u_2 + a_2 \Delta u_3$$

...

$$\hat{c}_N = a_N \Delta u_1 + a_{N-1} \Delta u_2 + a_{N-2} \Delta u_3 \quad (5.23)$$

As equações acima podem ser generalizadas para N valores da resposta e H_c variações na variável manipulada. Escrevendo o modelo de convolução na forma matricial:

$$\begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \\ \hat{c}_3 \\ \vdots \\ \hat{c}_{H_c} \\ \vdots \\ \hat{c}_N \end{bmatrix} = \begin{bmatrix} a_1 & 0 & 0 & \dots & 0 \\ a_2 & a_1 & 0 & \dots & 0 \\ a_3 & a_2 & a_1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{H_c} & a_{H_c-1} & a_{H_c-2} & \dots & a_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_N & a_{N-1} & a_{N-2} & \dots & a_{N-H_c+1} \end{bmatrix} \begin{bmatrix} \Delta u_1 \\ \Delta u_2 \\ \Delta u_3 \\ \vdots \\ \vdots \\ \vdots \\ \Delta u_{H_c} \end{bmatrix} \quad (5.24)$$

$$\hat{C}_{N \times 1} = A_{N \times H_c} \Delta u_{H_c \times 1} \quad (5.25)$$

5.3.2. ALGORITMO DE CONTROLE PARA SISTEMAS MONOVARIÁVEIS

Em sistemas monovariáveis SISO, uma variável de saída (controlada) é relacionada com uma variável de entrada (manipulada) ao longo do tempo.

O valor predito da variável de saída $\hat{c}$ para o instante de amostragem k , é dado pela equação:

$$\hat{c}_k = \sum_{i=1}^N h_i u_{k-i} \quad (5.26)$$

Onde u_{k-1} é o valor da variável manipulada no instante $k-1$, N é o horizonte do modelo.

O coeficiente de resposta ao impulso h_i é igual à diferença entre dois coeficientes de resposta ao degrau sucessivos.

$$h_i = a_i - a_{i-1} \quad (5.27)$$

O coeficiente de resposta ao degrau a_i representa o valor da variável de saída em resposta ao degrau.

Para o instante de amostragem $k+1$, tem-se:

$$\hat{c}_{k+1} = \sum_{i=1}^N h_i u_{k+1-i} \quad (5.28)$$

Subtraindo a Equação 5.26 da Equação 5.28 obtém-se, em termos de variações incrementais:

$$\hat{c}_{k+1} = \hat{c}_k + \sum_{i=1}^N h_i \Delta u_{k+1-i} \quad (5.29)$$

$$\Delta u_k = u_k - u_{k-1} \quad (5.30)$$

Para H_p instantes futuros, seja $j=1,2,3,\dots, H_p$

$$\hat{c}_{k+j} = \hat{c}_{k+j-1} + \sum_{i=1}^N h_i \Delta u_{k+j-i} \quad (5.31)$$

H_p é denominado horizonte de predição.

A predição do modelo dada pela Equação 5.31 é corrigida recursivamente, comparando o valor medido c_k no instante de amostragem k , com o valor predito da saída $\hat{c}_k$. A diferença d calculada entre os dois valores é utilizada para corrigir futuras predições porque os valores futuros de d não são disponíveis (DOUGHERTY e COOPER, 2003). Assim, as predições para os instantes $k+1$ até $k+P$ são corrigidas pela diferença d entre o valor medido do processo e o valor predito da saída, no instante atual de amostragem k .

$$d = c_k - \hat{c}_k \quad (5.32)$$

$$c_{k+1}^c = \hat{c}_{k+1} + (c_k - \hat{c}_k) = \hat{c}_{k+1} + d \quad (5.33)$$

$$c_{k+2}^c = \hat{c}_{k+2} + (c_{k+1}^c - \hat{c}_{k+1}) = \hat{c}_{k+2} + ((\hat{c}_{k+1} + d) - \hat{c}_{k+1}) = \hat{c}_{k+2} + d \quad (5.34)$$

$$c_{k+3}^c = \hat{c}_{k+3} + (c_{k+2}^c - \hat{c}_{k+2}) = \hat{c}_{k+3} + ((\hat{c}_{k+2} + d) - \hat{c}_{k+2}) = \hat{c}_{k+3} + d \quad (5.35)$$

...

Para o instante $K+H_p$, tem-se:

$$c_{k+NP}^c = \hat{c}_{k+NP} + d \quad (5.36)$$

Generalizando para os H_p instantes futuros:

$$c_{k+j-1}^c = \hat{c}_{k+j-1} + d \text{ para } j=1,2,\dots,H_p \text{ e} \quad (5.37)$$

$$c_{k+j}^c = \hat{c}_{k+j} + d \text{ para } j=1,2,\dots,H_p \quad (5.38)$$

As correções na predição dos valores da saída são atualizadas a cada instante de amostragem, assim sempre são utilizadas medidas atuais da planta introduzindo retroalimentação ou *feedback* no controlador.

Combinando-se as Equações 5.36, 5.37 e 5.38, obtém-se:

$$c_{k+j}^c = c_{k+j-1}^c + \sum_{i=1}^N h_i \Delta u_{k+j-i} \quad \text{para } j=1,2,\dots, Hp \quad (5.39)$$

A Equação 5.39 escrita na forma vetor-matriz para Hp instantes futuros, é:

$$\begin{bmatrix} c_{k+1}^c \\ c_{k+2}^c \\ \vdots \\ c_{k+Hp-1}^c \\ c_{k+Hp}^c \end{bmatrix} = \begin{bmatrix} a_1 & 0 & \cdots & 0 & 0 \\ a_2 & a_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{Hp-1} & a_{Hp-2} & \cdots & a_1 & 0 \\ a_{Hp} & a_{Hp-1} & \cdots & a_2 & a_1 \end{bmatrix} \begin{bmatrix} \Delta u_k \\ \Delta u_{k+1} \\ \vdots \\ \Delta u_{k+Hp-2} \\ \Delta u_{k+Hp-1} \end{bmatrix} + \begin{bmatrix} c_k + P_1 \\ c_k + P_2 \\ \vdots \\ c_k + P_{Hp-1} \\ c_k + P_{Hp} \end{bmatrix} \quad (5.40)$$

$$a_i = \sum_{j=1}^i h_j \quad (5.41)$$

$$\hat{c}_k = \sum_{i=1}^N h_i u_{k-1} \quad (5.42)$$

$$P_i = \sum_{m=1}^i S_m \quad i=1,2,\dots, Hp \quad (5.43)$$

$$S_m = \sum_{i=m+1}^N h_i \Delta u_{k+m-i} \quad m=1,2,\dots, Hp \quad (5.44)$$

Em geral, os valores desejados da variável controlada c_{k+j}^d podem ser especificados por uma trajetória de referência dada pela equação:

$$c_{k+j}^d = \alpha^j c_k + (1 - \alpha^j) c_{sp_k} \quad (5.45)$$

Onde $j=1,2,\dots, Hp$, $0 \leq \alpha < 1$ e c_{sp_k} e o set point atual.

Na forma matricial:

$$\begin{bmatrix} c_{k+1}^d \\ c_{k+2}^d \\ \vdots \\ c_{k+Hp-1}^d \\ c_{k+Hp}^d \end{bmatrix} = \begin{bmatrix} \alpha^1 c_k + (1-\alpha^1) c_{sp_k} \\ \alpha^2 c_k + (1-\alpha^2) c_{sp_k} \\ \vdots \\ \alpha^{Hp-1} c_k + (1-\alpha^{Hp-1}) c_{sp_k} \\ \alpha^{Hp} c_k + (1-\alpha^{Hp}) c_{sp_k} \end{bmatrix} \quad (5.46)$$

Para $\alpha = 0$, o valor desejado da variável controlada é igual ao set point atual:

$$c_{k+j}^d = c_{sp_k} \quad (5.47)$$

Portanto, a Equação 5.46 reduz-se a:

$$\begin{bmatrix} c_{k+1}^d \\ c_{k+2}^d \\ \vdots \\ c_{k+Hp-1}^d \\ c_{k+Hp}^d \end{bmatrix} = \begin{bmatrix} c_{sp_k} \\ c_{sp_k} \\ \vdots \\ c_{sp_k} \\ c_{sp_k} \end{bmatrix} \quad (5.48)$$

Subtraindo a Equação 5.40 da 5.48, tem-se:

$$\begin{bmatrix} c_{k+1}^d - c_{k+1}^c \\ c_{k+2}^d - c_{k+2}^c \\ \vdots \\ c_{k+Hp-1}^d - c_{k+Hp-1}^c \\ c_{k+Hp}^d - c_{k+Hp}^c \end{bmatrix} = \begin{bmatrix} a_1 & 0 & \cdots & 0 & 0 \\ a_2 & a_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{Hp-1} & a_{Hp-2} & \cdots & a_1 & 0 \\ a_{Hp} & a_{Hp-1} & \cdots & a_2 & a_1 \end{bmatrix} \begin{bmatrix} \Delta u_k \\ \Delta u_{k+1} \\ \vdots \\ \Delta u_{k+Hp-2} \\ \Delta u_{k+Hp-1} \end{bmatrix} + \begin{bmatrix} c_{sp_k} - c_k + P_1 \\ c_{sp_k} - c_k + P_2 \\ \vdots \\ c_{sp_k} - c_k + P_{Hp-1} \\ c_{sp_k} - c_k + P_{Hp} \end{bmatrix} \quad (5.49)$$

Definindo-se:

$$E = \begin{bmatrix} c_{k+1}^d - c_{k+1}^c \\ c_{k+2}^d - c_{k+2}^c \\ \vdots \\ c_{k+Hp-1}^d - c_{k+Hp-1}^c \\ c_{k+Hp}^d - c_{k+Hp}^c \end{bmatrix} \quad (5.50)$$

E é um vetor de erro predito, referido como uma predição em malha fechada baseada em ações de controle corrente e futuras.

$$M = \begin{bmatrix} a_1 & 0 & \cdots & 0 & 0 \\ a_2 & a_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{Hp-1} & a_{Hp-2} & \cdots & a_1 & 0 \\ a_{Hp} & a_{Hp-1} & \cdots & a_2 & a_1 \end{bmatrix} \quad (5.51)$$

$$\Delta u = \begin{bmatrix} \Delta u_k \\ \Delta u_{k+1} \\ \vdots \\ \Delta u_{k+Hp-2} \\ \Delta u_{k+Hp-1} \end{bmatrix} \quad (5.52)$$

$$E' = \begin{bmatrix} e_k - P_1 \\ e_k - P_2 \\ \vdots \\ e_k - P_{Hp-1} \\ e_k - P_{Hp} \end{bmatrix} \quad (5.53)$$

E' é o desvio predito da saída em relação à trajetória desejada baseado nas ações de controle passadas.

$$e_k = c_{sp_k} - c_k \quad (5.54)$$

A equação matricial é:

$$E_{Hp \times 1} = -M_{Hp \times Hp} \Delta u_{Hp \times 1} + E'_{Hp \times 1} \quad (5.55)$$

Quando a saída predita é igual à saída desejada, $E = 0$, então:

$$0 = -M \Delta u + \hat{E}' \quad (5.56)$$

consequentemente:

$$\Delta u = (M)^{-1} + \hat{E}' \quad (5.57)$$

(este resultado é denominado de protótipo mínimo)

A estratégia do DMC consiste em obter um sistema sobredeterminado reduzindo a dimensão do vetor Δu para o tamanho do horizonte de controle, H_c , admitindo que $\Delta u_{k+j} = 0$, para $j \geq H_c$. Sendo que $H_p > H_c$. Assim, apenas H_c ações futuras de controle são calculadas e a Equação 5.40 passa a ser:

$$\begin{bmatrix} c_{k+1}^c \\ c_{k+2}^c \\ \vdots \\ c_{k+Hp-1}^c \\ c_{k+Hp}^c \end{bmatrix} = \begin{bmatrix} a_1 & & \cdots & 0 & 0 \\ a_2 & a_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{Hp-1} & a_{Hp-2} & \cdots & a_{Hp-Hc} & 0 \\ a_{Hp} & a_{Hp-1} & \cdots & a_{Hp-Hc+1} & a_1 \end{bmatrix} \begin{bmatrix} \Delta u_k \\ \Delta u_{k+1} \\ \vdots \\ \Delta u_{k+Hc-2} \\ \Delta u_{k+Hc-1} \end{bmatrix} + \begin{bmatrix} c_k + P_1 \\ c_k + P_2 \\ \vdots \\ c_k + P_{Hp-1} \\ c_k + P_{Hp} \end{bmatrix} \quad (5.58)$$

A Equação 5.55 pode ser escrita como:

$$E = -A\Delta u + E' \quad (5.59)$$

As dimensões das matrizes são:

$$E_{Hp \times 1} = -A_{Hp \times Hc} \Delta u_{Hc \times 1} + E'_{Hp \times 1} \quad (5.60)$$

Uma notação compacta para a equação matricial é (DOUGHERTY e COOPER, 2003):

$$\bar{E} = -A\Delta u \quad (5.61)$$

Onde $\bar{E}$ é o vetor de erros preditos nos próximos NP instantes de amostragem, A é a matriz dinâmica de dimensão $Hp \times Hc$, que relaciona as entradas com as saídas do processo. Δu é o vetor das ações de controle a ser determinado.

O sistema da Equação 5.61 não tem solução exata, desde que o número de equações exceda os graus de liberdade ($Hp > Hc$). Mas é possível encontrar a melhor solução minimizando o índice de desempenho:

$$\min_{\Delta u} J = [\bar{E} - A\Delta u]^T [\bar{E} - A\Delta u] \quad (5.62)$$

A solução ótima é:

$$\Delta u = (A^t A)^{-1} A^t \bar{E} \quad (5.63)$$

A equação 5.63 representa a lei de controle do DMC. O uso da Equação 5.63 pode resultar em ações de controle muito bruscas, por isso se utiliza um índice de performance modificado, penalizando os movimentos da variável manipulada. A função objetivo modificada tem a forma:

$$\min_{\Delta u} J = [\bar{E} - A\Delta u]^t [\bar{E} - A\Delta u] + \lambda^2 [\Delta u^t \Delta u] \quad (5.64)$$

Onde λ é um fator de supressão para mudanças bruscas nas variáveis manipuladas.

No caso sem restrições, a função objetivo modificada tem como solução (LUYBEN, 1990; ONGUNNAIKE, 1994; SHRIDHAR e COOPER, 1997; DOUGHERTY e COOPER, 2003):

$$\Delta u = (A' A + \lambda^2 I)^{-1} A' \bar{E}' = K_c \bar{E}' \quad (5.65)$$

Na Equação 5.65, se A é constante, K_c se calcula apenas uma vez.

A lei de controle da Equação 5.65 indica quais serão os próximos H_c movimentos da variável manipulada, mas se aplica apenas a primeira ação de controle Δu_k . O problema de minimização é resolvido a cada instante de amostragem.

5.3.3. RESPOSTA AO DEGRAU- MIMO

Considerando o caso de um processo com duas entradas e duas saídas, com as seguintes seqüências de perturbação degrau nas variáveis manipuladas $u_1: \Delta u_{1,1}, \Delta u_{1,2}, \Delta u_{1,3}, \dots, \Delta u_{1,NC}$; e $u_2: \Delta u_{2,1}, \Delta u_{2,2}, \Delta u_{2,3}, \dots, \Delta u_{2,NC}$, os N valores da resposta para as duas variáveis são os seguintes:

$$\begin{aligned} \hat{c}_{1,1} &= a_{11,1} \Delta u_{1,1} + a_{12,1} \Delta u_{2,1} \\ \hat{c}_{1,2} &= a_{11,2} \Delta u_{1,1} + a_{11,2} \Delta u_{1,2} + a_{12,2} \Delta u_{2,1} + a_{12,2} \Delta u_{2,2} \\ \hat{c}_{1,3} &= a_{11,3} \Delta u_{1,1} + a_{11,2} \Delta u_{1,2} + a_{11,1} \Delta u_{1,3} + a_{12,3} \Delta u_{2,1} + a_{12,2} \Delta u_{2,2} + a_{12,1} \Delta u_{2,3} \\ &\dots \\ \hat{c}_{1,N} &= a_{11,N} \Delta u_{1,1} + \dots + a_{11,N-H_c+1} \Delta u_{1,H_c} + a_{12,N} \Delta u_{2,1} + \dots + a_{12,N-H_c+1} \Delta u_{2,H_c} \\ &\dots \\ \hat{c}_{2,1} &= a_{21,1} \Delta u_{1,1} + a_{22,1} \Delta u_{2,1} \\ \hat{c}_{2,2} &= a_{21,2} \Delta u_{1,1} + a_{21,1} \Delta u_{1,2} + a_{22,2} \Delta u_{2,1} + a_{22,1} \Delta u_{2,2} \\ \hat{c}_{2,3} &= a_{21,3} \Delta u_{1,1} + a_{21,2} \Delta u_{1,2} + a_{21,1} \Delta u_{1,3} + a_{22,3} \Delta u_{2,1} + a_{22,2} \Delta u_{2,2} + a_{22,1} \Delta u_{2,3} \\ &\dots \\ \hat{c}_{2,N} &= a_{21,N} \Delta u_{1,1} + \dots + a_{21,N-H_c+1} \Delta u_{1,H_c} + a_{22,N} \Delta u_{2,1} + \dots + a_{22,N-H_c+1} \Delta u_{2,H_c} \end{aligned} \quad (5.66)$$

Generalizando para NI entradas e NO saídas, com NC perturbações nas variáveis e escrevendo o modelo de convolução na forma matricial, tem-se:

$$\begin{bmatrix} \hat{c}_{1,1} \\ \hat{c}_{1,2} \\ \vdots \\ \hat{c}_{1,N} \\ \vdots \\ \hat{c}_{NO,1} \\ \hat{c}_{NO,2} \\ \vdots \\ \hat{c}_{NO,N} \end{bmatrix} = \begin{bmatrix} a_{11,1} & 0 & \dots & 0 & \dots & a_{1NI,1} & 0 & \dots & 0 \\ a_{11,2} & a_{11,1} & \dots & 0 & \dots & a_{1NI,2} & a_{1NI,1} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{11,N} & a_{11,N-1} & \dots & a_{11,N-Hc+1} & \dots & a_{1NI,N} & a_{1NI,N-1} & \dots & a_{1NI,N-Hc+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{NO,1,1} & 0 & \dots & 0 & \dots & a_{NO,NI,1} & 0 & \dots & 0 \\ a_{NO,1,2} & a_{NO,1,1} & \dots & 0 & \dots & a_{NO,NI,2} & a_{NO,NI,1} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{NO,N} & a_{NO,N-1} & \dots & a_{NO,N-Hc+1} & \dots & a_{NO,NI,N} & a_{NO,NI,N-1} & \dots & a_{NO,NI,N-Hc+1} \end{bmatrix} \begin{bmatrix} \Delta u_{1,1} \\ \Delta u_{1,2} \\ \vdots \\ \Delta u_{1,Hc} \\ \vdots \\ \Delta u_{NI,1} \\ \Delta u_{NI,2} \\ \vdots \\ \Delta u_{NI,Hc} \end{bmatrix} \quad (5.67)$$

As dimensões do modelo de convolução para o caso MIMO são:

$$\hat{C}_{[NO \times N, 1]} = A_{[NO \times N, NI \times Hc]} \Delta u_{[NI \times Hc, 1]}$$

5.3.4. ALGORITMO DE CONTROLE PARA SISTEMAS MULTIVARIÁVEIS

Utilizando o princípio da superposição, os resultados desenvolvidos para sistemas monovariáveis, podem ser estendidos para sistemas multivariáveis com M entradas e M saídas, nesse caso o modelo de convolução torna-se:

$$\begin{aligned} \hat{c}_{1,k+1} &= \sum_{i=1}^N h_{11,i} u_{1,k+1-i} + \dots + \sum_{i=1}^N h_{1NI,i} u_{NI,k+1-i} \\ \hat{c}_{2,k+1} &= \sum_{i=1}^N h_{21,i} u_{1,k+1-i} + \dots + \sum_{i=1}^N h_{2NI,i} u_{NI,k+1-i} \\ &\vdots \\ \hat{c}_{NO,k+1} &= \sum_{i=1}^N h_{NO1,i} u_{1,k+1-i} + \dots + \sum_{i=1}^N h_{NONI,i} u_{NI,k+1-i} \end{aligned} \quad (5.68)$$

$\hat{c}_{NO,k+1}$, denota os valores preditos para o instante de amostragem k+1, para as NO saídas $c_1, \dots, c_M$.

Estendendo para H_p instantes futuros, com $j=1, 2, \dots, H_p$, tem-se.

$$\begin{aligned} \hat{c}_{1,k+j} &= \hat{c}_{1,k+j-1} + \sum_{i=1}^N h_{11,i} \Delta u_{1,k+j-i} + \dots + \sum_{i=1}^N h_{1NI,i} \Delta u_{NI,k+j-i} \\ \hat{c}_{2,k+j} &= \hat{c}_{2,k+j-1} + \sum_{i=1}^N h_{21,i} \Delta u_{1,k+j-i} + \dots + \sum_{i=1}^N h_{2NI,i} \Delta u_{NI,k+j-i} \\ &\vdots \\ \hat{c}_{NO,k+j} &= \hat{c}_{NO,k+j-1} + \sum_{i=1}^N h_{NO1,i} \Delta u_{1,k+j-i} + \dots + \sum_{i=1}^N h_{NONI,i} \Delta u_{NI,k+j-i} \end{aligned} \quad (5.69)$$

Corrigindo o valor com base no valor atual:

$$\begin{aligned}
c_{1,k+j}^c &= \hat{c}_{1,k+j} + (c_{1,k+j-1}^c - \hat{c}_{1,k+j-1}) \\
c_{2,k+j}^c &= \hat{c}_{2,k+j} + (c_{2,k+j-1}^c - \hat{c}_{2,k+j-1}) \\
&\vdots \\
c_{NO,k+j}^c &= \hat{c}_{NO,k+j} + (c_{NO,k+j-1}^c - \hat{c}_{NO,k+j-1})
\end{aligned} \tag{5.70}$$

Para $j=1,2,\dots,Hp$

$$\begin{aligned}
c_{1,k}^c &= c_{1,k} \\
c_{2,k}^c &= c_{2,k} \\
&\vdots \\
c_{NO,k}^c &= c_{NO,k}
\end{aligned} \tag{5.71}$$

O modelo corrigido é obtido é:

$$\begin{aligned}
c_{1,k+j}^c &= c_{1,k+j-1}^c + \sum_{i=1}^N h_{11,i} \Delta u_{1,k+j-i} + \dots + \sum_{i=1}^N h_{1NI,i} \Delta u_{NI,k+j-i} \\
c_{2,k+j}^c &= c_{2,k+j-1}^c + \sum_{i=1}^N h_{21,i} \Delta u_{1,k+j-i} + \dots + \sum_{i=1}^N h_{2NI,i} \Delta u_{NI,k+j-i} \\
&\vdots \\
c_{NO,k+j}^c &= c_{NO,k+j-1}^c + \sum_{i=1}^N h_{NO1,i} \Delta u_{1,k+j-i} + \dots + \sum_{i=1}^N h_{NONI,i} \Delta u_{NI,k+j-i}
\end{aligned} \tag{5.72}$$

Escrevendo na forma vetor-matriz para os Hp instantes futuros, resulta:

$$\begin{bmatrix} c_1^c \\ c_2^c \\ \vdots \\ c_{NO-1}^c \\ c_{NO}^c \end{bmatrix} = \begin{bmatrix} A'_{11} & A'_{11} & \dots & A'_{11} & A'_{1NI} \\ A'_{21} & A'_{22} & \dots & A'_{21} & A'_{2NI} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ A'_{NO-11} & A'_{NO-12} & \dots & A'_{NO-1NI-1} & A'_{NO-1NI} \\ A'_{NO1} & A'_{NO2} & \dots & A'_{NONI-1} & A'_{NONI} \end{bmatrix} \begin{bmatrix} \Delta u_k \\ \Delta u_2 \\ \vdots \\ \Delta u_{NI-1} \\ \Delta u_{NI} \end{bmatrix} + \begin{bmatrix} c_{1,k} F + P_1 \\ c_{2,k} F + P_2 \\ \vdots \\ c_{NO-1,k} F + P_{NO-1} \\ c_{NO,k} F + P_{NO} \end{bmatrix} \tag{5.73}$$

Cada partição A'_{ij} é triangular:

$$A'_{ij} = \begin{bmatrix} a_{ij,1} & 0 & \dots & 0 & 0 \\ a_{ij,2} & a_{ij,1} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{ij,Hp-1} & a_{ij,Hp-2} & \dots & a_{ij,1} & 0 \\ a_{ij,Hp} & a_{ij,Hp-1} & \dots & a_{ij,2} & a_{ij,1} \end{bmatrix} \tag{5.74}$$

Os elementos a_{ij} são definidos como:

$$a_{ij,m} = \sum_{l=1}^m h_{ij,l} \quad (5.75)$$

$$c_n^c = [c_{n,k+1}^c \quad c_{n,k+2}^c \quad \dots \quad c_{n,k+Hp}^c]^t \quad n=1,2,\dots,M \quad (5.76)$$

$$\Delta u_n = [\Delta u_{n,k} \quad \Delta u_{n,k+1} \quad \dots \quad \Delta u_{n,k+L-1}] \quad n=1,2,\dots,M \quad (5.77)$$

$$F = [1 \quad 1 \quad \dots \quad 1]^t \quad (5.78)$$

$$P_n = [P_{n,1} \quad P_{n,2} \quad \dots \quad P_{n,Hp}]^t \quad n=1,2,\dots,M \quad (5.79)$$

$$P_{n,j} = \sum_{m=1}^j S_{n,m} \quad j=1,2,\dots,Hp \quad (5.80)$$

$$S_{n,m} = \sum_{i=m+1}^N \sum_{l=1}^M h_{nl,i} \Delta u_{l,k+m-i} \quad m=1,2,\dots,Hp \quad (5.81)$$

A trajetória desejada é:

$$\begin{aligned} c_{1,k+j}^d &= \alpha_1^j c_{1,k} + (1 - \alpha_1^j) c_{sp1,k} \\ c_{2,k+j}^d &= \alpha_2^j c_{2,k} + (1 - \alpha_2^j) c_{sp2,k} \\ &\vdots \\ c_{NO,k+j}^d &= \alpha_{NO}^j c_{NO,k} + (1 - \alpha_{NO}^j) c_{spNO,k} \end{aligned} \quad (5.82)$$

Para $\alpha = 0$, o valor desejado da variável controlada é igual ao setpoint atual:

$$c_{1,k+j}^d = c_{sp1,k}, \quad c_{2,k+j}^d = c_{sp2,k}, \dots, c_{NO,k+j}^d = c_{spNO,k} \quad (5.83)$$

Na forma matricial é:

$$\begin{bmatrix} c_1^d \\ c_2^d \\ \vdots \\ c_{NO-1}^d \\ c_{NO}^d \end{bmatrix} = \begin{bmatrix} c_{sp1,k} F \\ c_{sp2,k} F \\ \vdots \\ c_{spNO-1,k} F \\ c_{spNO,k} F \end{bmatrix} \quad (5.84)$$

Onde:

Subtraindo a Equação 5.73 da Equação 5.84.

$$E = -A' \Delta u + E' \quad (5.85)$$

$$E = \begin{bmatrix} c_1^d - c_1^c \\ c_2^d - c_2^c \\ \vdots \\ c_{NO-1}^d - c_{NO-1}^c \\ c_{NO}^d - c_{NO}^c \end{bmatrix} \quad (5.86)$$

$$E' = \begin{bmatrix} E_{1,k}F - P_1 \\ E_{2,k}F - P_2 \\ \vdots \\ E_{NO-1,k}F - P_{NO-1} \\ E_{NO,k}F - P_{NO} \end{bmatrix} \quad (5.87)$$

Onde:

$$\begin{aligned} E_{1,k} &= c_{sp1,k} - c_{1,k} \\ E_{2,k} &= c_{sp2,k} - c_{2,k} \\ &\vdots \\ E_{NO,k} &= c_{spNO,k} - c_{NO,k} \end{aligned} \quad (5.88)$$

Reduzindo a dimensão do vetor Δu de $[NI \times Hp, I]$ para $[NI \times Hc, I]$, e procedendo de maneira análoga ao caso SISO, chega-se a:

$$E = -A\Delta u + E' \quad (5.89)$$

E e E' são vetores de dimensão $[NO \times Hp, I]$

$$E = \begin{bmatrix} E_{1,k+1} & \dots & E_{1,k+Hp} & \dots & E_{NO,k+1} & \dots & E_{NO,k+Hp} \end{bmatrix}^t \quad (5.90)$$

$$E' = \begin{bmatrix} E'_{1,k+1} & \dots & E'_{1,k+Hp} & \dots & E'_{NO,k+1} & \dots & E'_{NO,k+Hp} \end{bmatrix}^t \quad (5.91)$$

Δu é um vetor de dimensão $[Hc \times NI, I]$

$$\Delta u = \begin{bmatrix} \Delta u_{1,k+1} & \dots & \Delta u_{1,k+Hc} & \dots & \Delta u_{NI,k+1} & \dots & \Delta u_{NI,k+Hc} \end{bmatrix}^t \quad (5.92)$$

Para NI entradas e NO saídas a matriz A é:

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1NI-1} & A_{1NI} \\ A_{21} & A_{22} & \cdots & A_{2NI-1} & A_{2NI} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{NO-1,1} & A_{NO-1,2} & \cdots & A_{NO-1,NI-1} & A_{NO-1,NI} \\ A_{NO1} & A_{NO2} & \cdots & A_{NO,NI-1} & A_{NO,NI} \end{bmatrix} \quad (5.93)$$

Os elementos A_{ij} são definidos como:

$$A_{ij} = \begin{bmatrix} a_{ij,1} & 0 & \cdots & 0 & 0 \\ a_{ij,2} & a_{ij,1} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{ij,Hp-1} & a_{ij,Hp-2} & \cdots & a_{ij,Hp-Hc} & 0 \\ a_{ij,Hp} & a_{ij,Hp-1} & \cdots & a_{ij,Hp-Hc+1} & a_{ij,1} \end{bmatrix} \quad (5.94)$$

Cada elemento A_{ij} tem dimensão $Hp \times Hc$, por tanto a matriz dinâmica A possui dimensão $[NO \times Hp, NI \times Hc]$.

Para o caso DMC MIMO, o problema de otimização da função objetivo modificada que penaliza também os movimentos da variável manipulada é (SHRIDHAR e COOPER, 1997 e 1998):

$$\underset{\Delta u}{Min} J = [\bar{E} - A\Delta u]^T \Gamma^T \Gamma [\bar{E} - A\Delta u] + \Delta u^T \Lambda^T \Lambda \Delta u \quad (5.95)$$

A solução ótima é a lei de controle para o DMC MIMO sem restrições cuja equação resultante é:

$$\Delta u = (A^T \Gamma^T \Gamma A + \Lambda^T \Lambda)^{-1} A^T \Gamma^T \Gamma \bar{E} \quad (5.96)$$

Onde, Δu é o vetor das ações de controle a ser determinadas, a matriz dinâmica multivariável A possui número de linhas igual ao produto do número de variáveis controladas, NO , pela dimensão do horizonte de predição, Hp ; e número de colunas igual ao produto de número de variáveis manipuladas, NI , pela dimensão do horizonte de controle, Hc .

$\Gamma^T \Gamma$ é uma matriz de coeficientes de ponderação definida como:

$$\Gamma' \Gamma = \begin{bmatrix} \gamma_1^2 I_{H_{px}H_p} & 0 & 0 & 0 \\ 0 & \gamma_1^2 I_{H_{px}H_p} & 0 & 0 \\ 0 & 0 & \gamma_1^2 I_{H_{px}H_p} & 0 \\ 0 & 0 & 0 & \gamma_1^2 I_{H_{px}H_p} \end{bmatrix}_{NO.H_{px}NO.H_p} \quad (5.97)$$

$\Lambda' \Lambda$ é a matriz de coeficientes de supressão definida como:

$$\Lambda' \Lambda = \begin{bmatrix} \lambda_1^2 I_{H_{cx}H_c} & 0 & 0 & 0 \\ 0 & \lambda_1^2 I_{H_{cx}H_c} & 0 & 0 \\ 0 & 0 & \lambda_1^2 I_{H_{cx}H_c} & 0 \\ 0 & 0 & 0 & \lambda_1^2 I_{H_{cx}H_c} \end{bmatrix}_{NI.H_{cx}NI.H_c} \quad (5.98)$$

Os coeficientes de supressão têm um propósito duplo na lei de controle do DMC, de suprimir ações de controle agressivas e de acondicionamento da matriz do sistema antes da inversão.

5.4. CONTROLADOR DE MATRIZ DINÂMICA QUADRÁTICO (QDMC)

O QDMC foi desenvolvido por Garcia e Morshedi (1986), usa as mesmas equações de predição do DMC e consiste na solução on-line de um programa quadrático (QP), que minimiza a soma dos desvios quadrados das predições dos valores futuros das variáveis controladas em relação a seus setpoints, sujeito a manter as previsões das variáveis controladas e manipuladas restringidas dentro de certos limites. O QDMC fornece uma maneira sistemática para incluir restrições de entrada e de saída, esta é realizada ao colocar o problema de otimização como um programa quadrático.

Segundo Quin e Badwell (2003), as principais características do QDMC incluem:

- 1) Modelo de resposta ao degrau linear para a planta;
- 2) Objetivo de desempenho quadrático sobre um horizonte de previsão finito.
- 3) Comportamento futuro da saída da planta especificado tentando seguir o setpoint tão perto quanto possível, sujeito a um termo que penaliza os desvios.
- 4) Entradas ótimas calculadas como solução para um programa quadrático.

Para o QDMC, a melhor solução para o sistema da Equação 5.59 obtém-se minimizando a função objetivo:

$$J(\Delta u) = \frac{1}{2} E' E + \frac{1}{2} \Delta u' R \Delta u \quad (5.99)$$

Onde $R = \lambda I$, e λ é o fator de supressão. Substituindo Equação 5.59 na Equação 5.99:

$$J(\Delta u) = \frac{1}{2} [(-A\Delta u + E')'(-A\Delta u + E')] + \frac{1}{2} \Delta u' R \Delta u \quad (5.100)$$

$$J(\Delta u) = \frac{1}{2} [\Delta u' (A' A + R) \Delta u] - E'^t A \Delta u + \frac{1}{2} E'^t E' \quad (5.101)$$

$$J(\Delta u) = \frac{1}{2} [\Delta u' (A' A + R) \Delta u] - E'^t A \Delta u \quad (5.102)$$

A minimização da função objetivo pode ser escrita como (GATTU e ZAFIROU, 1992; OGUNNAIKE e REY, 1994):

$$\min_{\Delta u} J = \frac{1}{2} \Delta u' H \Delta u + c' \Delta u \quad (5.103)$$

Onde H é a matriz Hessiana

$$H = A' A + R \quad (5.104)$$

$$c' = E'^t A \quad (5.105)$$

Pode-se incluir as seguintes restrições (PORFÍRIO, 2001):

1) Restrições de limites nas variáveis manipuladas.

$$u_{\min} \leq u \leq u_{\max} \quad (5.106)$$

2) Restrições na magnitude dos movimentos das variáveis manipuladas.

$$-\Delta u_{\max} \leq \Delta u \leq \Delta u_{\max} \quad (5.107)$$

3) Restrições nas variáveis controladas.

$$c_{\min} \leq c \leq c_{\max} \quad (5.108)$$

1) Restrições de limites nas variáveis manipuladas

A equação matricial para as variáveis manipuladas (KWONG, 2005) é:

$$\begin{bmatrix} u_k \\ u_{k+1} \\ \vdots \\ u_{k+Hc-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \Delta u_k \\ \Delta u_{k+1} \\ \vdots \\ \Delta u_{k+Hc-1} \end{bmatrix} + \begin{bmatrix} u_{k-1} \\ u_{k-1} \\ u_{k-1} \\ u_{k-1} \end{bmatrix} \quad (5.109)$$

ou também:

$$u = N_u \Delta u + u_{k-1} F \quad (5.110)$$

onde $F = [1 \ 1 \ \cdots \ 1]^t$ tem dimensão apropriada.

Na equação (5.106) tem-se dois casos: $u \leq u_{\max}$ e $u \geq u_{\min}$

Para o caso $u \leq u_{\max}$ da Equação (5.110) tem-se:

$$N_u \Delta u \leq (u_{\max} - u_{k-1}) F \quad (5.111)$$

Para o caso $u \geq u_{\min}$ ou $-u \leq -u_{\min}$ da Equação (5.110) tem-se:

$$-N_u \Delta u \leq (u_{k-1} - u_{\min}) F \quad (5.112)$$

As condições dadas pelas eqs. (5.111) e (5.112) podem ser resumidas em:

$$\begin{bmatrix} N_u \\ -N_u \end{bmatrix} \Delta u \leq \begin{bmatrix} (u_{\max} - u_{k-1}) F \\ (u_{k-1} - u_{\min}) F \end{bmatrix} \quad (5.113)$$

2) Restrições na magnitude dos movimentos das variáveis manipuladas.

A restrição (5.113) apresenta dois casos: $\Delta u \leq \Delta u_{\max}$ e $-\Delta u_{\max} \leq \Delta u$.

Para o caso $\Delta u \leq \Delta u_{\max}$, tem-se:

$$I \Delta u \leq \Delta u_{\max} F \quad (5.114)$$

Para o caso $-\Delta u_{\max} \leq \Delta u$, tem-se:

$$-I \Delta u \leq \Delta u_{\max} F \quad (5.115)$$

Compondo as duas condições dadas pelas eqs. (5.114) e (5.115), tem-se:

$$\begin{bmatrix} I \\ -I \end{bmatrix} \Delta u \leq \begin{bmatrix} \Delta u_{\max} F \\ \Delta u_{\max} F \end{bmatrix} \quad (5.116)$$

3) Restrições nas variáveis controladas.

A restrição 5.108 apresenta dois casos: $c \leq c_{\max}$ e $c_{\min} \leq c$.

Para o caso $c \leq c_{\max}$ tem-se: $c^c \leq c_{\max}$, da Equação 5.58 $c^c = A\Delta u + C$, sendo

$C = [c_k + P]_{N_{px} \times 1}$, então:

$$A\Delta u \leq c_{\max} F - C \quad (5.117)$$

Para o caso $c_{\min} \leq c$ ou $-c \leq -c_{\min}$, tem-se:

$$-c^c \leq -c_{\min}$$

então:

$$-A\Delta u \leq C - c_{\min} F \quad (5.118)$$

compondo as duas condições dadas pelas Eqs. 5.117 e 5.118, tem-se:

$$\begin{bmatrix} A \\ -A \end{bmatrix} \Delta u \leq \begin{bmatrix} c_{\max} F - C \\ C - c_{\min} F \end{bmatrix} \quad (5.119)$$

Assim, o problema de otimização do QDMC tem como função objetivo a Equação 5.103 e as restrições dadas pelas Eqs. 5.113, 5.116 e 5.119. O problema de otimização do QDMC é um problema de programação quadrática que é resolvido em cada instante de amostragem.

5.5. CONCLUSÃO

Neste capítulo, foram apresentadas as características do controle preditivo com modelo, MPC, que utiliza técnicas de otimização para calcular os melhores valores das ações de controle a serem implementadas, a fim de que a variável controlada coincida com a trajetória de referência, e foram revisados os fundamentos matemáticos de técnicas de controle representativas da tecnologia MPC, tais como o controle de matriz dinâmica DMC e o controle de matriz quadrática QDMC, com otimização linear e não linear.

A implementação com sucesso destas técnicas em plantas industriais baseada na confiabilidade das suas características, justifica a aplicação das mesmas conjuntamente com modelos de processo e preditivos baseados em

redes neurais, para o controle de procesos complexos, como para o controle de um processo de fermentação extrativa contínua com flash, de complexidade reconhecida na literatura, por constituir uma classe de sistemas não lineares que apresentam características especiais pela interação reação-separação.

5.6. REFERÊNCIAS BIBLIOGRÁFICAS

1. Arulalan, G.R. and Deshpande, P.B. (1987). Simplified Model Predictive Control. *Industrial & Engineering Chemistry Research*, v. 26, pp. 347-356.
2. Camacho, E. F. and Bordons, C. **Model Predictive Control**. London, UK: Springer-Verlag, 2004.
3. Clarke, D.W., Mohtadi, C.;Tuffs, P.S. (1987a) Generalized Predictive Control, Part. I: The Basic Algorithm, *Automatica*, v.23,p.137-148.
4. Clarke, D.W., Mohtadi, C.;Tuffs, P.S. (1987b) Generalized Predictive Control, Part. II: Extensions and Interpretations. *Automatica*, v. 23, pp. 149-160.
5. Costa, A.C.,Meleiro, L.A.C., Maciel Filho, R.(2002). Non-Linear predictive control of an extractive alcoholic fermentation process. *Process Biochemistry*, v. 38, pp. 743-750.
6. Cutler, C. R. and Ramaker, D. L. (1980). Dynamic Matrix Control-A Computer Control Algorithm. *Proceedings of the Joint Automatic Control Conference*, San Francisco, CA, 1980.
7. Dougherty, D., and Cooper, D. A. (2003) Practical multiple model adaptive strategy for single-loop MPC. *Control Eng. Pract.*, v. 11, pp. 141–159.
8. Fisher, D.G.(1991). Process control: an overview and personal perspective. *The Canadian Journal of Chemical Engineering*, v. 69, pp. 5-26.
9. Gattu, G. and Zafiriou, E. (1992). Nonlinear Quadratic Dynamic Matrix Control with State Estimation. *Industrial & Engineering Chemistry Research*, v. 31, pp. 1096-1104
10. García, C.E. and Morari, M. (1982). Internal Model Control 1. A Unifying Review and Some New Results, *Industrial & Engineering Chemistry Process Design and Development*, v. 21, pp. 308-323.
11. García, C.E. (1984). Quadratic Dynamic Matrix Control of nonlinear processes: An application to a batch reaction process. *AIChE Annual Meeting*, San Francisco, U.S.A.
12. García, C.E. and Morari M. (1985a) Internal Model Control 2. Design Procedure for Multivariable Systems. *Industrial & Engineering Chemistry Process Design and Development*, v. 24, pp. 472-484.

13. García, C.E. and Morari M. (1985b) Internal Model Control 3. Multivariable Control Law Computation and Tuning Guidelines, *Industrial & Engineering Chemistry Process Design and Development*, v. 24, pp. 484-494.
14. García, C.E. and Morshedi, A.M. (1984). Solution of the dynamic matrix control problem via quadratic programming. *Proceedings of the Conference of the Canadian Industrial Computing Society*, Ottawa, Canada, pp. 13.1-13.3.
15. Garcia, C.E. and A.M. Morshedi, A.M. (1986). Quadratic Programming Solution of Dynamic Matrix Control (QDMC). *Chemical Engineering Communications*, v. 46 (1-3), pp. 73-87.
16. Himmelblau, D.M. and Edgar, T.F. **Optimization of Chemical Process**. New York, NY: MacGraw-Hill, 2001.
17. Luyben, W. *Process modeling, simulation and control for chemical engineers*; Second Edition; New York, NY: McGraw-Hill , 1990.
18. Kwong, W.H. **Introdução ao Controle Preditivo com Matlab**. São Carlos, SP. Brasil: EdUFSCar, 2005.
19. Meleiros, L.A.C., Costa, A.C. and Maciel Filho, R. (2005). Non-Linear Multivariable Predictive Control of an Alcoholic Fermentation Process Using Functional Link Networks. *Brazilian Archives of Biology and Technology*, v. 48, pp. 7-18.
20. Ogunnaike, B.A., Ray, W.H. (1994). Process **Dynamic Modeling and Control**. New York NY, USA : Oxford University Press. Oxford.
21. Porfírio, C.A. (2001). **Implantação de um controlador multimodelos em uma coluna depropanizadora industrial**. Tese de Mestrado. Escola Politécnica da Universidade de São Paulo. Brasil.
22. Qin,S.J. and Badgwell, T.A. (2003). A survey of industrial model predictive control technology. *Control Engineering Practice*, v. 11, pp. 733–764
23. Richalet, J., Rault, A., Testud, J. L., and Papon, J. (1978). Model predictive heuristic control: Applications to industrial processes. *Automatica*, v. 14, pp. 413–428.
24. Seborg, D.E., Edgar, T.F. and Mellichamp, D.A. **Process Dynamics and Control**. Hoboken, New Jersey, USA: Jhon Wiley & Sons, Inc., 2004.
25. Shridhar, R. and Cooper, D. (1997). A Tuning Strategy for Unconstrained SISO Model Predictive Control. *Industrial & Engineering Chemistry Research*, v. 36, pp. 729-746.
26. Shridhar, R. e Cooper, D. (1998). A Tuning Strategy for Unconstrained Multivariable Model Predictive Control. *Industrial & Engineering Chemistry Research*, v. 37, pp. 4003-4016.
27. Soertemboek. **Predictive Control. A unified approach**. Doctor Thesis. Delft University of Technology. Delft, Netherlands, 1990.

28. Taha, H.A. **Operations Research: An introduction**. Seventh Edition. Upper Saddle River, NJ: Prentice Hall, 2003.

CAPÍTULO 6. ENZYME LOADING OPTIMIZATION IN THE SUGARCANE BAGASSE HYDROLYSIS FOR SECOND GENERATION ETHANOL PRODUCTION USING ARTIFICIAL INTELLIGENCE TECHNIQUES

Trabalho apresentado em Congresso Internacional:

Edwin G. Boza Condorena, Laura G. Fuentes, Sarita C. Rabelo, and Aline Carvalho da Costa. 32nd Symposium on Biotechnology for Fuels and Chemicals. Clearwater Beach, Florida. USA. 2010.

Prefácio

Neste capítulo é apresentada uma aplicação do uso de redes neurais e algoritmos genéticos para modelagem e otimização de processos. O caso de estudo é um processo de hidrólise enzimática de bagaço de cana para obtenção de glicose na produção de etanol de segunda geração. Não existia um modelo do processo disponível, apenas dados experimentais, obtidos na tese de Fuentes (2009), desenvolvida no mesmo grupo de pesquisa, em que o presente trabalho vem sendo desenvolvido. Assim, as redes neurais foram usadas para modelagem do processo e os algoritmos genéticos foram usados para, a partir do modelo determinado, determinar a carga de enzimas necessária para maximizar o rendimento de glicose.

Abstract

BACKGROUND: The task of finding an optimal combination of enzymes is extremely important to get the highest glucose production in enzymatic hydrolysis. In this study artificial neural networks (ANN) and genetic algorithms (GA), two powerful tools of artificial intelligence (AI), were integrated for modeling and optimization. Bayesian regularization, one of the best alternatives for improving generalization and avoiding overfit was used to develop the dynamic model.

RESULTS: The optimal region found for the maximum glucose production was: cellulase in the range 420 - 600 FPUL⁻¹, β -glucosidase in the range 540-850 CBU L⁻¹ and reaction time in the range 3000 min (50 h) - 3720 min (62 h). The global optimum was: 482 FPU L⁻¹ (16.1 FPU g⁻¹ bagasse) cellulase and 688 CBU L⁻¹ (22.9 CBU g⁻¹ bagasse) β -glucosidase for a reaction time of 57 h. This corresponds to a glucose concentration of 16.725 g L⁻¹ and a yield of 83%.

CONCLUSIONS: In spite of the complexity of the multi-component enzymatic action, by integrating ANN and GA, the optimum coordinates (loads of enzymes, reaction time and glucose production) were successfully found in a 4-dimensional hyperspace; therefore they can be valuable tools for improving processes of ethanol production from cellulosic biomass.

Keywords: Enzymatic hydrolysis; sugarcane bagasse; modeling; optimization; neural networks.

6.1. INTRODUCTION

The global economic growth and the consequent increased demand for energy are promoting a series of social and environmental problems. Within this context there is an unprecedented interest focused on research and development of technologies for the sustainable production of energy from biomass.

Among several potential sources of biomass, sugarcane bagasse is a highly energetic material and its chemical composition makes it a material of high

technological potential, especially for bioproduction of second generation ethanol. Sugarcane ethanol is an attractive biofuel from a greenhouse gas (GHG) emissions perspective, because of its relatively large GHG emissions reduction potential¹.

The use of sugarcane bagasse for the production of ethanol as a substitute of gasoline could have a highly positive effect because it would be capable of increasing the supply of alcohol without the need to increase the acreage of sugarcane. Bagasse offers others big advantages: to have minimum cost and to come processed from the mills and ready to use at the production place².

However, to obtain ethanol from biomass for commercial use is still a great challenge. The biomass hydrolysis is currently performed by two routes: the acid or enzymatic hydrolysis. The second approach is promising because it can achieve high yields of sugar working in less drastic conditions of temperature, pressure and aggressive chemicals, eliminating the need to use large amounts of chemicals to treat the formation of inhibitory by-products as happens in acid hydrolysis, but this route has as disadvantages the high prices of enzymes and the hydrolysis time^{3,4}.

The enzymatic hydrolysis of cellulosic biomass is a promising route and methods are being developed to carry out this step for obtaining second generation ethanol in an efficient and low-cost manner, which makes research on this subject very important.

Cellulose enzymatic hydrolysis according to the most accepted mechanism⁵ is performed by synergistic action of different enzymes: exo and endoglucanases and β -glucosidase; in presence of substrate and product inhibitions⁶. The main product of the action of exo and endoglucanases is the formation of cellobiose, which is an important intermediate product, while the action of β -glucosidase produces glucose from cellobiose. Thus, low levels of glucose production are obtained with an enzyme loading poor in exo and endoglucanases or in β -glucosidase. Many commercial cellulase preparations (including that obtained from *Trichoderma reesei*) contain low levels of β -glucosidase^{7,8} and then it is necessary

to supplement β -glucosidase from other sources to increase glucose production. Therefore, the task of finding an optimal combination of loads of enzymes is extremely important, to get the highest production of glucose and consequently to optimize the ethanol production for commercial purposes.

The availability of an accurate mathematical model is important to find the optimal glucose production and combination of enzymes. In order to obtain response surface models for optimization of bioprocesses, researchers usually make use of experimental designs⁹⁻¹¹. However, it is known that the behavior of enzymatic reactions with varying enzyme loadings in hydrolysis process is extremely complex, therefore a response surface strongly nonlinear is expected^{7,11}. The response surface methodology based on experimental designs has limitations for modeling and optimization when the process is strongly nonlinear.

One alternative that can be successful under these conditions is the use of artificial intelligence (AI) techniques. Artificial neural networks (ANN) for modeling and genetic algorithms (GA) in the optimization step. These techniques have been proved successfully to optimize a fermentation medium for the production of the enzyme hydantoinase by *Agrobacterium radiobacter*¹² and the results suggest that both artificial intelligence techniques can be effectively integrated to create a powerful tool to process modeling and optimization. On the other hand, the capabilities of artificial neural networks (ANN) and response surface methodology (RSM) based on experimental designs were compared in the literature for the optimization of biochemical processes. The results show higher prediction accuracy of ANN compared to RSM based on experimental designs^{13,14}.

The aims of the present work were: 2²

- 1) To build an accurate model based on ANN for the evolution of the glucose concentration during the enzymatic hydrolysis of sugarcane bagasse, using a central composite design (CCD) 2² plus star to obtain a small experimental data set at each sampling time.

- 2) To find in the 4-dimensional hyperspace the optimum region and the optimal coordinates for glucose production, enzyme loadings and reaction time; using AI techniques: ANN for modeling and GA for optimization, as well as verifying the performance in optimization by comparing the results with those obtained using simulated annealing and threshold acceptance; and with experimental data.

6.2. MATERIAL AND METHODS

6.2.1. SUBSTRATE

The raw material for all experiments was sugarcane bagasse (*Saccharum officinarum*) from the sugar plant Usina da Pedra (São Paulo, Brazil). The sugarcane harvest was mechanized, without burning and the bagasse came from the last mill.

The bagasse was previously dried at room temperature for four days. After drying, the particles size was reduced in a hammer mill (General Electronic) and subsequently ground in a knife mill (Wiley Mill Model No. 3) for 10 minutes at each mill. After this physical treatment, the material was separated using Taylor sieves of 32, 35, 80, 100, and 150 meshes. The material retained on the sieves 35, 80, 100 and 150, which met with the desired size (0.5 mm or less), was mixed and then stored in freezer in airtight bags to prevent the growth of microorganisms.

The composition of sugarcane bagasse before and after pretreatment is shown in Table 6.1. On dry basis, 58.73 g of lime pretreated bagasse were obtained from 100 g of raw bagasse after pretreatment. Determination of chemical composition was made according to the laboratory analytical procedures (LAPs) for standard biomass analysis developed by the National Renewable Energy Laboratory (NREL) of the U.S. Department of Energy¹⁵⁻¹⁷.

Table 6.1. Composition of raw material and bagasse pretreated with lime		
Composition	Raw bagasse 100 g (%)	Lime pretreated bagasse 58.73 g (%)
Ash	1.79 ± 0.48	-
Cellulose	37.35 ± 0.51	60.45± 0.88
Hemicellulose	23.66 ± 0.88	21.33± 0.58
Lignin	25.10 ± 0.48	15.60± 0.24
Extractives	3.25 ± 0.20	-

6.2.2. LIME PRETREATMENT

4 g of dry bagasse were added in 500 mL Erlenmeyer flasks with 100 mL of distilled water and stirred in shaker at 150 rpm for 90 hours at 90 °C and with a lime loading of 0.4 w/w dry bagasse. After completion of the shaker cycle, the reaction was stopped with cold water and the pretreated biomass was neutralized with hydrochloric acid 5N to pH 7.0, washed with distilled water and oven dried at 40 °C for 48 h.

6.2.3. ENZYMATIC HYDROLYSIS

3g of lime pretreated bagasse was added to 100 mL of citrate buffer with 0.07 g of azide in erlenmeyers and the pH was adjusted to 4.8. The flasks were sealed and stirred in shaker at 150 rpm at a temperature of 50 °C.

For the enzymatic hydrolysis were used cellulase from *Trichoderma reesseei* ATCC 26921 (Sigma-Aldrich), and β -glucosidase from *Aspergillus niger* (Novozym 188). The concentrations of cellulase and β -glucosidase were determined for each assay according to runs of the design matrix (Table 6.2).

Table 6.2. Runs of the experimental design central composite 2^2 plus star (coded levels in parenthesis) and glucose (g L^{-1}) at three sampling times during the enzymatic hydrolysis.

Run	Cellulase (FPU L^{-1})	β -glucosidase (CBU L^{-1})	Glucose (g L^{-1}) time : 10 min (0.167 h)	Glucose (g L^{-1}) time: 2160 min (36 h)	Glucose (g L^{-1}) time : 3600 min (60 h)
1	175 (-1)	220 (-1)	2.042	14.652	14.028
2	175 (-1)	1280 (+1)	2.744	14.009	14.022
3	775 (1)	220 (-1)	2.704	14.355	15.821
4	775 (1)	1280 (+1)	2.608	14.729	14.235
5	50 (-1.4142)	750 (0)	3.520	12.128	12.265
6	900 (+1.4142)	750 (0)	2.697	14.443	14.878
7	475 (0)	0 (-1.4142)	2.383	13.831	13.072
8	475 (0)	1500 (+1.4142)	3.015	14.339	14.779
9,10,11	475 (0)	750 (0)	2.087 ± 0.522 (*)	15.580 ± 0.401 (*)	16.685 ± 0.226 (*)

* Average of three observations at the central point

Aliquots were taken at 10 min, 20 min, 30 min, 40 min, 50 min, 1 h, 3 h, 6 h, 12 h, 24 h, 36 h, 48 h, 60 h and 72 h of reaction and inactivated by increasing the temperature to 80 °C for 15 min. The concentration of glucose was determined for each sample by high-performance liquid chromatography (HPLC) using a SUGAR-PAK (Waters) column at 70 °C, eluted with deionized water at 0.5 ml/min and using a temperature-controlled refractive index detector (Model 410 waters, pump model 515) at 40 °C.

Cellulase activity (T. Reesseei, Sigma) was measured by the filter paper method as described by Ghose¹⁸ and was of 64.106 FPU mL^{-1} . β -glucosidase activity (Novozyme 188) was measured as described by Wood and Bhat¹⁹ and was of 309.92 CBU mL^{-1} .

6.2.4. DYNAMIC MODELING OF ENZYMATIC HYDROLYSIS

6.2.4.1. ANN CONFIGURATIONS

Neural networks are composed of simple elements (neurons) distributed in layers and operating in parallel. In between the input layer and the output layer are the hidden layers of the network. The inputs are propagated through network neurons being modified by weights and bias and following by neuron transfer functions of the network hidden layers. The final result is the network output, which is obtained from the neurons of the output layer. In this study after several tests of different ANN multilayer models a feedforward network with two layers, log-sigmoid transfer function on the hidden layer and linear transfer function on the output layer was used. The Bayesian regularization method for backpropagation training was used because it is one of the best methods to avoid overfit and improve generalization performance of the neural networks²⁰.

The data set of inputs and targets was normalized before training to make it more efficient, in the range [-1,1], as follows:

$$x_{iN} = \frac{2 * (x_i - x_{\min})}{x_{\max} - x_{\min}} - 1, \quad (6.1)$$

where: $x_{\min}$, $x_{\max}$ and x_{iN} are minimum, maximum and normalized values of the variable x_i .

The input x_i , weights W_{ji} , bias b_j^1 , transfer functions f and the value of the output variable ge_k^O are related by the following equation:

$$ge_k^O = f^2 \left(\sum_{j=1}^N W_{kj} f^1 \left(\sum_{i=1}^M W_{ji} x_i + b_j^1 \right) + b_k^2 \right) \quad (6.2)$$

Where: $j = 1, \dots, N$; $i = 1, \dots, M$; $k = 1, \dots, K$

The weight W_{ji} , have a source on the i th input (x_i), and a destination on the j th neurons on the hidden layer (layer 1). The weight W_{kj} , have a source on the j th neuron of the hidden layer, and a destination on the k th neurons of the output layer (layer 2). Superscript 1 indicates that the bias b , or the activation function f is associated with the first layer. Superscript 2 indicates layer 2. The glucose estimated in the k th neuron of the output (o) layer is represented by ge_k^o .

The training procedure requires a set of inputs and outputs, so that a particular input leads to a specific target output. During the training process, weights and biases are iteratively adjusted to minimize a performance function (F), which is expressed in terms of the mean of squared error (mse) that exist between the neural network estimated output and the specified targets. Typically, many input / target pairs are needed to train a network.

One of the problems that occur during neural network training is called overfitting. The error on the training set is driven to a very small value, but when new data is presented to the network the error is large. The network has memorized the training examples, but it has not learned to generalize to new situations. There are three alternatives to improving generalization²⁰:

- 1) Use a small enough network to provide an adequate fit and not overfit the data.
- 2) Use the early stopping method, in which case it is necessary to divide the available data into three subsets, training, validation and test sets.
- 3) Use the regularization method.

In the present work, the regularization method was used in order to prevent overfitting. According to Beale²⁰, when the data set is small, Bayesian regularization provides better generalization performance than early stopping.

The implementation of the Bayesian regularization method for improving generalization requires modification of the usual performance function: $F=mse$, by using, besides the mean of squared error (mse), the mean of the sum of squares of

the network weights and biases (m_{swb}). Thus, the performance function equation used in Bayesian regularization is:

$$F = \gamma.mse + (1 - \gamma).m_{swb} \quad (6.3)$$

Where:

$$mse = \frac{1}{N} \sum_{k=1}^N (t_k - ge_k^O)^2 \quad (6.4)$$

$k = 1, \dots, N$

t_k is the k th *target output value*.

ge_k^O is the k th *net estimated output value*.

N is the number of *target output values*.

$$m_{swb} = \frac{1}{(n + m)} \left(\sum_{u=1}^n W_u^2 + \sum_{v=1}^m b_v^2 \right) \quad (6.5)$$

W_u = weight values. $u = 0, 1, \dots, n$

b_v = bias values. $v = 0, 1, \dots, m$

n = total number of weights, m = total number of bias.

γ is the performance ratio, its optimal value is set by the Bayesian regularization training algorithm.

6.2.4.2. THE DATA SET FOR DYNAMIC MODELING BASED ON ANN

The inputs for the neural network were cellulase (CBU L⁻¹) and β -glucosidase (FPU L⁻¹) concentrations, as well as the reaction time (min). The output was glucose concentration (gL⁻¹). The data set for modeling based on ANN contains an experimental data set (data set A) and two interpolated data sets (data set B and data set C) with a total of 355 input/output patterns. 224 input/output patterns (63%) were used for training and 131 input/output patterns (37%) were used for validation.

6.2.4.3. THE EXPERIMENTAL DATA SET A

In this data set the concentrations of the inputs (cellulase and β -glucosidase) were determined according to the experimental runs of a central composite design (CCD) 2^2 plus star (Table 6.2). The glucose concentration (output) was measured at 14 different reaction times during the evolution of the enzymatic hydrolysis for each experimental run. The ninth experimental run contains the average of three observations for the central point of the experimental design.

The experimental data set A contains 126 input /output patterns distributed in 9 runs. The third run was used as the validation data set 1 (14 input /output patterns), the remaining 112 input /output patterns distributed in 8 experimental runs were part of the training data set. The training data set was integrated by experimental data (50%) and interpolated data (50%) in a total of 224 input/output patterns.

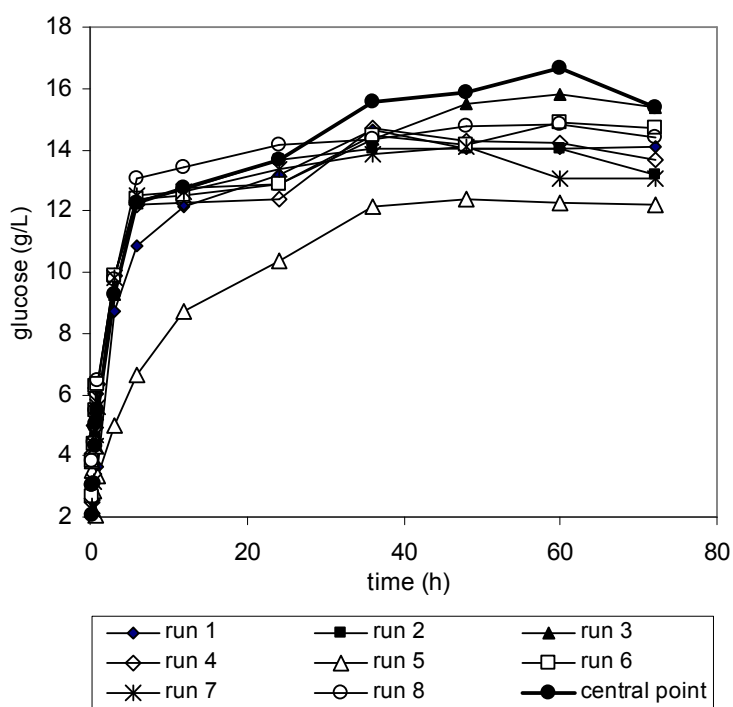


Figure 6.1. Experimental behaviour of glucose (g L^{-1}) for runs in the design matrix on 14 different times

Figure 6.1 shows the evolution of glucose concentration for all the runs of the experimental design. The best result for glucose concentration (16.685 gL^{-1}) corresponds to the ninth experimental run for a reaction time of 60 h. The fifth experimental run contains the lowest glucose concentration for each sampling time. In none of the experimental runs a significant increase in glucose concentration after 60 hours was identified, so it is estimated that this reaction time is very close to the end of the reaction.

6.2.4.4. THE INTERPOLATED DATA SET B

The use of the central composite experimental design $2^2 + \text{star}$ to obtain information of the dynamics of the process have the advantage that less experimental effort is needed, however the limited number of experimental points and the locations of these points in the experimental design space increase the probability of the neural network model to give unrealistic trends between experimental points²¹. Therefore, in order to improve the prediction ability of the ANN model at the center region, 8 additional runs were considered for intermediate points around the central point of the experimental design, each run provides 8 additional interpolated values of the output (glucose gL^{-1}) for each of the above mentioned 14 sampling times. Thus, by using a procedure of interpolation based on ANN, 112 additional input/output patterns (data set B) were obtained, which were integrated to the training data set.

6.2.4.5. THE INTERPOLATED DATA SET C

This data set contains interpolated values of glucose (g L^{-1}) which correspond to the following 13 intermediate times between sampling times: 15 min, 25 min, 35 min, 45 min, 55 min, 1.5 h, 4.5 h, 9 h, 18 h, 30 h, 42 h, 54 h, and 66 h; for each one of the 9 experimental runs. Therefore, data set C contains 117 input/output patterns and was used as the validation data set 2.

6.2.4.6. ARTIFICIAL NEURAL NETWORK BASED INTERPOLATION

Lanouette, et al.²¹ suggested the use of regularization techniques to improve the process modeling with neural networks when using small experimental data sets. One of these suggested techniques was used as interpolation procedure in this study. A neural network with topology 3:18:1 and the experimental data set (data set A) with 126 input/output patterns were used to predict unknown outputs (interpolated points). The interpolation procedure consists of repeating the identical learning procedure a number of times using different initial weights and then to calculate an average value of the predicted outputs. In the present study the averages of 100 repetitions were used as interpolated values of glucose (g L^{-1}) in data sets B and C.

6.2.5. SELECTION OF THE BEST DYNAMIC MODEL BASED ON ANN FOR ENZYMATIC HYDROLYSIS

The appropriate number of neurons in the hidden layer was found using the cross-validation technique in order to avoid model overfitting and to achieve good generalization from the training data set. This technique splits the data sample into a training data set and a validation data set. Then, neural network models with different numbers of neurons in the hidden layer are trained with the training data set and their generalization capabilities are evaluated by mean of predictions of the validation data set. In this study the appropriate number of neurons in the hidden layer was determined with the validation data set 1 using as criterion the smaller average obtained by adding the mean of squared error (mse) and the mean of the sum of squares of the network weights and biases ($mswb$).

$$sa = \frac{(mse + mswb)}{2} \quad (6.6)$$

Where: sa is the smaller average, mse was calculated by equation (6.4) and $mswb$ was calculated by equation (6.5).

6.2.6. OPTIMIZATION USING THE DYNAMIC MODEL BASED ON NEURAL NETWORKS

The optimization methods used to find the optimum of glucose concentration in the dynamic model based on neural networks were the following:

- 1) Genetic algorithms (GA). It is a stochastic global optimization technique that belongs to one of the most active areas of artificial intelligence. The first book on genetic algorithms was written by Holland²², then contributions were made by others researchers, among them Goldberg²³ and Michalewicz²⁴. The procedure is broadly detailed in the literature²⁵ and involves the use of a population of input values (individuals), which are subject to evaluation using a fitness function. If the optimization criteria are met, the best individuals give the solution. If not, the best individuals are selected and from them we obtain new generations using genetic operators for mutation and crossover. Finally, the global maximum or minimum is obtained. The new population of potential solutions that replace the former contains more and better features than the previous one and therefore allow convergence in an optimal solution of the problem within the search space.

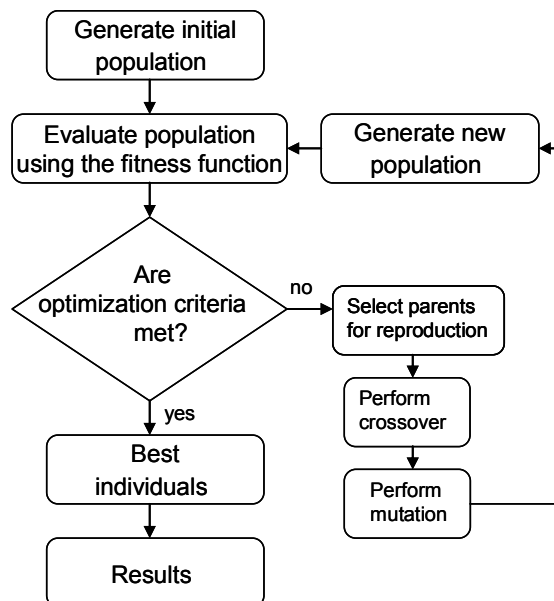


Figure 6.2. Structure of the basic genetic algorithm

- 2) Simulated annealing. It is a meta-heuristic search algorithm for global optimization problems, which is explained by Kirkpatrick²⁶.
- 3) Threshold acceptance. This algorithm is a variation of Simulated Annealing and was proposed by Dueck and Scheuer²⁷.

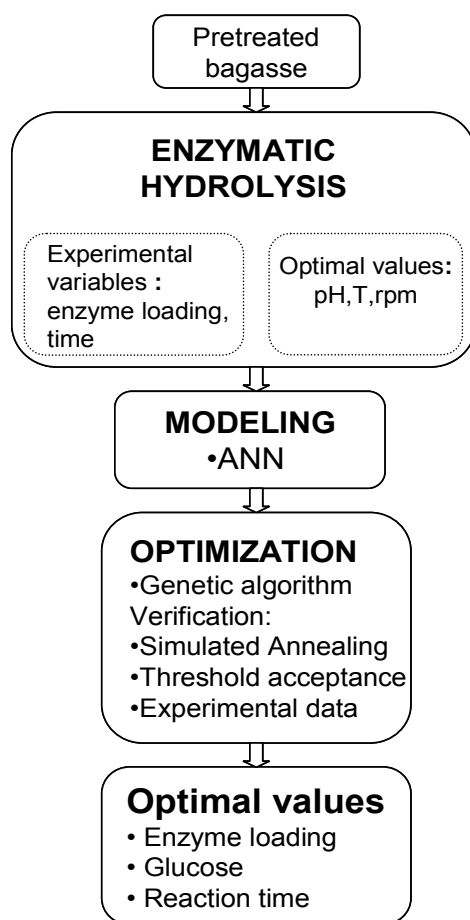


Figure 6.3. Modeling and optimization based on AI techniques for enzymatic hydrolysis

Figure 6.3 resumes the modeling and optimization based on artificial intelligence (AI) techniques for the load of enzymes on the enzymatic hydrolysis of sugarcane bagasse.

The modeling and optimization were developed using the Neural Network Toolbox and the Genetic Algorithm and Direct Search Toolbox of MATLAB® software.

6.3. RESULTS AND DISCUSSION

6.3.1. SELECTION OF THE BEST NEURAL NETWORK FOR DYNAMIC MODELING

Neural networks of configurations 3:N:1, where N is the number of neurons in the hidden layer ($N \in [12, 25]$), were tested in order to select the best neural network based model of the process. A neural network topology 3:20:1 was selected, because this configuration presented the lowest values of *mse*, *mswb* (Equation 6.5) and *sa* (Equation 6.6). From the results, it was observed that *mswb* is a better criterium than R^2 for testing the generalization capacity of a neural network based model, because high values of R^2 could be associated with levels of overfitting, therefore with lower generalization capacity. Figure 6.4 shows the optimal ANN structure. The values of the criteria parameter: *mse*, *mswb* and *sa*; are shown in Table 6.3. The weights and biases are presented in Table 6.4.

Table 6.3. Values of the criteria parameters to select the best neural network model using the validation data set 1.				
Nº of neurons	Mse	R^2	Mswb	Sa
16	0.293	0.984	8.879	4.586
17	0.286	0.992	7.929	4.108
18	0.159	0.996	8.100	4.130
19	0.268	0.996	7.347	3.808
20	0.107	0.996	7.208	3.672
21	0.369	0.997	10.470	5.420
22	0.268	0.995	8.727	4.498
23	0.207	0.998	9.859	5.033
24	0.157	0.996	9.008	4.583

R^2 = determination coefficient, *mse* = mean square error, *mswb* = mean of the sum of squares of the network weights and biases, *sa* = smaller average

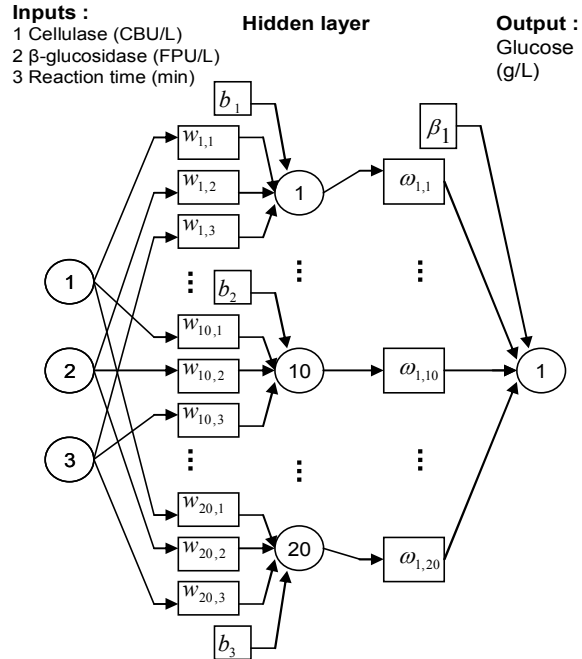


Figure 6.4. Optimal ANN structure

Table 6.4. Weight and biases in the optimal ANN structure							
Hidden layer					Output layer		
Weights (W_{ij})	Source (inputs) : $J = 1,2,3$			Bias	Weights (W_{ij})	Bias	
Destination neurons : $I = 1 \dots 20$	$J = 1$	$J = 2$	$J = 3$	B_{i1}	Source Neurons: $J = 1 \dots 20$	Destination Neuron 1 $i = 1$	$B_{11} = 0.240568$
1	2.887349	-0.126027	-0.13979	-1.716624	1	1.963993	
2	-3.078070	0.133617	0.456221	-2.140195	2	3.128506	
3	0.736484	-0.113152	-10.74188	-12.59915	3	-9.885024	
4	1.555672	0.117979	-5.790305	-6.069625	4	-4.357880	
5	-0.106578	1.059486	-2.841512	2.903664	5	2.193869	
6	0.928131	0.176929	-3.299657	-0.446936	6	4.028185	
7	-0.267750	0.174052	-0.245136	-0.064412	7	0.537941	
8	-3.246703	-0.150519	1.745204	-0.356010	8	2.317011	
9	1.531672	0.609076	1.052487	-0.556734	9	2.175127	
10	-2.904962	-0.374612	-1.178366	-1.188146	10	-2.681491	
11	-0.684839	-0.570582	-0.043015	-0.459031	11	0.848389	
12	0.929529	-1.014153	-2.385174	-0.889401	12	-1.813388	
13	-1.151492	2.750053	-0.259380	-1.318910	13	-1.306382	
14	-0.340389	0.238888	-0.279015	-0.151141	14	0.680965	
15	2.874219	0.112451	-2.376671	-2.087021	15	-3.506905	
16	3.066710	0.123564	4.219750	4.522366	16	3.012662	
17	-4.850232	-0.272447	-8.651839	-11.859123	17	-3.983665	
18	0.260890	1.977195	-1.407222	1.627312	18	-2.050918	
19	1.811436	1.092898	-2.815696	2.477191	19	1.768837	
20	-0.143841	-0.090744	6.121499	5.342421	20	-7.461825	

6.3.2. ESTIMATED RESPONSES WITH THE DYNAMIC MODEL BASED ON ANN

Figure 6.4 shows the results of the training and validation stages for the dynamic model based on ANN. Figure 6.4a presents the parity plot between the values of the output variable (glucose g L^{-1}) predicted by the selected neural network and the values of the same variable in the training data set (targets) when the training stage is completed. The value of the determination coefficient is $R^2 = 0.997$. According to Figure 6.4a ANN learning is best for values that are in the top half of the distribution of values. The value of the mean square error for values less than 8 gL^{-1} of glucose is $mse = 0.1265 \text{ g}^2 \text{ L}^{-2}$, while for values greater than 8 gL^{-1} and lower than 15 gL^{-1} is $mse = 0.0172 \text{ g}^2 \text{ L}^{-2}$, and for values greater than 15 gL^{-1} is $mse = 0.0097 \text{ g}^2 \text{ L}^{-2}$. This is a good result for optimization purposes because the values of glucose concentration are greater than 15 gL^{-1} in the optimum region.

Figure 6.4b, presents the parity plot for glucose (g L^{-1}) predicted by the selected ANN and experimental values of glucose (g L^{-1}) in the validation set 1 (run 3), these values were not considered in the training stage. The determination coefficient is $R^2 = 0.996$, which means that the optimal ANN model explain 99.6% of the variability of the experimental data of the validation data set 1. The parity plots (Figures 6.4a and 6.4b) show a similar accuracy between the results obtained with both the training data set and the validation data set 1.

Figure 6.4c presents the parity plot of the predicted values for glucose (g L^{-1}) obtained with the optimal ANN and the values of the same variable that are contained in the validation set 2 (data set C), the determination coefficient is $R^2 = 0.999$. As was already mentioned, data set C contains 117 data points of glucose (g L^{-1}) for intermediate points between sampling times, which were unknown to the selected ANN in the training stage. The parity plots (Figures 6.4a and 6.4c) show a similar accuracy between the results obtained with both the training data set and validation data set 2.

The performance in the validation stage is crucial to test the model generalization capacity. The results in Figure 6.5 shows that the optimal ANN

presented a good performance in the validation stage, therefore the model has a good generalization capacity.

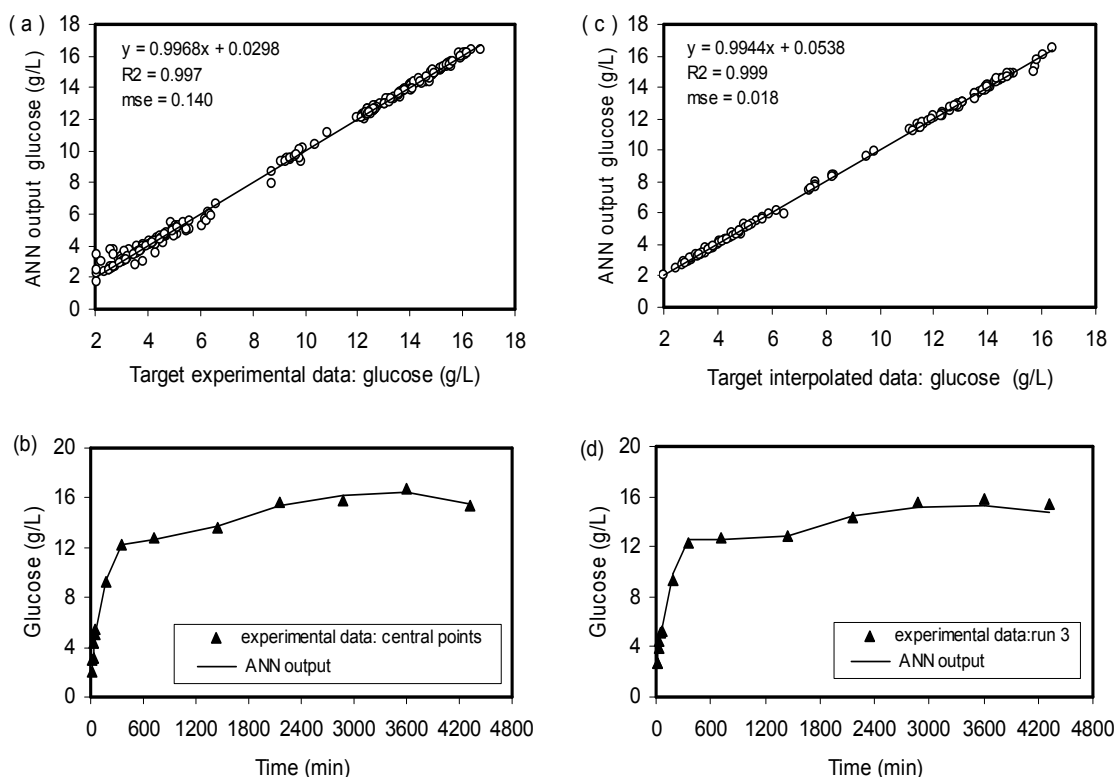


Figure 6.5. ANN output and target output after training and validation. (a) , (b) training, (c) validation with interpolated data points; d) validation using experimental data: run 3.

The experimental data in case (b) of Figure 6.5 corresponds to the average of three replicates at the center point of the experimental design.

6.3.3. IDENTIFICATION OF THE OPTIMUM REGION USING THE MODEL BASED ON ANN

In order to visualize the identified optimum region, Figure 6.6 presents response surfaces (3-D) and contour plots (2-D) plotted using the dynamic model based on ANN developed for glucose prediction (g L^{-1}) in the enzymatic hydrolysis of lime pretreated sugarcane bagasse.

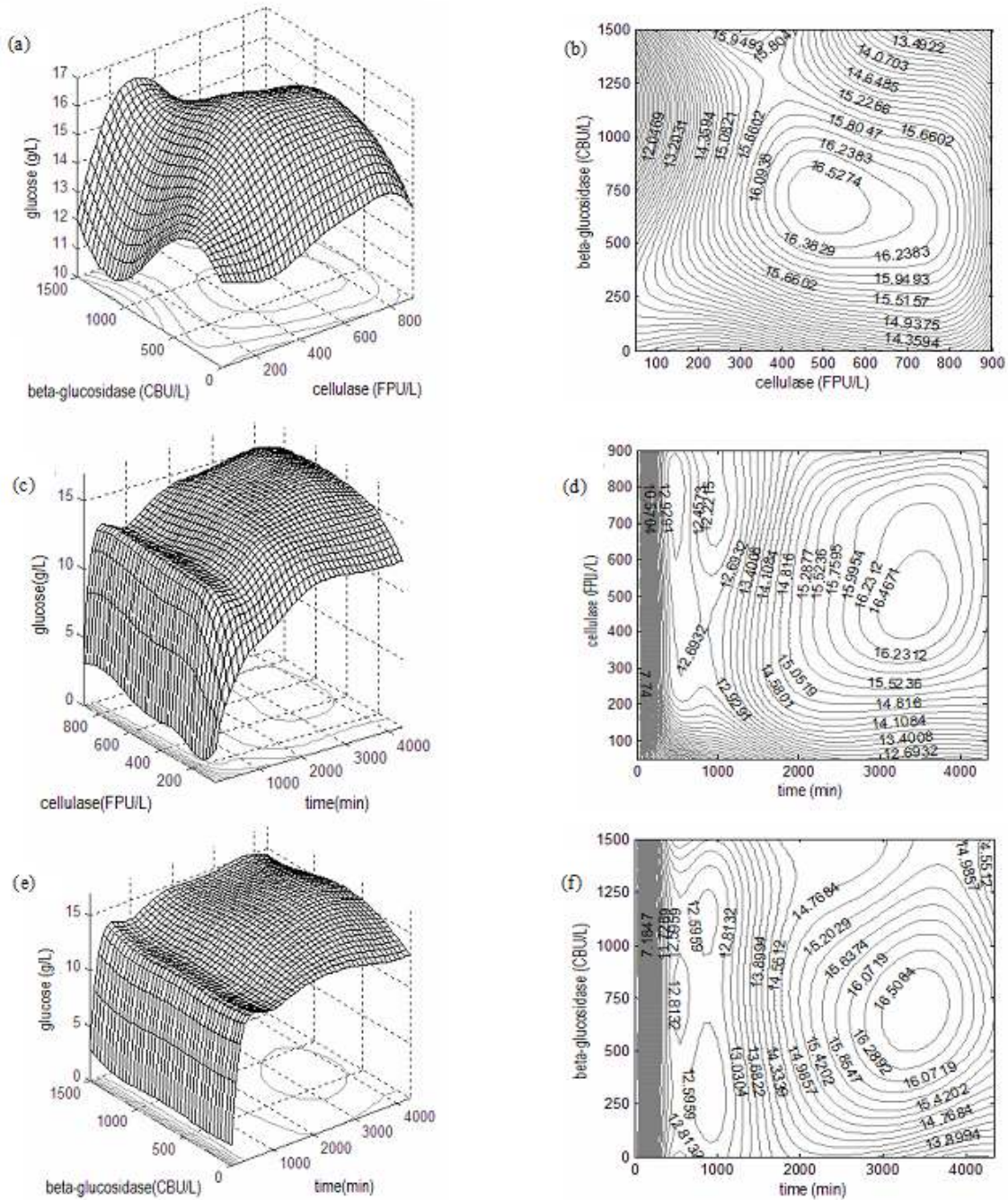


Figure 6.6. Optimal region for the maximum glucose production in the enzymatic hydrolysis of lime pretreated sugarcane bagasse: (a), (c) and (e) response surfaces and (b), (d) and (f) contour plots of the ANN based dynamic model. Constant factors: (a) and (b) time = 60 h (3600 min), (c) and (d) β -glucosidase = 750 CBU/L, (e) and (f) cellulase = 475 FPU/L; pH = 4.8 and T = 50 °C for all the experiments.

Figure 6.6a and 6.6b show the relative effects of different loadings of cellulase and β -glucosidase for a reaction time of 60 h (3600 min). Figure 6.6c and 6.6d show the relative effects of the reaction time in the range of 0.167 h (10 min) to 72 h (4320 min) and different concentrations of cellulase (FPU L^{-1}) while β -glucosidase is maintained constant at 750 CBU L^{-1} . Figure 6.5e and 6.5f show the relative effects of the reaction time in the same range above mentioned and different concentrations of β -glucosidase (CBU L^{-1}) while cellulase is maintained constant at 475 FPU L^{-1} .

From these graphics, an estimation of the optimal region for the maximum glucose production can be made: cellulase in the range $420 - 600 \text{ FPU L}^{-1}$, β -glucosidase in the range $540-850 \text{ CBU L}^{-1}$ and the optimal reaction time is in the range 3000 min (50 h) - 3720 min (62 h).

Rivera⁷ studied the effects of cellulase (FPU L^{-1}) and β -glucosidase (CBU L^{-1}) loads on glucose yield (%) after enzymatic hydrolysis of sugarcane bagasse pretreated with hydrogen peroxide and sodium hydroxide. Temperature 50°C , substrate concentration 3% (w/v) and pH 4.8, were used for enzymatic hydrolysis. The results show that yields of glucose above 99% after hydrolysis could be obtained with cellulase in the range $460 - 580 \text{ FPU L}^{-1}$ and β -glucosidase in the range $750 - 1140 \text{ CBU L}^{-1}$, at a reaction time of 72 h.

In spite of using a different pretreatment it is observed that the ranges above mentioned for the optimum region fully overlap in case of cellulase concentration and partially overlap in case of β -glucosidase concentration with the identified ranges in the present research.

6.3.4. OPTIMIZATION BY INTEGRATING AI TECHNIQUES

Table 6.5 shows that the optimal dynamic model based on ANN gives estimated values of glucose (g L^{-1}) inside the ranges of the experimental error.

Table 6.5. Estimated global optimum using different methods for enzyme loading and glucose production in enzymatic hydrolysis of sugarcane bagasse pretreated with lime.

	Method	Glucose (g L ⁻¹)	Cellulase (FPU L ⁻¹)	β-glucosidase (CBU L ⁻¹)	Time -min-
Global Optimum	ANN + genetic algorithm	16.725	481.540	688.227	56.734 h
	ANN + simulated annealing	16.725	485.342	689.775	56.853 h
	ANN + threshold acceptance	16.725	481.688	688.594	56.747 h

The 4-dimensional hyperspace defined by the ANN based model was explored using genetic algorithms and other optimization algorithms to verify the obtained results. The global optimization results obtained using genetic algorithms, simulated annealing and threshold acceptance were similar in both glucose produced (g L⁻¹) and combination of enzymes in the load, and are shown in Table 6.5. From these results it is clear that genetic algorithms were capable of finding efficiently the global optimum.

The solution for the global optimum is achieved when the best combination of input values that leads to the highest value of the output is found in the search space. According to the results in the optimal point, 16.725 g L⁻¹ of glucose (a yield of 83 %) can be obtained in around 57 hours of hydrolysis using concentrations of cellulase and β-glucosidase of around 482 FPU L⁻¹ (16.1 FPU g⁻¹ bagasse) and 688 CBU L⁻¹ (22.9 CBU g⁻¹ bagasse), respectively.

A literature review shows that in enzymatic hydrolysis, the yield in glucose and processing times vary according to: the pretreatment used, the enzyme loading, the nature and amount of the raw material used and the conditions in which hydrolysis is performed.

According to Fuentes²⁸, which studied the influence of solid substrate concentration in the hydrolysis of sugarcane bagasse pretreated with lime using 1,2,3,4,5% (w/v) of substrate concentration, glucose yield of 97.8 % and glucose concentration of 6.5 g L⁻¹ were obtained with 1% of initial substrate concentration for a reaction time of 60 h, 500 FPU L⁻¹ of cellulase and 500 CBU L⁻¹ of β -glucosidase.

Mosier²⁹, reported that similar results on enzyme digestibility of the cellulose are obtained if the pretreatment of sugarcane with lime is realized by using higher temperatures (e.g. 120 °C) with shorter reaction times (1- 4 h) or lower temperatures (e.g. 25 °C) with longer reaction times (up to 192 h). They obtained glucose yield of 88% of cellulose as glucose after 7 days at an enzyme loading of 25 FPU per gram of biomass.

According to Zhao³¹, pretreating bagasse with 10% NaOH at 90 °C for 1.5 h and further delignifying with 10% peracetic acid at 75 °C for 2.5 h, led to a yield of reducing sugars of 92.04 % and a glucose yield of 56.23% after enzymatic hydrolysis for 120 h with cellulase loading of 15 FPU g⁻¹ solid.

6.4. CONCLUSIONS

The neural network model with topology 3:20:1, with 3 neurons in the first layer, 20 neurons in the hidden layer, and 1 neuron in the output layer, is able to describe successfully the dynamic behavior of glucose concentration in the enzymatic hydrolysis of sugarcane bagasse.

Although the results obtained for the optimal set are specific to the pretreatment, raw material, types of enzyme used and enzymatic hydrolysis conditions, the application of the same methodology to other operational conditions is straightforward. The method of dynamic modeling based on artificial intelligence techniques provided response surfaces of glucose concentration as a result of the action of enzymes for any hydrolysis time until the final reaction time, with far less effort than would be required if the classic RMS based on experimental designs

was used. In this last method, to obtain a response surface an experimental design is required for each hydrolysis time, thus, usually only static results at the estimated final time of the reaction are obtained.

Acknowledgments

The authors acknowledge CAPES, CNPq, Fapesp and OEA for financial support.

6.5. REFERENCES

- 1 Smeets E, Junginger M, Faaij A, Walter A and Dolzan P, *Sustainability of Brazilian bio-ethanol*.
http://www.therenewablecorp.com/resources/white_papers/4.%20Sustainability%20of%20Brazilian%20Ethanol%20-%20Copernicus%20Institute.pdf.
[accessed 10 December, 2009].
- 2 Schlittler LAFS, Engineering of a bioprocess for the production of ethanol from sugar cane. MSc thesis, Federal University of Rio Janeiro, Brazil (2006).
- 3 Farone WA and Cuzens JE, Method of Separating Acids and sugars resulting from Strong Acid Hydrolysis. U.S. Patent 5 580 389 (1996).
- 4 Laopaiboon P, Thani A, Leelavatcharamas V, Laopaiboon L, Acid hydrolysis of sugarcane bagasse for lactic acid production. *Bioresour Technol* 101:1036–1043 (2010).
- 5 Castro RJH, Saccharification of hemicellulose from sugarcane bagasse and its fermentation by *Pachysolen tannophilus*. PhD thesis, State University of Campinas, SP, Brazil (1985).
- 6 Philippidis GP, Smith TK and Wyman CE, Study of the enzymatic hydrolysis of cellulose for production of fuel ethanol by the simultaneous saccharification and fermentation process. *Biotechnol Bioeng* 41:846–853 (1993).
- 7 Rivera EC, Rabelo SC, Garcia D, Filho R and Costa AC, Enzymatic hydrolysis of sugarcane bagasse for bioethanol production: determining optimal enzyme loading using neural networks. *J Chem Tech Biotechnol* 85(7): 983-992 (2010).
- 8 Zhang Q, Collection of *Trichoderma reesei* cellulase by foaming. PhD thesis, University of Akron, Ohio, USA (2007).

- 9 Isar J, Lata A, Saurabh S and Rajendra SA, Statistical method for enhancing the production of succinic acid from *Escherichia coli* under anaerobic conditions. *Bioresour Technol* 97: 1443–1448 (2006).
- 10 Kim JK, Oh BR, Shin HJ and Eom CY, Statistical optimization of enzymatic saccharification and ethanol fermentation using food waste. *Process Biochem* 43: 1308-1312 (2008).
- 11 Rabelo SC, Maciel Filho R and Costa AC, Lime pretreatment of sugarcane bagasse for bioethanol production. *Appl Biochem Biotechnol* 153: 139-150 (2009).
- 12 Nagata Y and Chu KH, Optimization of a fermentation medium using neural networks and genetic algorithms. *Biotechnol Lett* 25: 1837–1842 (2004).
- 13 Deniz B and Boyacı IH, Modeling and optimization II: Comparison of estimation capabilities of response surface methodology with artificial neural networks in a biochemical reaction. *J Food Eng* 78: 846–854(2007).
- 14 Dutta JR, Dutta PK, and Banerjee R. Optimization of culture parameters for extracellular protease production from a newly isolated *Pseudomonas* sp. using response surface and artificial neural network models. *Process Biochem* 39, 2193–2198 (2004).
- 15 Sluiter A, Ruiz R, Scarlata C, Sluiter J and Templeton D, National Renewable Energy Laboratory, Golden, CO, USA (2005).
<http://www.nrel.gov/biomass/pdfs/42619.pdf> [accessed 5 March 2009].
- 16 Sluiter A, Hames B, Ruiz R, Scarlata C, Sluiter J, and Templeton D, National Renewable Energy Laboratory, Golden, CO, USA (2005).
<http://www.nrel.gov/biomass/pdfs/42622.pdf> [accessed 24 March 2009].
- 17 Sluiter A, Hames B, Ruiz R, Scarlata C, Sluiter J, Templeton D, and Crocker D, National Renewable Energy Laboratory, Golden, CO, USA (2008).
<http://www.nrel.gov/biomass/pdfs/42618.pdf> [accessed 7 April 2009].
- 18 Ghose TK, Measurement of cellulose activities. *Pure Appl Chem* 59, 257-268(1987).
- 19 Wood TM and Bhat KM, Methods for measuring cellulase activities, in *Methods in Enzymology*, ed by Wood WA and Kellog ST, Academic Press, San Diego, CA, vol 160, pp 81-112 (1988).
- 20 Beale MH, Hagan MT, Demuth HB, *Neural Network Toolbox User's Guide*.
http://www.mathworks.com/help/toolbox/nnet/nnet_product_page.html
[accessed 14 January, 2010].

- 21 Lanouette R, Thibault J, Valade JL, Process modeling with neural networks using small experimental datasets. *Comput Chem Eng* 23: 1167-1176 (1999).
- 22 Holland JH, *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor, Michigan (1975).
- 23 Golberg DE, *Genetic Algorithms in Search Optimization and Machine Learning*. Addison-Wesley, Reading, Massachusetts (1989).
- 24 Michalewicz Z, *Genetic Algorithms + Data Structures = Evolution programs*. Springer, New York (1996).
- 25 Alvarez VH, Larico R, Ianos Y and Aznar M, Parameter estimation for VLE calculation by global minimization: The genetic algorithm. *Braz J Chem Eng* 25: 409-418 (2008).
- 26 Kirkpatrick S, Gelatt CD and Vecchi MP, Optimization by Simulated Annealing. *Science*, New Series 220: 671-680 (1983).
- 27 Dueck G and Scheuer T, Threshold Accepting: A General Purpose Optimization Algorithm Appearing Superior to Simulated Annealing. *J Comput Phys* 90: 161-175 (1990).
- 28 Fuentes LG, Determination of kinetic data of sugarcane bagasse deslignification and of the enzymatic hydrolysis with pretreatment with calcium hydroxide. MSc thesis, State University of Campinas, SP, Brazil (2009).
- 29 Mosier N, Wyman C, Dale B, Elander R, Lee YY, Holtzapple M and Ladisch M, Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresour Technol* 96: 673–686 (2005).
- 30 Karr WE, Holtzapple T, Using lime pretreatment to facilitate the enzymatic hydrolysis of corn stover. *Biomass and Bioenergy* 18: 189–199 (2000).
- 31 Zhao X, Peng F, Cheng K and Liu D, Enhancement of the enzymatic digestibility of sugarcane bagasse by alkali–peracetic acid pretreatment. *Enzym Microb Technol* 44: 17–23 (2009).

CAPÍTULO 7. MODELAGEM EMPÍRICA DE PROCESSOS: UM PROCESSO DE FERMENTAÇÃO CONTINUA COM EVAPORADOR FLASH COMO CASO DE ESTUDO

7.1. INTRODUÇÃO

Uma revisão detalhada do processo de fermentação alcoólica extrativa foi mostrada no capítulo 2. Neste capítulo, a modelagem empírica desse processo a partir de dados de operação de uma planta experimental será estudada. Um modelo fenomenológico que descreva o comportamento do processo não está disponível pela dificuldade de modelagem do equilíbrio líquido-vapor no tanque flash, assim serão usadas redes neurais artificiais.

A identificação de processos utilizando redes neurais é uma alternativa que permite determinar o modelo do processo só a partir de dados experimentais das variáveis de entrada e saída. O modelo empírico de processo obtido é uma ferramenta valiosa na análise e no controle de processos, porque permite determinar os valores que as variáveis de saída adotam para diferentes valores das variáveis de entrada.

O modelo empírico de processo faz possível analisar o comportamento do processo por simulação para diferentes condições de operação estabelecidas pelos valores das variáveis de entrada de forma mais rápida, segura e econômica do que realizar testes em uma planta real. No controle de processos, o modelo de processo fornece uma alternativa de controle sugerida na literatura para processos complexos (SEBORG *et al.*, 2004), em que o modelo de processo incorporado à lei de controle permite otimizar as ações de controle para rejeitar perturbações e estabilizar o estado operacional do processo segundo o comportamento desejado.

Entretanto, como os dados são obtidos sempre através de instrumentos físicos que apresentam um nível de precisão, estão naturalmente corrompidos por erros que podem ser aleatórios ou grosseiros. Por outro lado, durante a operação do processo existem diversas variabilidades que afetam a qualidade dos dados, fazendo que sejam portadores de informações erradas do comportamento dinâmico do processo. Portanto, para garantir que os dados forneçam informações precisas e retratem o processo de forma fidedigna, é necessário utilizar procedimentos de pré-tratamento (retificação) de dados (PRATA, 2009). No

entanto, esses procedimentos devem ser utilizados adequadamente para não remover informação útil da variação de interesse.

Segundo Cinar *et al.*, (2005), os *outliers* (erros grosseiros anômalos) em dados e o ruído do sinal podem ter uma forte influência sobre os valores dos parâmetros e a estrutura dos modelos empíricos desenvolvidos. Consequentemente, a detecção de erros grosseiros, *outliers*, e o pré-tratamento de dados usando técnicas de reconciliação e suavização de dados são críticos para o desenvolvimento de modelos empíricos precisos. Eliminada a influência dos valores espúrios, um resultado mais robusto pode ser esperado.

Neste artigo, é feita uma aplicação da identificação de processos, incluindo: 1) O pré-tratamento dos dados: detecção de erros grosseiros e *outliers*, reconciliação empírica de dados e *denoising*, 2) A caracterização estática básica dos dados, com base na análise de correlações, 3) A caracterização dinâmica dos dados com base no cálculo de médias e variâncias utilizando janelas móveis de diferente duração, e a identificação dos intervalos de estado estacionário e quase-estacionário para cada variável, 4) O desenvolvimento de dois modelos empíricos baseados em redes neurais artificiais para o monitoramento das concentrações de: etanol no fermentador e etanol condensado obtido na unidade de separação flash. No primeiro modelo são consideradas variáveis de entrada que requerem determinação *off-line* no laboratório. No segundo modelo, os valores das variáveis de entrada consideradas podem ser obtidos com instrumentação *on-line*, portanto são variáveis úteis no desenvolvimento de *software sensors* para monitoramento das concentrações de etanol no processo.

7.2. CASO DE ESTUDO

Como caso de estudo foi utilizado um processo de bancada de fermentação alcoólica extrativa contínua desenvolvido por Atala (2006) no laboratório de Engenharia de Bioprocessos da Faculdade de Engenharia de Alimentos da Universidade Estadual de Campinas, Brasil. Este consiste de quatro unidades interconectadas:

- 1) Fermentador – Unidade onde se produz a transformação dos açúcares em etanol por ação da levedura (*Saccharomyces cerevisiae*), possui um sistema de resfriamento para manter temperatura constante através de circulação da água da rede a temperatura ambiente em seu encamisamento.
- 2) Filtro - O filtro é uma unidade utilizada para a reciclagem das células da levedura, consiste de uma fina membrana permeável suportada em material cerâmico, que retém as células e permeia outras partículas menores dissolvidas ou suspensas no caldo fermentativo, incluindo o solvente.
- 3) Tanque flash - O tanque flash é um equipamento que trabalha em condições de vácuo e onde a extração do etanol do caldo fermentativo ocorre devido a diferença de volatilidades do etanol em relação a outros componentes.
- 4) Condensador – A condensação dos vapores do etanol extraído no tanque flash é feita através de uma armadilha de vácuo (trap) com sistema de refrigeração para uma temperatura de trabalho de - 25°C.

A Figura 7.1 mostra o diagrama esquemático do processo de fermentação contínua com extração flash a vácuo identificado neste estudo. Na figura são mostradas as 17 variáveis tomadas em consideração: 1) Em **negrito e sublinhado**, as variáveis etanol condensado, E_c (g/L), e etanol no fermentador E_f (g/L) ; 2) Em **negrito**, as variáveis cujo valor pode ser conhecido *on-line*: Vazão de entrada no tanque flash, F_{in} (L/h); Temperatura no flash, T_{flash} (°C); Pressão no tanque flash, P_{flash} (mmHg); Temperatura no fermentador, T_{ferm} (°C); Taxa de diluição no fermentador, D (h^{-1}); Turbidez no fermentador, T_b (%); pH no fermentador, pH; Vazão de alimentação, F_o (mL/h); Vazão de purga, F_{pu} (mL/h); Vazão de permeado no filtro; F_p (mL/h); Vazão de retorno ao fermentador, F_{re} (L/h); 3) Sem **negrito e sem sublinhado**, as variáveis cujo valor é determinado *off-line* no laboratório e são as concentrações de: Biomassa (X), Substrato (S), Glicerol (G); e a Viabilidade celular (C_v).

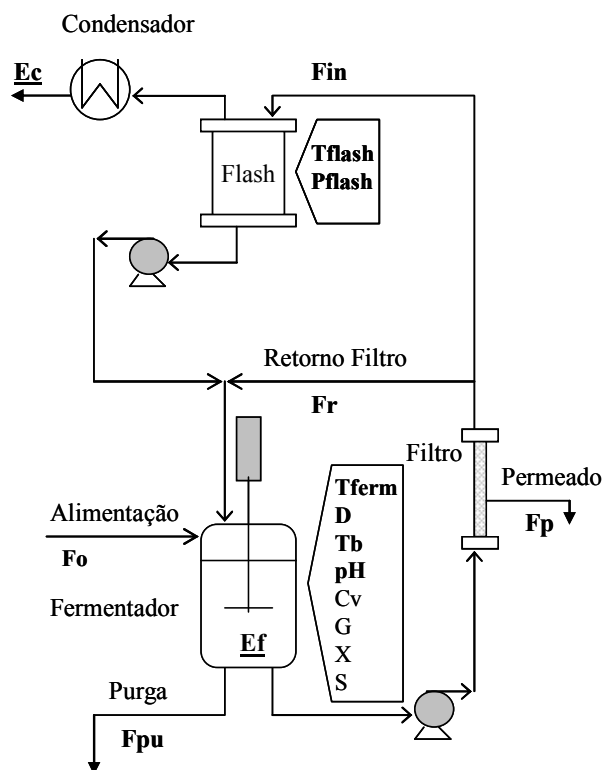


Figura 7.1. Diagrama esquemático de processo de fermentação com unidade de separação flash.

Na etapa experimental, o meio utilizado na fermentação foi o melaço de cana-de-açúcar, cedido por uma usina industrial e contendo cerca de 77% de pureza em açúcar (ART). No processo experimental, com base em resultados de trabalhos anteriores (SILVA, 1998, ATALA, 2001, ATALA 2004) a concentração de substrato na alimentação (melaço de cana-de-açúcar diluído) foi mantida constante em 180 g/L de açúcares redutores totais (ART). A vazão de alimentação foi ajustada para manter o volume constante no sistema durante toda a fermentação.

A etapa inicial de batelada extrativa do processo fermentativo foi iniciada após a adição do inóculo contendo a levedura *Saccharomyces cerevisiae* e outros componentes. O objetivo desta etapa foi promover o consumo total de substrato. Com o fim da etapa batelada, foi iniciada a etapa contínua extrativa pelo acionamento da bomba peristáltica acoplada ao sistema de filtração. A vazão desta bomba foi ajustada para a retirada do mosto filtrado correspondente a uma

taxa de diluição inicial de $0,1 \text{ h}^{-1}$, o que representa um tempo de residência de 10 horas. Quando a concentração celular no fermentador atingiu valores de 10 % v/v, a bomba de purga foi acionada e a vazão da bomba de permeado foi diminuída mantendo a taxa de diluição inicial de $0,1 \text{ h}^{-1}$. A taxa de diluição foi mantida por cerca de 3 tempos de residência (neste caso, cerca de 30 horas). Após o término desta etapa, as vazões de entrada e de saída foram ajustadas para fornecer uma taxa de diluição de $0,15 \text{ h}^{-1}$, o que corresponde a um tempo de residência de 6,67 horas, mantendo as mesmas condições de concentração de substrato na alimentação. Este procedimento foi repetido para as taxas de diluição de 0,20 e $0,35 \text{ h}^{-1}$, que correspondem a tempos de residência de 5 e 2,85 horas, respectivamente. Em cada caso, as taxas de diluição foram mantidas por vários tempos de residência.

O fermentador operou com temperatura entre $33 \text{ }^{\circ}\text{C}$ e $33,5 \text{ }^{\circ}\text{C}$, vazão de alimentação entre 136,83 mL/h e 853,87 mL/h e vazão de purga entre 0,00 mL/h e 165,24 mL/h; para uma concentração de etanol entre 51,19 g/L e 67,28 g/L. A vazão do permeado no filtro esteve no intervalo de 122,1 mL/h a 869,4 mL/h.

O tanque flash operou com temperatura entre $34,19 \text{ }^{\circ}\text{C}$ e $33,38 \text{ }^{\circ}\text{C}$, pressão de vácuo entre 131,66 mmHg e 177,56 mmHg, e vazão de entrada entre 160,57 L/h e 204,86 L/h. A concentração de etanol condensado esteve entre 378,27 g/L e 410,68 g/L.

O tempo total de operação considerado neste estudo é de cerca de 400 h, período no qual o processo atingiu a produtividade máxima de 25 g/L h com uma taxa de diluição de $0,35 \text{ h}^{-1}$, que corresponde a um tempo de residência de 2,85 horas. Esta produtividade é três vezes maior do que a produtividade obtida no processo tradicional.

Durante o tempo de operação, amostras do caldo fermentativo de aproximadamente 30 mL foram recolhidas periodicamente para a realização de análises das concentrações de açúcares, glicerol, massa seca, viabilidade e concentração de álcool. Na solução condensada foram realizadas análises de concentração de etanol e volume recolhido. As concentrações de açúcares, de

etanol e de glicerol foram determinadas através de análises em HPLC modelo Varian 9010.

Os resultados analíticos obtidos das amostras retiradas durante o processo fermentativo complementam os dados armazenados no computador pelo sistema de aquisição do programa de supervisão desenvolvido em LABVIEW versão 6.1 (National Instruments, Austin, TX). Todos estes dados foram obtidos no trabalho de Atala (2006) e disponibilizados pelo autor para uso no presente trabalho.

7. 3. IDENTIFICAÇÃO DE PROCESSOS

A identificação de processos envolve a construção dos modelos com dados de entrada/saída obtidos experimentalmente, sem fazer uso de pressupostos teóricos relativos à natureza do processo (LJUNG, 1996). Portanto, a precisão do modelo empírico depende da qualidade da informação fornecida pelos dados experimentais.

7.3.1. RECONCILIAÇÃO DE DADOS E *DENOISING*

Segundo Cinar *et al.*, (2005), é importante a detecção de erros grosseiros e o pré-tratamento de dados usando técnicas de reconciliação e eliminação do ruído dos sinais para o desenvolvimento de modelos precisos.

Neste estudo, a reconciliação de dados foi realizada utilizando diferentes critérios e informações, tais como: as margens de erro dos instrumentos, a faixa de variação dos valores das variáveis em torno do ponto de ajuste (setpoint), os relatórios de pesquisa que informam as dificuldades e comportamento de equipamentos e processo na etapa experimental. Para a retificação dos erros grosseiros anômalos (*outliers*) e dos erros grosseiros relacionados com mau funcionamento de instrumentos, ou com falhas nos equipamentos de processo, a técnica utilizada foi a filtragem de dados, nos casos isolados foi utilizado um filtro de ordem zero e nos casos de existência do ruído nos sinais, o *denoising* foi feito com filtros digitais chamados de anti-causais (OPPENHEIM AND SCHAFER, 1989), os quais eliminam as distorções de fase altamente não lineares no

processo de filtragem. Na aplicação dos filtros foram avaliadas diversas alternativas de filtragem e as consequências da sua aplicação para evitar a perda de informação.

7.3.2. CARACTERIZAÇÃO BÁSICA DOS DADOS

Esta etapa teve como objetivo determinar os referenciais estatísticos do comportamento das variáveis na operação do processo. Segundo Freitas (2009), para este propósito são úteis os cálculos das médias e das matrizes de covariância e de correlação. No presente trabalho, nesta etapa, se deu maior ênfase à interpretação dos coeficientes de correlação, deixando a interpretação de médias e variâncias para a etapa de caracterização dinâmica.

Os coeficientes de correlação foram calculados utilizando a equação:

$$r_{ij} = \frac{V_{ij}^2}{\sqrt{V_{ii}^2 V_{jj}^2}} \quad (7.1)$$

Onde as variâncias V_{ii}^2 , V_{jj}^2 e covariâncias V_{ij}^2 foram calculadas com a equação:

$$V_{ij}^2 = \frac{\sum_{k=1}^N (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{N-1} \quad (7.2)$$

Onde x_{ik} e x_{jk} são as variáveis no instante k , $\bar{x}_i$ e $\bar{x}_j$ são valores médios das variáveis, e N é o número de amostragens disponíveis.

7.3.3. CARACTERIZAÇÃO DINÂMICA DOS DADOS

Nesta etapa foram investigadas as características dinâmicas do comportamento das variáveis utilizando médias e variâncias calculadas em janelas móveis (FREITAS, 2009). As médias móveis foram obtidas com a seguinte equação:

$$\bar{x}m_{i,j} = \frac{1}{TJ} \sum_{k=i-TJ+1}^i x_{k,j} \quad (7.3)$$

Onde j indica a variável, i indica o instante de amostragem superior, TJ é o tamanho da janela e $\bar{x}m_{i,j}$ é a média móvel da variável x_j na janela móvel i .

As variâncias móveis foram calculadas com a seguinte equação:

$$vm_{i,j}^2 = \frac{1}{TJ-1} \sum_{k=i-TJ+1}^i (x_{k,j} - \bar{x}m_{i,j})^2 \quad (7.4)$$

Onde j indica a variável, i indica instantes de amostragem, TJ é o tamanho da janela, $\bar{x}m_{i,j}$ é a média móvel calculada para a variável j na janela móvel i , x são as variáveis e $vm_{i,j}^2$ representa a variância móvel da variável j na janela de amostragem.

No presente trabalho, foram utilizadas as médias e variâncias móveis para identificação dos estados estacionários, quase estacionários e transientes das variáveis no processo.

Nos diferentes capítulos da tese foram desenvolvidos modelos empíricos de redes neurais artificiais para o processo de fermentação contínua extrativa com flash, cada um para determinados propósitos. Neste capítulo, são desenvolvidos dois modelos de redes neurais, o primeiro inclui variáveis de entrada cuja determinação é *off-line*, e o segundo só inclui variáveis de entrada cuja determinação é *on-line*.

7.3.4. MODELOS EMPÍRICOS BASEADOS EM REDES NEURAIS ARTIFICIAIS

Os modelos empíricos desenvolvidos utilizam variáveis de entrada e saída, segundo esquemas que se mostram nas seguintes figuras:

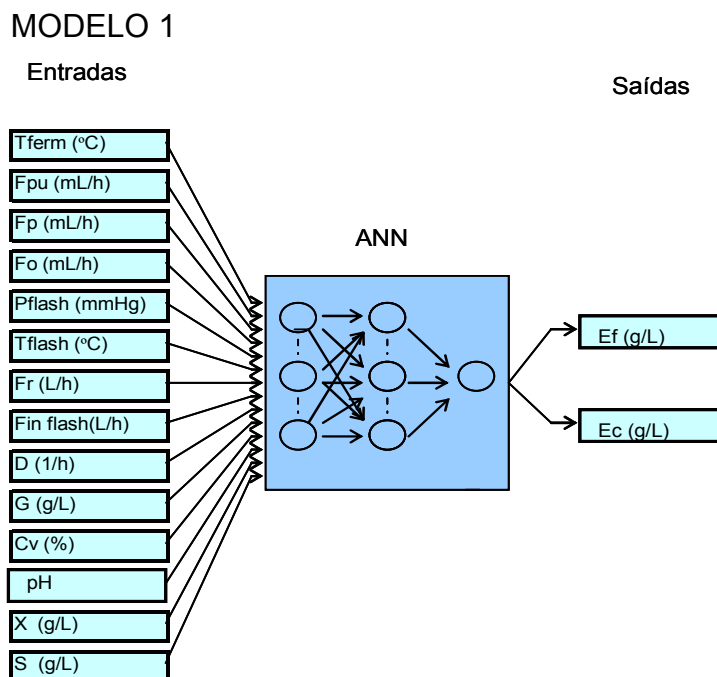


Figura 7.2. Esquema do modelo empírico do processo com variáveis de entrada (G, Cv, X e S) que requerem determinação *off-line*.

As variáveis de entrada do modelo 1 foram utilizadas para descrever o comportamento do processo de fermentação extrativa com flash em estudos realizados por Atala (2001, 2004 e 2006), a maioria delas também estão presentes em diferentes modelos propostos na literatura (Jarzebski, *et al.*, 1989; Silva, 1998; Duarte, 2007; Rivera, *et al.*, 2010), pelo que se considera que são parâmetros importantes da fermentação alcoólica no processo em estudo.

MODELO 2

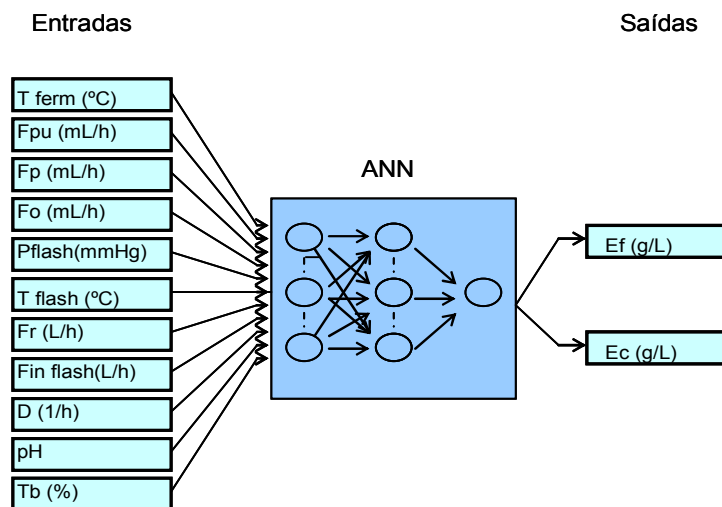


Figura 7. 3. Esquema do modelo empírico do processo com variáveis de entrada que não requerem determinação *off-line*.

Das variáveis de entrada do modelo 2 : Pflash (mmHg), Tferm (°C), Tflash (°C), Fin flash (L/h), D (1/h), pH, Tb (%); já foram propostas para monitoramento das concentrações de etanol no fermentador e etanol condensado em estudos realizados por Rivera *et al* (2010), neste trabalho foram adicionadas as variáveis que descrevem fluxos na operação da planta piloto de fermentação extrativa com flash: fluxo de purga (Fpu), fluxo de permeado (Fp), fluxo de alimentação do fermentador (Fo), e fluxo de retorno ao fermentador (Fr). Todas essas variáveis podem ser monitoradas ON LINE, e se considera que são potencialmente importantes para explicar o comportamento do processo.

Para a modelagem foram utilizadas redes neurais *feedforward* e avaliadas diferentes topologias destas redes com uma camada escondida, representadas por: NE: NCE: NS. Onde NCE é o número de neurônios na camada escondida, NE é o número de entradas e NS é o número de saídas.

As redes *feedforward* são adequadas para modelado de sistemas não lineares complexos, com muitas aplicações em engenharia devido a sua arquitetura de fácil entendimento e implementação (RIVERA, 2010).

A expressão matemática geral dos modelos que relaciona as entradas, x_i , pesos, W_{ji} , bias, b_j^1 , funções de transferência, f , e as variáveis de saída E_f e E_c , é:

$$E_k = f^2 \left(\sum_{j=1}^N W_{kj} f^1 \left(\sum_{i=1}^M W_{ji} x_i + b_j^1 \right) + b_k^2 \right) \quad (7.5)$$

Onde: $j = 1, \dots, N$; $i = 1, \dots, M$; $k = 1, \dots, K$. M é o número de entradas, N é o número de neurônios na camada oculta e K é o número de saídas.

O peso W_{ji} , tem fonte na ***i-ésima*** entrada (x_i), e destino no ***j-ésimo*** neurônio na camada escondida (camada 1). O peso W_{kj} tem fonte no ***j-ésimo*** neurônio da camada escondida, e destino no ***k-ésimo*** neurônio da camada de saída (camada 2). Sobrescrito 1 indica que o bias b , ou a função de ativação f é associada com a primeira camada. Sobrescrito 2 indica camada 2. A concentração de etanol no fermentador é representada por E_1 com $k=1$, e a concentração de etanol condensado é representada por E_2 , com $k=2$.

A técnica da regularização bayesiana (MACKAY, 1992; FORESEE e HAGAN, 1997; DIAS *et al.*, 2003) e o algoritmo de Levenverg – Marquardt foram utilizados para o treinamento das redes. A regularização bayesiana evita o *overfitting* e melhora a capacidade de generalização sem utilizar a técnica de validação cruzada. Neste trabalho, foi utilizado um grupo de teste diferente do grupo treinamento para comparar o desempenho das redes com as diferentes estruturas e selecionar a melhor em cada caso. O grupo de teste foi integrado por

25% de padrões de entrada/saída contendo 25% dos dados experimentais obtidos por medições *off-line* das concentrações de etanol no fermentador e etanol condensado procedente do tanque flash. O total de padrões representativos utilizados foi de 710 que correspondem a aproximadamente 350 h de operação.

A fim de eliminar a interferência das diferentes magnitudes das variáveis no processo de aprendizagem da rede, todos os dados foram normalizados segundo as seguintes equações no intervalo [+1 , - 1] :

$$x_n(i) = \left[\frac{x_{máx} - x(i)}{x_{máx} - x_{mín}} \right] L_{inf} + \left[\frac{x(i) - x_{mín}}{x_{máx} - x_{mín}} \right] L_{sup} \quad (7.6)$$

Para retornar o vetor de saídas normalizadas y_n do modelo após validação às grandezas originais foi utilizada a equação:

$$y(i) = \left[\frac{L_{máx} - y_n(i)}{L_{máx} - L_{inf}} \right] y_{mín} + \left[\frac{y_n(i) - L_{inf}}{L_{máx} - L_{inf}} \right] y_{máx} \quad (7.7)$$

Os dados originais foram retificados e reconciliados a fim de eliminar erros grosseiros e o ruído, que podem distorcer o sentido da variação no comportamento dinâmico da variável.

O número apropriado de neurônios na camada escondida foi determinado utilizando os seguintes critérios computados com os integrantes do grupo de teste:

1) O valor mais baixo do MSE (mean squared error) calculado utilizando os erros de previsão segundo a equação 7.8 (RAGHAVARAO e PADGETT, 2005).

$$MSE = \frac{(Y - \hat{Y})'(Y - \hat{Y})}{n_t} \quad (7.8)$$

Onde: Y é o vetor das n_t respostas que correspondem ao conjunto de teste. $\hat{Y}$ é o vetor de respostas estimadas para o conjunto de teste baseadas no modelo obtido com os n pares entrada-saída utilizados para treinar a rede e determinar seus parâmetros.

2) O maior valor do coeficiente de determinação R^2

3) O maior valor do coeficiente de correlação CORR, calculado utilizando a equação (MILTON e ARNOLD, 1990):

$$CORR = (1 - \frac{SEE}{STT}) * 100 \quad (7.9)$$

Onde: $SEE = \sum_{i=1}^N (y_e(i) - y(i))^2$, $STT = \sum_{i=1}^N (y_e(i) - \bar{y}_e)^2$, $y_e(i)$ é a saída experimental, $y_e(i)$ é a saída da rede, $\bar{y}_e$ é o valor medio das saídas experimentais e N é o número de integrantes do grupo de teste.

4) O menor número de épocas de treinamento.

Os algoritmos foram desenvolvidos em MATLAB.

7.4. RESULTADOS E DISCUSSÃO

7.4.1. PRÉ-TRATAMENTO DOS DADOS

Como foi mencionado antes, foram identificados e retificados os erros grosseiros nos dados e foi realizada a eliminação do ruído, principalmente nos sinais das medições da pressão de operação do tanque flash e na vazão de retorno do filtro para o fermentador. Os resultados se mostram nas seguintes figuras.

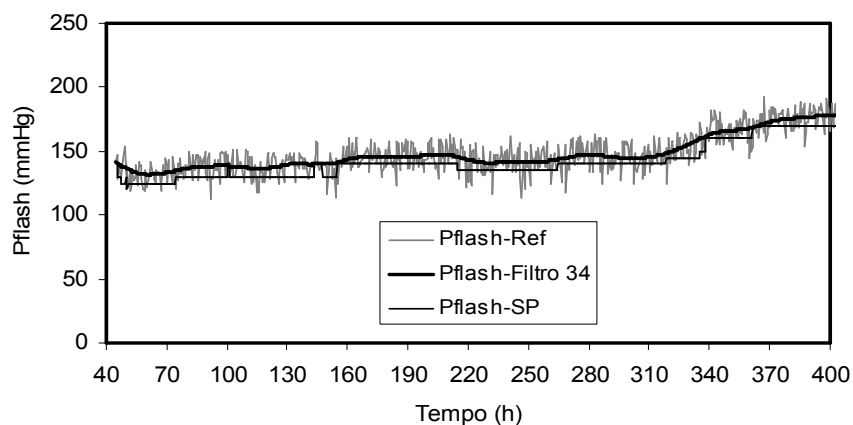


Figura 7.4. Eliminação do ruído do sinal para a pressão de operação do tanque flash, utilizando filtro anticausal de fase zero.

A Figura 7.4 mostra o sinal que fornece informação do comportamento da variável pressão no tanque flash, depois de eliminar o ruído que contém o sinal original utilizado como referência (Pflash-Ref). O sinal filtrado (Pflash-Filtro 34) foi otimizado tendo como referência o set point (Pflash-SP) que estabelece os valores desejados para a pressão de operação no tanque flash. O sinal filtrado possui informação mais precisa sobre o comportamento dinâmico da variável, que é interpretada com ajuda do cálculo das variâncias móveis, conforme mostrado abaixo na caracterização dinâmica.

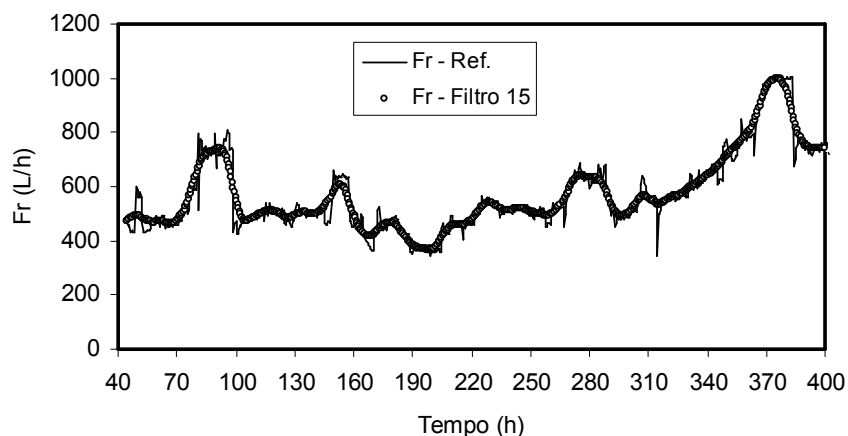


Figura 7.5. Eliminação do ruído do sinal para a vazão de retorno do filtro, utilizando filtro anticausal de fase zero.

A Figura 7.5 mostra o comportamento do sinal que fornece informação da vazão de retorno do filtro após eliminar o ruído (Fr-Filtro 15) e o sinal original utilizado como referência (Fr- Ref).

A Figura 7.5 mostra que o conteúdo de ruído na vazão de retorno do filtro é menor do que o conteúdo de ruído na pressão de operação do tanque flash (Figura 7.4). Neste caso, foram comparados os efeitos na variância móvel de vários filtros alternativos, para obter uma melhora na identificação de características de comportamento dinâmico com menor perda de informação.

7.4.2. CARACTERIZAÇÃO BÁSICA DOS DADOS

Na Tabela 7.1 se apresentam as correlações lineares entre as diferentes variáveis do processo. A correlação linear indica o grau em que duas variáveis

estão associadas ou, em que medida ocorrem conjuntamente. O valor de R varia entre -1 e +1 e quanto mais se afasta de 0, mais forte é a relação.

Nos casos de valores do coeficiente de correlação linear zero ou próximo de zero, a ocorrência conjunta (associação) entre as variáveis é nula ou é muito fraca, nestos casos, não é eliminada a hipótese de que as variáveis podem estar correlacionadas de forma não linear.

Tabela 7.1. Matriz de correlações para as variáveis do processo de produção de etanol por fermentação contínua extrativa com evaporador flash.

Variável	Biomassa (X)	Substrato (S)	Produto 1 (Ef)	Produto 2 (Ec)
Tferm (°C)	0,816	0,126	-0,437	0,411
Fpu (mL/h)	0,655	0,468	-0,655	0,682
Fp (mL/h)	0,675	0,568	-0,729	0,710
Fo (mL/h)	0,705	0,543	-0,713	0,707
Pflash (mmHg)	0,516	0,507	-0,554	0,453
Tflash (°C)	0,824	0,250	-0,536	0,495
F(L/h)	0,217	0,667	-0,628	0,418
Fr (L/h)	0,220	0,672	-0,633	0,423
Fin Flash(L/h)	-0,187	-0,491	0,495	-0,346
D (h ⁻¹)	0,543	0,619	-0,737	0,638
G (g/L)	0,857	0,057	-0,398	0,528
Cv (%)	-0,846	-0,192	0,665	-0,727
Ph	-0,851	-0,053	0,451	-0,535
Tb (%)	0,960	0,226	-0,706	0,746
X (g/L)	1,000	0,202	-0,648	0,648
S (g/L)	0,202	1,000	-0,632	0,317
Ef (g/L)	-0,648	-0,632	1,000	-0,754
Ec (g/L)	0,648	0,317	-0,754	1,000

No sistema, as diversas associações entre variáveis que se apresentam estão limitadas às faixas de variação das variáveis durante a operação do processo, entre elas salientamos as seguintes:

- À medida que o pH aumenta, a biomassa (X) diminui.
- À medida que a biomassa (X) cresce, a viabilidade (Cv) decresce.
- À medida que a concentração do Etanol no produto condensado (Ec) aumenta, a concentração de Etanol no fermentador (Ef) diminui.
- À medida que a concentração de biomassa (X) aumenta, a turbidez (Tb) aumenta.

- À medida que a temperatura no fermentador (T_{ferm}) aumenta, a concentração de biomassa aumenta (X).
- À medida que a vazão de alimentação (F_o) aumenta, a concentração de biomassa (X) aumenta.
- À medida que a concentração de biomassa (X) aumenta, a pressão de operação no flash (P_{flash}) aumenta.
- À medida que a concentração de glicerol (G) aumenta, a concentração de biomassa (X) aumenta.
- À medida que a concentração do etanol no fermentador (E_f) aumenta, a concentração de biomassa (X) e a turbidez (T_b) diminuem.
- À medida que a concentração do etanol condensado (E_c) aumenta, a concentração de biomassa aumenta (X).
- À medida que a vazão de permeado aumenta (F_p), a concentração de etanol no fermentador (E_f) diminui.
- À medida que a vazão de permeado aumenta, a concentração de etanol condensado (E_c) aumenta.

7.4.3. CARACTERIZAÇÃO DINÂMICA DOS DADOS

Os dados do processo de fermentação com extração contínua de produto por evaporação flash abarcam um período de operação desde perto das 40 h até as 400 h. Para a caracterização dinâmica das variáveis do processo, foram utilizadas janelas móveis de diferentes durações, onde foram computadas médias e variâncias.

As figuras a seguir, mostram estados estacionários, quase-estacionários e transientes para cada variável do processo. Tendo em conta o comportamento dinâmico não linear das variáveis, são considerados como intervalos de estado estacionário, os intervalos mais amplos que têm variância zero e média constante identificados utilizando as janelas de 15 h, 10 h e 5 h, e são denominados como estados quase-estacionários, os que têm variância próxima a zero nas janelas. As figuras que mostram a média móvel fornecem informação dos valores das médias de referência nas janelas e, as figuras que mostram a variância móvel dão

informação da constância da variância nas janelas. Em detalhe, se indicam para cada caso, os intervalos identificados que correspondem a estados estacionários e quase estacionários. Os períodos de tempo não compreendidos nestes intervalos correspondem a estados transientes.

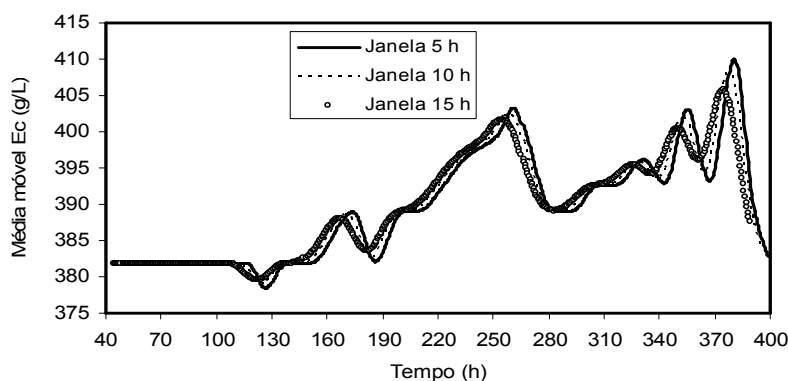


Figura 7.6. Média móvel da variável concentração de etanol condensado, Ec.

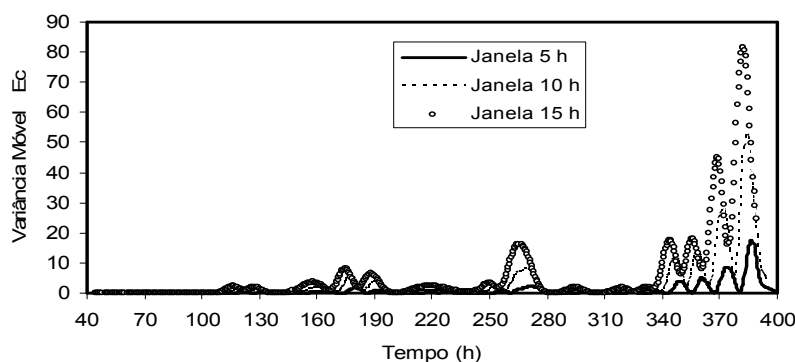


Figura 7.7. Variância móvel da variável concentração de etanol condensado, Ec.

As Figuras 7.6 e 7.7 mostram que a concentração de etanol condensado, tem um comportamento não linear com mudanças maiores nos períodos intermediários (aproximadamente de 144 h até 280 h) e final, desde aproximadamente 345 h de operação. A Figura 7.7 mostra que a partir de aproximadamente 340 h, a concentração de etanol condensado se torna mais instável e entra em um período de alta variabilidade.

Nas Figuras 7.6 e 7.7, a variável concentração de etanol condensado apresenta:

a) Intervalos de estado estacionário e médias de referência:

- 1) [45 h; 115,5 h] com média 381,87 g/L ;
- 2) [139,5 h; 150 h] com média 381,87 g/L;
- 3) [202,5 h; 208 h] com média 389,07 g/L;
- 4) [283,5 h; 292,5 h] com média 389,07g/L;
- 5) [307,5 h; 315,5 h] com média 392,67 g/L.

b) Intervalos de estado quase estacionário e médias de referência:

- 1) [116 h; 139 h] com média 380,34 g/L ;
- 2) [137 h; 156 h] com média 381,96 g/L;
- 3) [199 h; 215,5 h] com média 389,19 g/L;
- 4) [280 h; 297 h] com média 389,21 g/L;
- 5) [304 h; 319 h] com média 392,62 g/L.

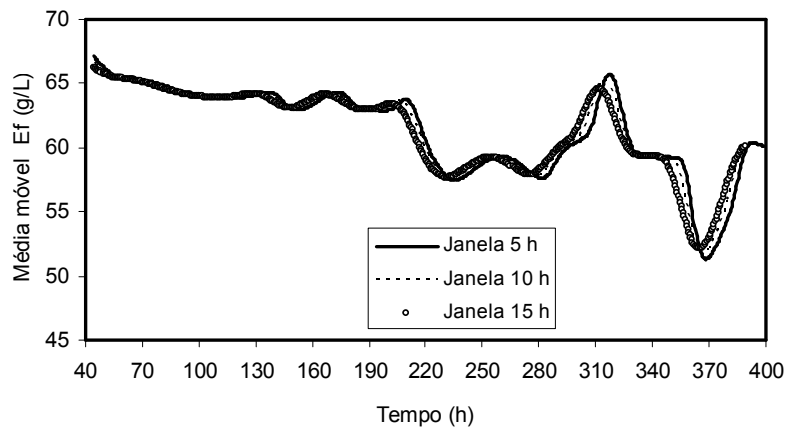


Figura 7.8. Média móvel da variável concentração de etanol no fermentador, Ef.

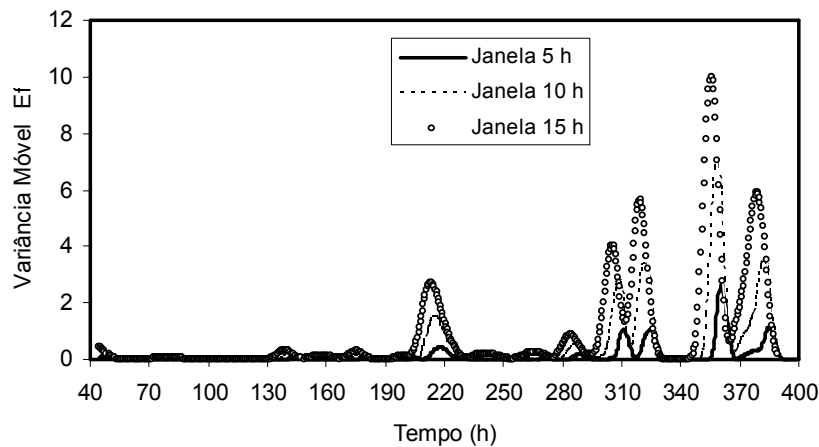


Figura 7.9. Variância móvel da variável concentração de etanol no fermentador, Ef.

As Figuras 7.8 e 7.9 mostram que durante o seu comportamento dinâmico não linear, a concentração de etanol no fermentador evolui de um estado para outro muito próximo rapidamente, pelo que possui estados estacionários de curta duração

(≤ 4 h), a variância móvel mostra que as mudanças maiores se apresentam a partir de aproximadamente 350 h de operação.

Nas Figuras 7.8 e 7.9, a variável concentração de etanol no fermentador apresenta:

a) Intervalos de estado estacionário e mediamédias de referência:

1) [108,5 h; 112,5 h] com média 63,87 g/L ; 2) [188 h; 191 h] com média 62,92 g/L.

Além deles existem outros estados estacionários com duração menor ou igual a 1,5 h.

b) Intervalos de estado quase estacionário e médias de referência:

1) [54 h, 80 h] com média 65.31 g/L;	5) [232 h, 242 h] com média 57,60 g/L;
2) [90 h, 140 h] com média 64.06 g/L;	6) [250 h, 264 h] com média 59,19 g/L;
3) [160 h, 170 h] com média 63.80 g/L;	7) [275 h, 286 h] com média 57,81 g/L;
4) [184 h, 202 h] com média 62.94 g/L;	8) [331 h; 354,5 h] com média 59,33 g/L.

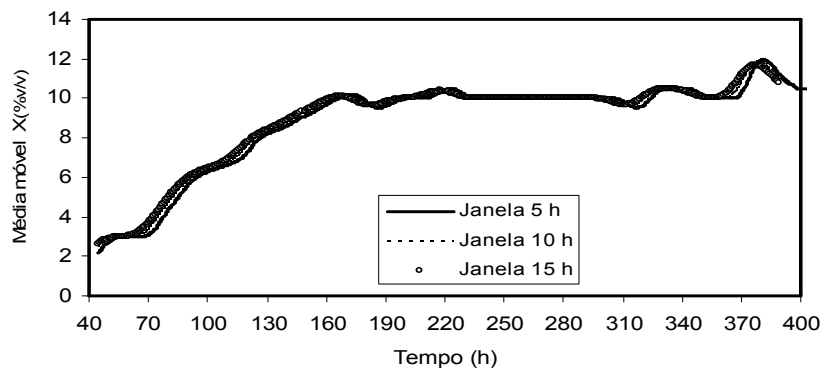


Figura. 7.10. Média móvel da variável concentração de biomassa no fermentador, X.

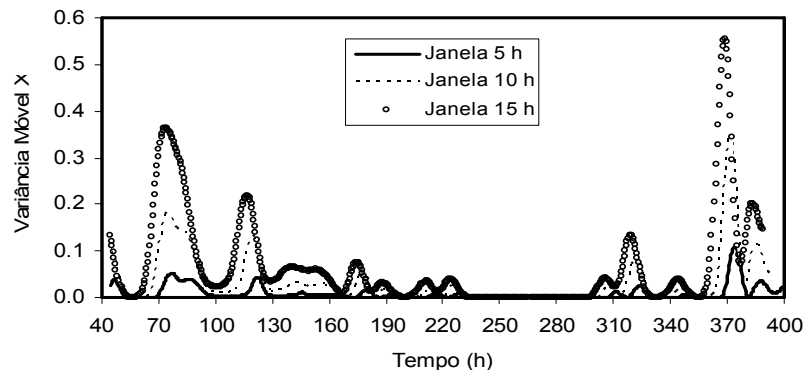


Figura 7.11. Variância móvel da variável concentração de biomassa no fermentador, X

De acordo com as média e variância móveis, a concentração de biomassa apresenta mudanças no período de tempo inicial até aproximadamente as 165 h, que são atribuíveis à formação de biomassa. As Figuras 7.10 e 7.11 mostram que a partir das 365 h de operação aproximadamente, a biomassa tem a mudança mais dinâmica.

Nas Figuras 7.10 e 7.11, a variável biomassa apresenta:

a) Intervalo de estado estacionário e média de referência:

1) [235 h, 305 h], com média 10,00 % v/v.

b) Intervalo de estado quase estacionário e média de referência:

1) [165 h, 365 h] com média 10,04 % v/v.

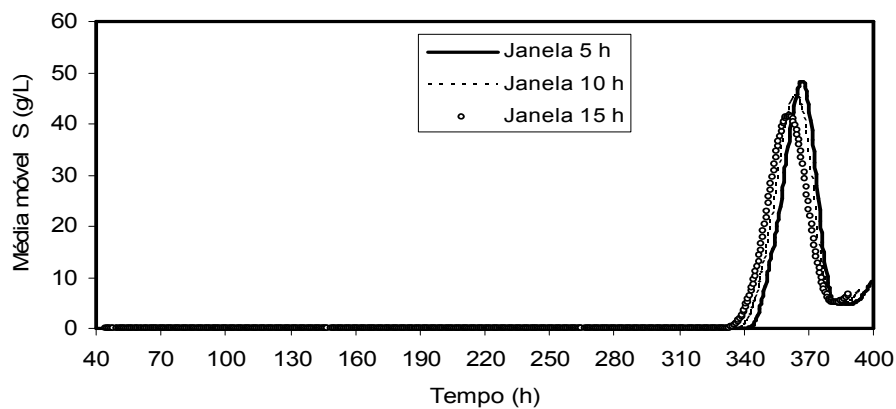


Figura 7.12. Média móvel da variável concentração de substrato no fermentador, S.

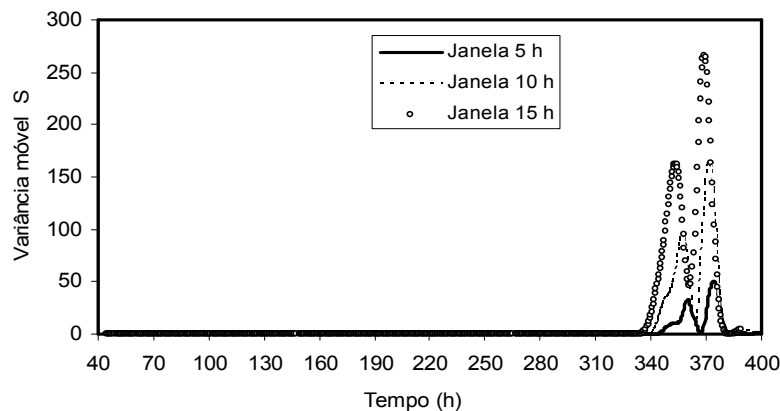


Figura 7.13. Variância móvel da variável concentração de substrato no fermentador, S.

Segundo o comportamento dinâmico apresentado nas figuras, a concentração do substrato no fermentador está em estado estacionário até aproximadamente as 345 horas de operação, após este período ocorrem mudanças mais dinâmicas na concentração de substrato no fermentador.

O substrato alimentado é consumido pela biomassa, de modo que, não há substrato residual (excedente) durante o intervalo de tempo que dura o estado estacionário para esta variável. De acordo com as média e variância móveis, o substrato não consumido se acumula rapidamente a partir de aproximadamente 345 h. A Figura 7.12 mostra que o substrato acumulado é consumido posteriormente, mas não de forma total no intervalo [380 h, 400 h].

Nas Figuras 7.12 e 7.13 para a variável concentração de substrato o intervalo de estado estacionário e média de referência é: [45 h, 345 h], com média 0,00 g/L.

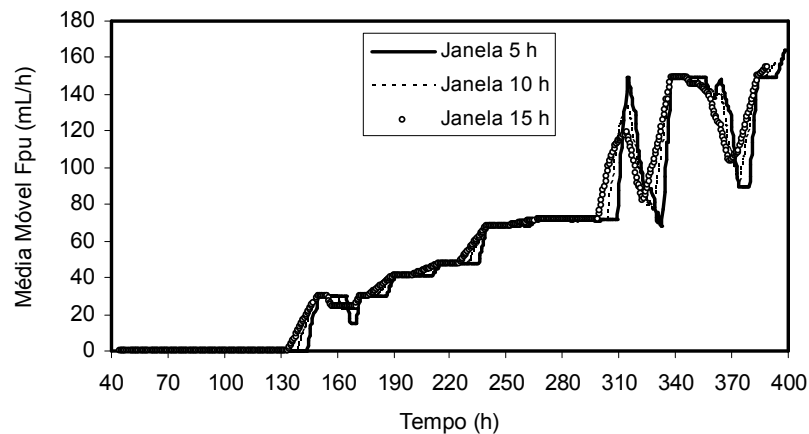


Figura 7.14. Média móvel da variável vazão de purga, Fpu.

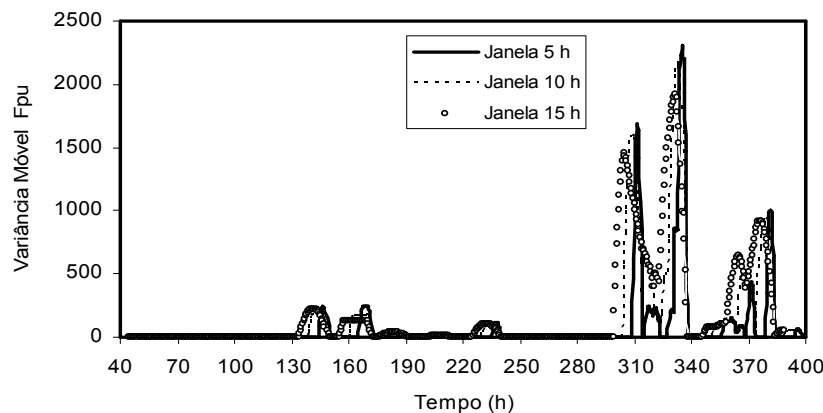


Figura 7.15. Variância móvel da variável vazão de purga, Fpu.

A Figura 7.14 mostra que a vazão de purga, aumenta durante o tempo de operação do processo, e a Figura 7.15 mostra que as maiores mudanças no seu comportamento dinâmico se produzem aproximadamente nos intervalos: 309 h a 338 h; e 356 h a 390 h.

Nas Figuras 7.14 e 7.15, a variável vazão de purga apresenta:

a) Intervalo de estado estacionário e média de referência:

[45 h, 144 h] com média 0,00 mL/h, nesse intervalo de tempo ainda não é utilizada a purga no processo.

b) Intervalos de estado quase estacionário e médias de referência:

1) [152 h, 165h] com média 29,43 mL/h;	5) [240 h, 262 h] com média 68,23 mL/h;
2) [173 h, 186h] com média 29,52 mL/h;	6) [268 h, 309 h] com média 71,63 mL/h;
3)[191h, 210.5 h] com média 41,39 mL/h;	7)[338 h, 356 h] com média 149,45 mL/h.
4) [218 h, 230 h] com média 47,91 mL/h;	

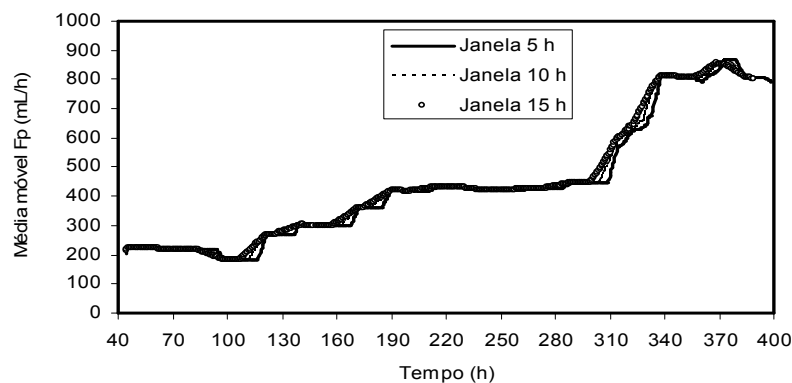


Figura 7.16. Média móvel da variável vazão de permeado, F_p .

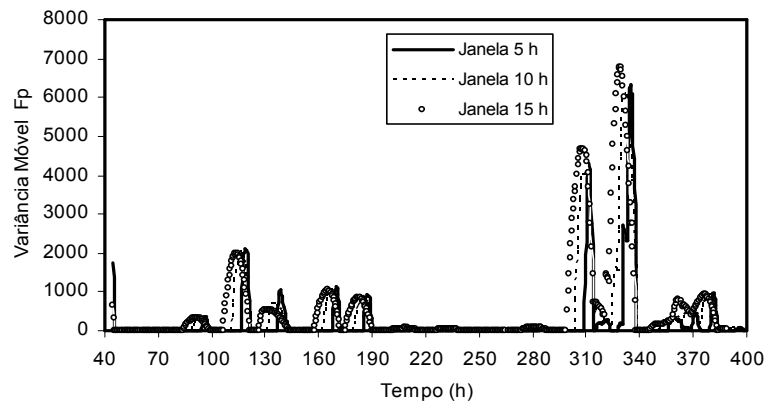


Figura 7.17. Variância móvel da variável vazão de permeado, F_p .

A Figura 7.16 mostra que a vazão de permeado aumenta durante o tempo de operação do processo, e a Figura 7.17 mostra que as maiores mudanças no seu comportamento dinâmico se produzem aproximadamente no intervalo: 309 h a 338 h.

Nas Figuras 7.16 e 7.17, a variável vazão de permeado no filtro, apresenta:

a) Intervalos de estado estacionário e médias de referência:

1) [45,5 h; 71 h] com média 221,16 mL/h;	4) [149 h;160,5] com média 299,88 mL/h;
2) [76 h, 94 h] com média 216,6 mL/h;	5) [267h; 283,5h] com média 426,60 mL/h;
3) [99 h, 116.5h] com média 180,48 mL/h;	6) [288,5 h; 309h] com média 445,56 mL/h

b) Intervalos de estado quase estacionário e médias de referência:

1) [122 h, 136 h] com média 268,22 mL/h ;	2) [161 h,167.5 h] com média 299,22 mL/h;
3) [174 h,185 h] com média 362,40 mL/h;	4) [190 h, 210 h] com média 418,86 mL/h;
5)[216 h,235 h] com média 433,14 mL/h;	6) [240 h,263 h] com média 421,31 mL/h;
7) [338 h, 356 h] com média 811,32 mL/h.	

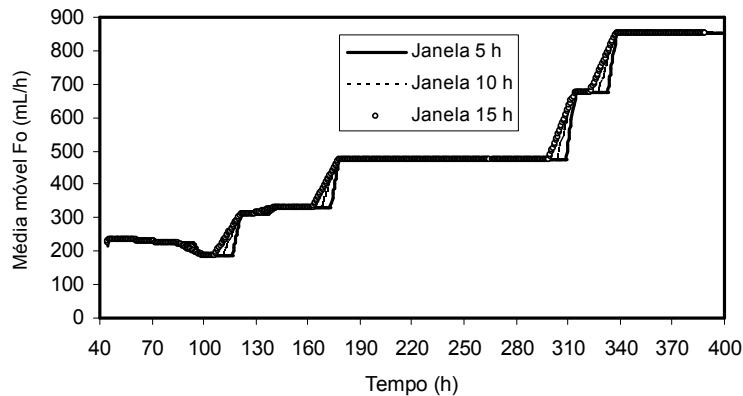


Figura 7.18. Média móvel da variável vazão de alimentação, F_o .

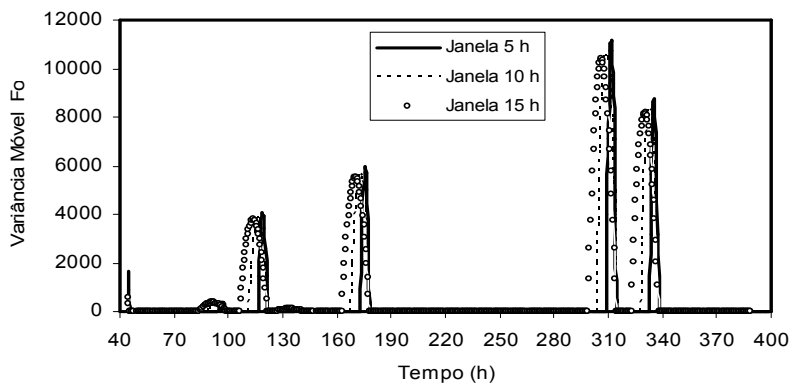


Figura 7.19. Variância móvel da variável vazão de alimentação, F_o .

As Figuras 7.18 e 7.19 mostram que, a vazão de alimentação aumenta durante o tempo de operação do processo, e que as maiores mudanças no seu comportamento dinâmico se produzem aproximadamente nos intervalos: 309 h - 314 h; 333 h - 338 h. Outras mudanças menores são observadas nos intervalos 94 h - 121 h e 173 h - 178 h, e assume-se que a variável tem um comportamento quase estacionário em outros intervalos de tempo diferentes aos anteriormente mencionados.

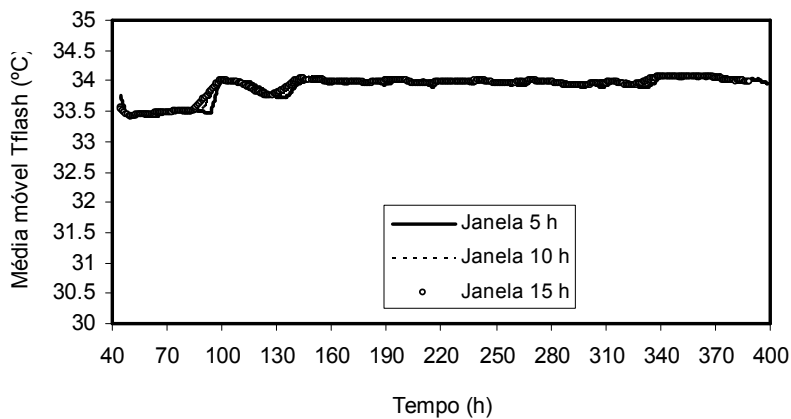


Figura 7.20. Média móvel da variável temperatura no tanque flash, T_{flash} .

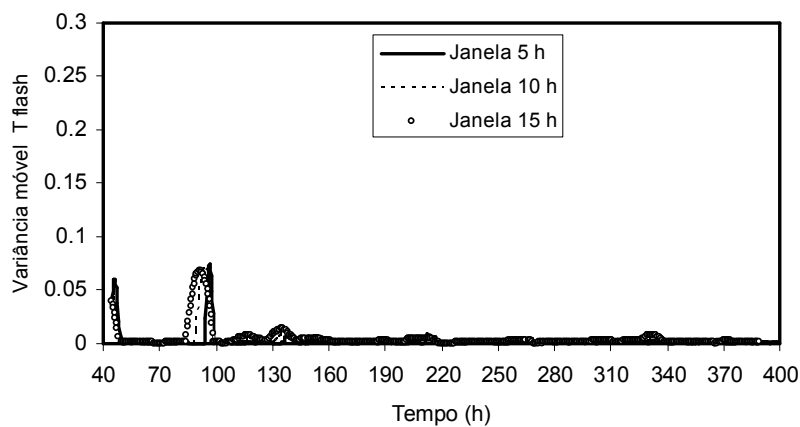


Figura 7.21. Variância móvel da variável temperatura no tanque flash, T_{flash} .

A Figura 7.21 mostra que a temperatura no tanque flash tem a maior mudança aproximadamente entre as 94 h e as 100 h de operação e que o seu comportamento é mais estável no intervalo de tempo de 140 h até 400 h.

A variável temperatura no tanque flash é uma variável controlada na planta de produção de etanol por fermentação extrativa contínua, de modo que, apresenta uma variância móvel muito próxima a zero a maior parte do tempo de operação.

Nas Figuras 7.20 e 7.21, a variável temperatura no tanque flash apresenta os seguintes intervalos de estado quase estacionário e médias de referência:

- 1) [48 h, 94 h] com média 33,47 °C;
- 2) [100 h, 116.5 h] com média 34 °C ;
- 3) [140 h, 400 h] com média 33,99 °C .

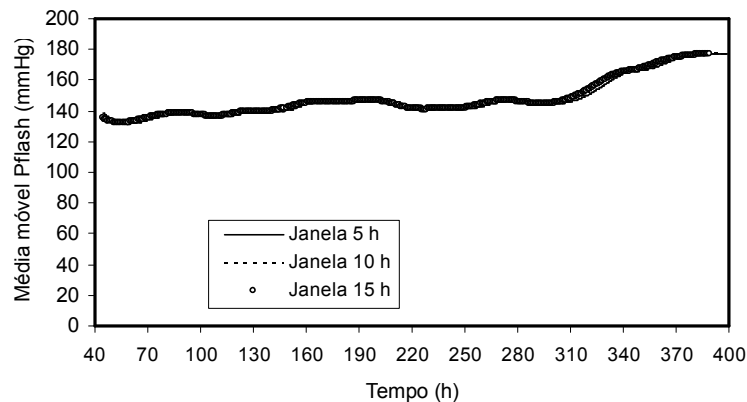


Figura 7.22. Média móvel da variável pressão de operação no tanque flash, Pflash.

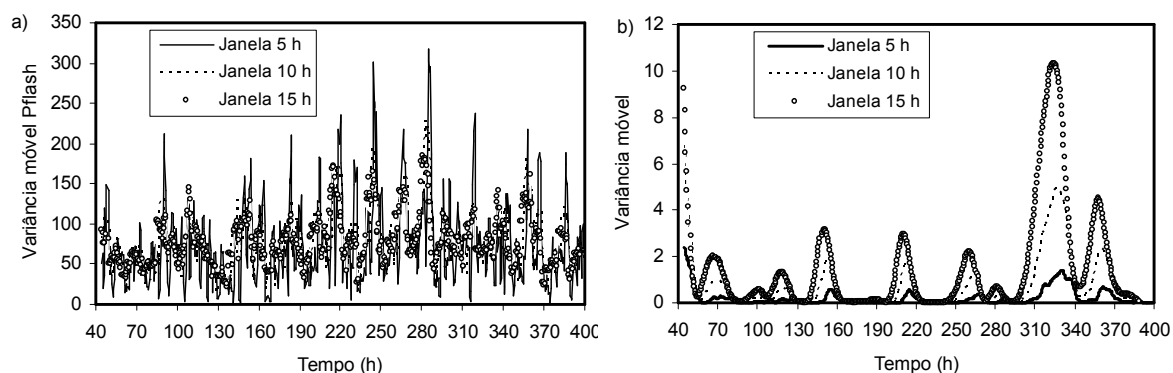


Figura 7.23. Variância móvel da variável pressão de operação no tanque flash, Pflash.

a) Com ruído do sinal. b) Após filtragem.

A Figura 7.23a mostra que o ruído do sinal não permite identificar as características do comportamento dinâmico da variável, na Figura 7.23b, depois de ter sido eliminado o ruído, se pode identificar claramente a dinâmica da variável durante a operação do processo.

A Figura 7.22 mostra que a pressão de operação no tanque flash aumenta muito lentamente durante o tempo de operação do processo até aproximadamente as 310 h. A Figura 7.23b mostra que a variável experimenta sua maior mudança no intervalo 310 h a 340 h de operação.

Nas Figuras 7.22 e 7.23, a variável pressão de operação no tanque flash apresenta os seguintes intervalos de estado quase estacionário e médias de referência:

1)[53 h, 65 h] com média 132,10 mmHg;	5)[220h,257h] com média 141,40 mmHg;
2)[77 h, 116 h] com média 137,60 mmHg;	6)[270h,310 h] com média 145,80 mmHg;
3)[128h,151h] com média 139,90 mmHg;	7)[341h,353h] com média 165,77 mmHg;
4)[162h,210 h] com média 146,12 mmHg;	8)[376h,400h] com média 176,68 mmHg.

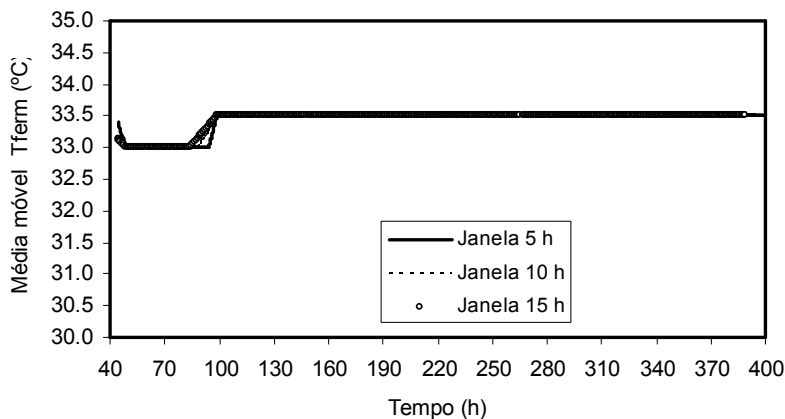


Figura 7.24. Média móvel da variável temperatura no fermentador, T_{ferm}.

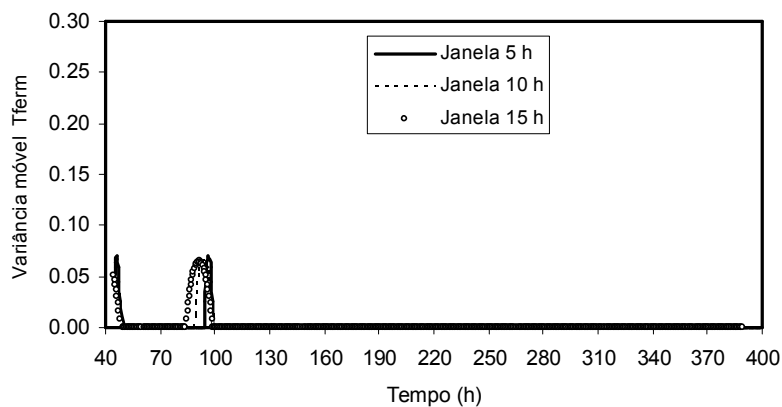


Figura 7.25. Variância móvel da variável temperatura no fermentador, T_{ferm} .

A Figura 7.25 mostra que a temperatura no fermentador tem a maior mudança aproximadamente entre as 94 h e as 99 h de operação e, que o seu comportamento é mais estável no intervalo de tempo de 99 h até 400 h.

A variável temperatura no fermentador é uma variável controlada na planta de produção de etanol por fermentação extrativa contínua com flash e apresenta uma variância móvel zero a maior parte do tempo de operação.

Nas Figuras 7.24 e 7.25, a variável temperatura no fermentador apresenta os seguintes intervalos de estado estacionário e médias de referência:

- 1) [49,5 h; 94 h] com média 33,0 °C;
- 2) [99 h, 400 h] com média 33,5 °C.

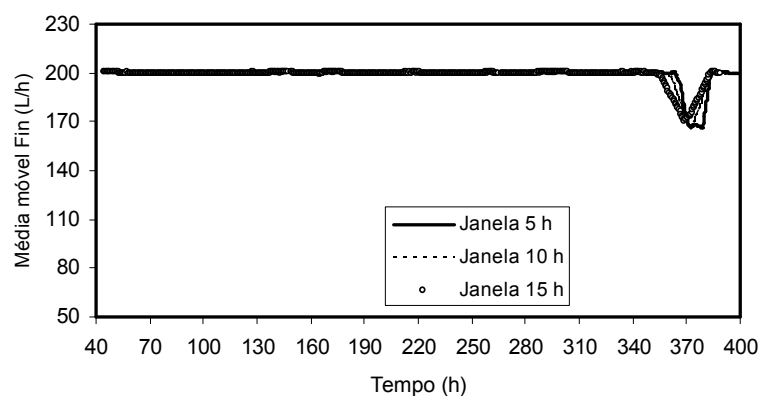


Figura 7.26. Média móvel da variável vazão de entrada no tanque flash, Fin .

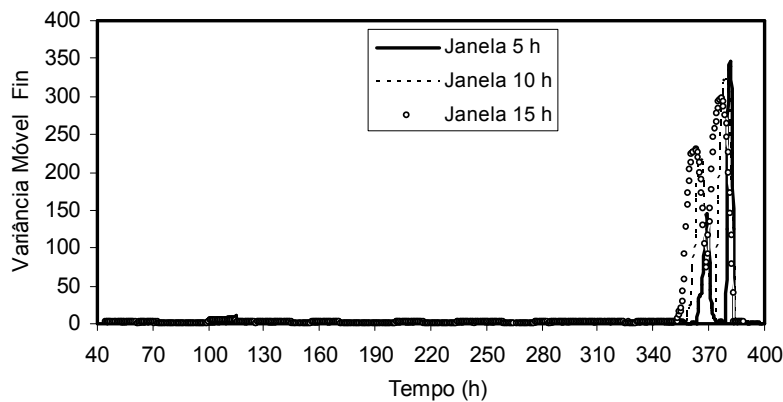


Figura 7.27. Variância móvel da variável vazão de entrada no tanque flash, Fin.

As Figuras 7.26 e 7.27 mostram que a variável vazão de entrada no tanque flash, tem a maior mudança no intervalo de tempo aproximado de 363 h até 385 h de operação e que o seu comportamento é estável em outros intervalos de tempo.

A variável vazão de entrada no flash é uma variável controlada na planta de produção de etanol por fermentação extrativa contínua com flash, assim apresenta uma variância móvel muito próxima de zero a maior parte do tempo de operação.

Nas figuras 7.26 e 7.27, a variável vazão de entrada no flash apresenta os seguintes intervalos de estado quase estacionário e médias de referência:

- 1) [45 h; 363,5 h] com média 200,01 L/h e
- 2) [385 h, 400 h] com média 200,15 L/h

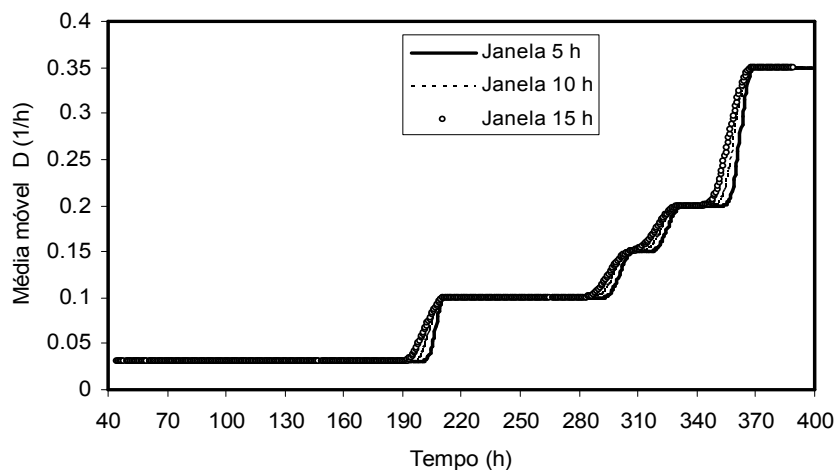


Figura 7.28. Média móvel da variável taxa de diluição no fermentador, D

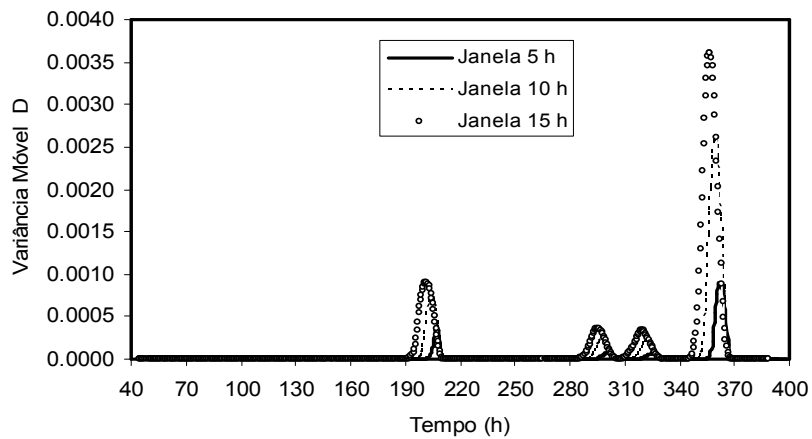


Figura 7.29. Variância móvel da variável taxa de diluição no fermentador, D.

As Figuras 7.28 e 7.29 mostram que a taxa de diluição no fermentador apresenta quatro mudanças no seu valor durante o tempo de operação. Segundo a Figura 7.29, a maior mudança ocorre aproximadamente as 350 h de operação.

Nas Figuras 7.28 e 7.29, a variável taxa de diluição apresenta os seguintes intervalos de estado estacionário e médias de referência:

- 1) [45 ; 202,5 h] com média $0,03 \text{ h}^{-1}$; 2) [210 h; 296 h] com média $0,10 \text{ h}^{-1}$;
- 3) [305 h; 320 h] com média $0,15 \text{ h}^{-1}$; 4) [329,5 h; 356 h] com média $0,20 \text{ h}^{-1}$;
- 5) [367,5 h; 400 h] com média $0,35 \text{ h}^{-1}$.

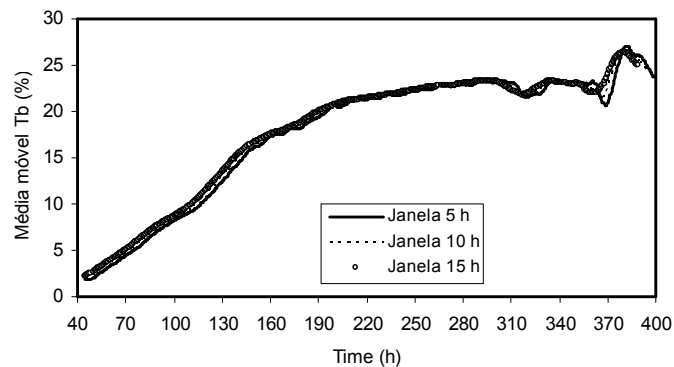


Figura 7.30. Média móvel da variável concentração de etanol condensado, Ec.

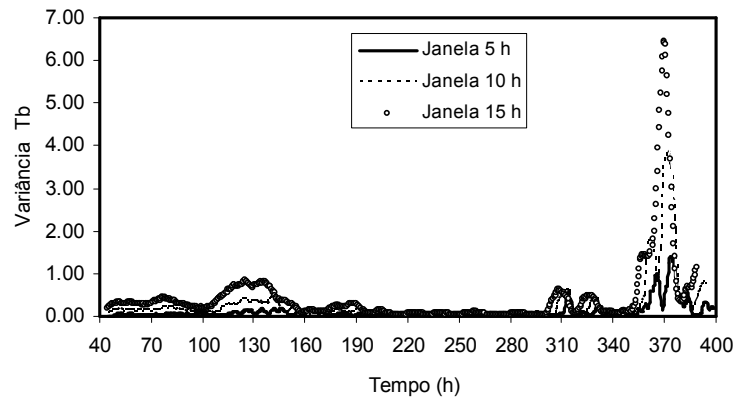


Figura 7.31. Variância móvel da variável turbidez no fermentador, Tb.

A Figura 7.30 mostra que a turbidez, Tb, aumenta continuamente até alcançar o valor próximo de 23%, no entanto, com ajuda da Figura 7.31 podem-se identificar vários intervalos com estado quase estacionário na sua trajetória dinâmica. Essa variável experimenta mudanças mais dinâmicas aproximadamente às 360 h de operação.

Segundo análise das correlações, a turbidez está muito associada com a concentração de biomassa no fermentador, o que mostra que esta medida on-line pode ser usada para estimar a concentração de biomassa, somente obtida off-line.

Nas Figuras 7.30 e 7.31, a variável turbidez apresenta os seguintes intervalos de estado quase estacionário e médias de referência:

- 1) [159 h, 168 h] com média 17,42% ;
- 2) [171 h, 179.5 h] com média 18,13% ;
- 3) [198 h, 207 h] com média 20,58% ;
- 4) [211 h, 227 h] com média 21,31% ;
- 5) [233 h, 244 h] com média 21,79% ;
- 6) [246.5 h, 261 h] com média 22,30% ;
- 7) [265 h, 286 h] com média 22,84% ;
- 8) [290 h, 310 h] com média 23,27% ;
- 9) [316 h, 329 h] com média 21,79 %;
- 10) [333.5 h, 343 h] com média 23,33 %.
- 11) [345 h, 363 h] com média 23,02 %.

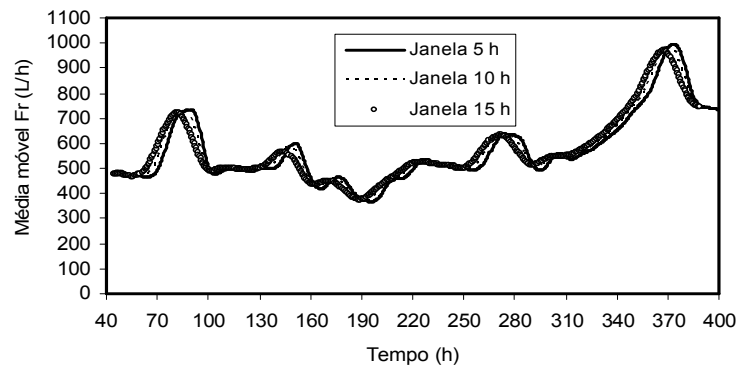


Figura 7.32. Média móvel da variável vazão de retorno ao fermentador, Fr.

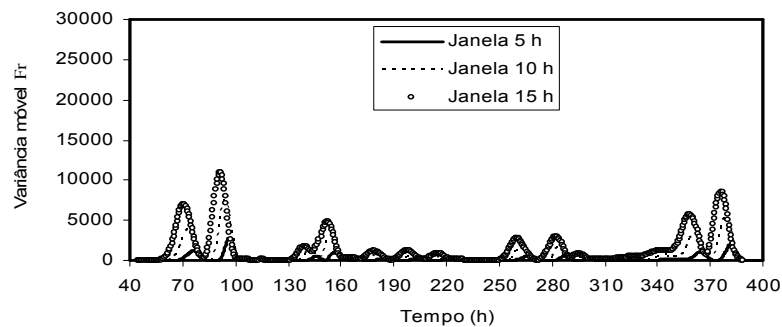


Figura 7.33. Variância móvel da variável vazão de retorno ao fermentador, Fr.

Para observar melhor a tendência das mudanças da variável foi feita uma filtragem dos dados iniciais utilizando um filtro digital anti-causal, a fim de eliminar o ruído do sinal. As Figuras 7.32 e 7.33 mostram que a vazão de retorno ao fermentador, Fr, em geral aumenta durante o tempo de operação, experimentando as maiores mudanças nos períodos inicial entre as 70 e 100 h e final aproximadamente as 350 h de operação. A vazão de saída do fermentador, F, tem o mesmo comportamento dinâmico ($Fr = 0,7 * F$), a Tabela 7.1 mostra que as variáveis F e Fr têm praticamente os mesmos valores nos coeficientes de correlação.

Nas Figuras 7.32 e 7.33, a variável vazão de retorno ao fermentador apresenta os seguintes intervalos de estado quase estacionário e médias de referência:

- 1) [112 h, 119 h] com média 506,74 L/h; 2) [129 h, 139 h] com média 501,10 L/h;
- 3) [233 h, 246.5 h] com média 515,11 L/h; 4) [250 h, 259 h] com média 497,96 L/h;
- 5) [271 h, 284 h] com média 633,43 L/h; 6) [310.5 h, 316 h] com média 540 L/h.

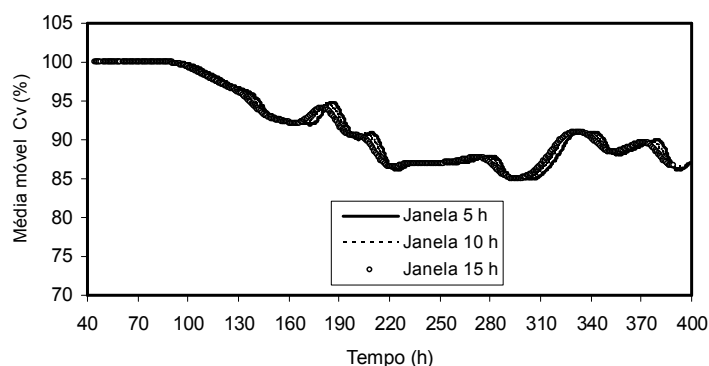


Figura 7.34. Média móvel da variável viabilidade celular, Cv.

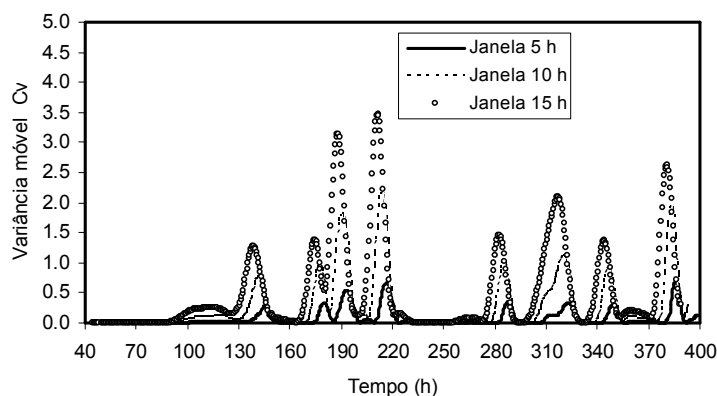


Figura 7.35. Variância móvel da variável viabilidade celular, Cv.

As Figuras 7.34 e 7.35 mostram que a viabilidade celular diminui durante as primeiras 200 h de operação do processo, tende a manter os seus valores entre as 200 h e 280 h, e logo tem uma dinâmica mais instável até as 400 h de operação.

Nas Figuras 7.34 e 7.35, a variável viabilidade celular apresenta:

a) Intervalos de estado estacionário e média de referência:

1) [45 h, 92 h] com média 100 %; 2) [235,5 h; 260,5 h] com média 87%;

b) Intervalos de estado quase estacionário e médias de referência:

1) [152 h, 179 h] com média 92,37 %; 2) [180,5 h;190,5 h] com média 94,42 %

3) [220 h, 235 h] com média 86,51 %; 4) [262 h; 286 h] com média 87,55 %

5) [289 h;313,5 h] com média 85,20 %; 6) [330 h; 347 h] com média 90,91 %.

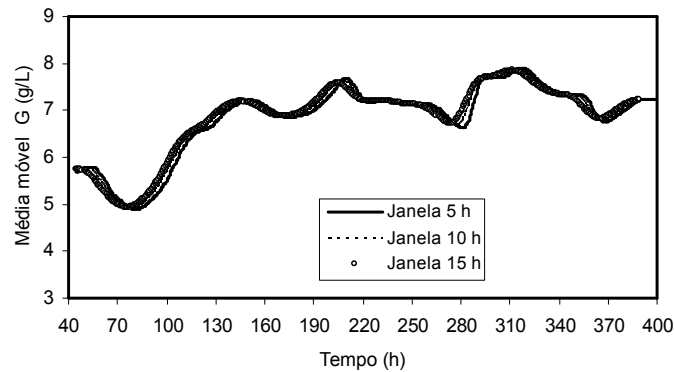


Figura 7.36. Média móvel da variável concentração de glicerol no fermentador.

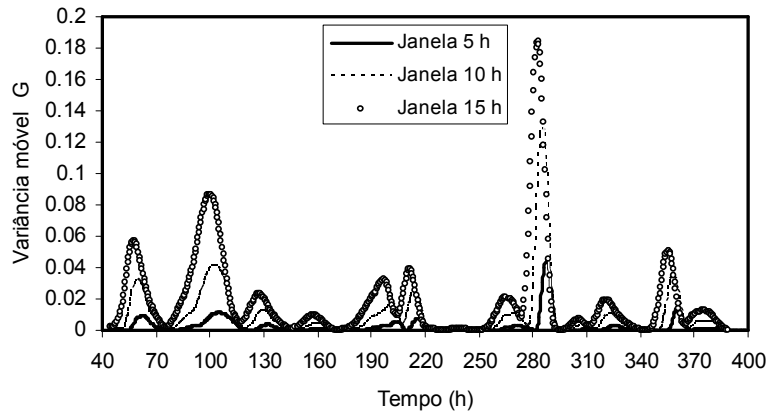


Figura 7.37. Variância móvel da variável concentração de glicerol no fermentador.

As Figuras 7.36 e 7.37 mostram que a concentração do glicerol, G tem uma dinâmica não linear com aumentos e diminuições na sua concentração, no entanto, em geral aumenta durante o tempo de operação. O glicerol é um dos principais subprodutos na produção de etanol por fermentação dos melaços de cana de açúcar, e tende a acumular-se, ao contrário do etanol, que é extraído utilizando evaporação flash.

Nas Figuras 7.36 e 7.37, a variável concentração do glicerol, G , apresenta os seguintes intervalos de estado quase estacionário e médias de referência:

- 1) [46 h, 58 h] com média 5,75 g/L; 2) [77,5 h; 86,5 h] com média 4,92 g/L;
- 3) [146 h, 156 h] com média 7,20 g/L; 3) [168 h, 187 h] com média 6,89 g/L;
- 4) [277 h, 284 h] com média 6,66 g/L; 5) [293 h, 307 h] com média 7,69 g/L;
- 6) [338 h, 352 h] com média 7,34 g/L.

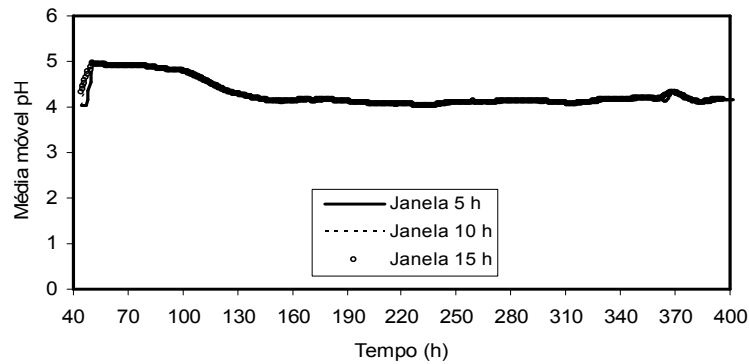


Figura 7.38. Média móvel da variável pH no fermentador .

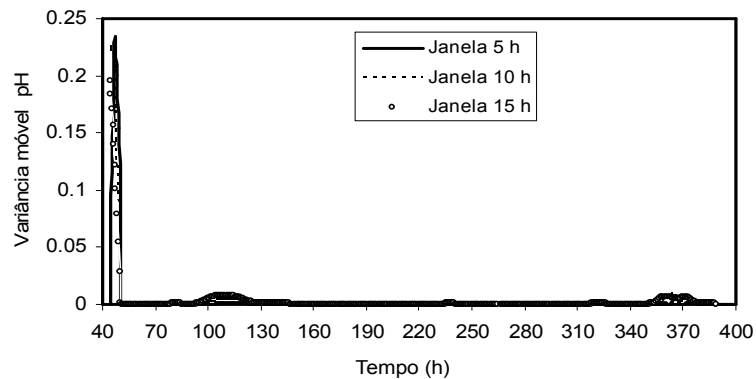


Figura 7.39. Variância móvel da variável pH no fermentador.

A Figura 7.38 mostra que o pH diminui lentamente de 5 para 4 aproximadamente entre as 75 h e 150 h de operação. No intervalo de 150 h até as 400 h de operação o pH evolui de um estado quase estacionário para outro muito próximo, sendo a média geral neste período de tempo de 4,12.

Nas Figuras 7.38 e 7.39, o pH em geral evolui de um estado quase-estacionário para outro, exceto no intervalo [61,5 h; 74,5 h], onde apresenta estado estacionário com média de referência 4,91.

7.4.4. COMENTÁRIOS FINAIS DO COMPORTAMENTO DINÂMICO DAS VARIÁVEIS DO PROCESSO.

Os resultados mostram que as mudanças mais importantes na dinâmica do processo se produzem no período final de operação, e começam no intervalo

aproximado de 310 h a 340h de operação, com mudanças maiores nas vazões: de alimentação F_o , de purga F_{pu} e de permeado F_p , que têm como consequência mudanças na taxa de diluição D . Observa-se, no entanto, que algumas variáveis são mais sensíveis do que outras a essas mudanças. Por exemplo, as concentrações de etanol no fermentador E_f e de etanol condensado E_c , a pressão de operação no tanque flash P_{flash} , a vazões de retorno ao fermentador F_r e de entrada ao tanque flash F_{in} , as concentrações no fermentador de substrato S e biomassa X e a turbidez T_b , todas elas mostram também maiores mudanças no intervalo de 310 h a 400 h de operação. Entretanto, outras variáveis, tais como as temperaturas de operação no fermentador e no flash, permanecem quase constantes perto dos seus valores de referência, assim como o pH. A viabilidade celular C_v mostra mudanças que estariam associadas também a outras fontes de variação e a concentração de glicerol G mostra outras características de comportamento.

Identificar o cumprimento dos intervalos de estado estacionário ou quase estacionário globais para o processo é complexo devido à variedade de intervalos no comportamento não linear das variáveis. Neste trabalho, se assume que o estado quase estacionário simultâneo das concentrações de etanol, biomassa e substrato corresponde a estado quase estacionário do processo. Infere-se que a maior parte do tempo de operação o processo esteve em estado transiente.

7.4.5. MODELAGEM EMPÍRICA BASEADA EM REDES NEURAIIS PARA MONITORAMENTO DA CONCENTRAÇÃO DE ETANOL

7.4.5.1. MODELO 1: modelo empírico com variáveis cuja determinação é off-line (Figura 7.2).

Tabela 7.2. MODELO 1: Número de neurônios na camada oculta, erro quadrático médio (mse), R^2 e CORR no grupo de teste com 100 épocas de treinamento.

Neurônios	MSE (Ef)	R^2 (Ef)	CORR(Ef)	MSE(Ec)	R^2 (Ec)	CORR(Ec)
5	0,055	0,9953	99,533	0,649	0,9884	98,832
6	0,102	0,9914	99,140	0,774	0,9861	98,607
7	0,042	0,9965	99,645	0,233	0,9958	99,580
8	0,030	0,9975	99,747	0,187	0,9967	99,663
9	0,022	0,9981	99,814	0,280	0,9950	99,496
10	0,009	0,9993	99,928	0,110	0,9980	99,803
11	0,007	0,9994	99,942	0,051	0,9991	99,908
12	0,005	0,9996	99,961	0,053	0,9991	99,904
13	0,005	0,9996	99,954	0,051	0,9991	99,908
14	0,003	0,9998	99,976	0,041	0,9993	99,926
15	0,004	0,9997	99,967	0,029	0,9995	99,948

Tabela 7.3. MODELO 1: Número de neurônios na camada oculta, erro quadrático médio (mse), R^2 e CORR no grupo de teste com 300 épocas de treinamento.

Neurônios	MSE (Ef)	R^2 (Ef)	CORR(Ef)	MSE(Ec)	R^2 (Ec)	CORR(Ec)
5	0,082	0,9931	99,305	0,866	0,9846	98,441
6	0,061	0,9948	99,483	0,528	0,9905	99,050
7	0,044	0,9963	99,631	0,262	0,9954	99,529
8	0,036	0,9970	99,698	0,277	0,9950	99,501
9	0,014	0,9988	99,882	0,314	0,9945	99,434
10	0,009	0,9992	99,923	0,094	0,9983	99,830
11	0,007	0,9994	99,943	0,034	0,9994	99,939
12	0,004	0,9997	99,966	0,026	0,9995	99,954
13	0,003	0,9998	99,975	0,026	0,9995	99,953
14	0,002	0,9998	99,982	0,017	0,9997	99,969
15	0,001	0,9999	99,990	0,014	0,9998	99,975

As Tabelas 7.2 e 7.3 mostram que uma melhor precisão do modelo de redes neurais artificiais na previsão dos valores experimentais é obtida com maior número de neurônios na camada escondida (no intervalo 5 a 15 neurônios) e tempo de treinamento no intervalo de 100 a 300 épocas (foram testadas diferentes épocas e os resultados apresentados foram os melhores obtidos). As Tabelas também mostram, que em geral, quando se aumenta o número de neurônios de um em um na camada escondida, as melhoras nos valores dos indicadores para modelos com 11 ou mais neurônios nessa camada já são menores do que nos casos com menos de 11 neurônios, e que nos casos de 11 ou mais neurônios os valores dos indicadores são melhores com 300 do que com 100 épocas de treinamento, portanto segundo os testes realizados considera-se que uma rede

neural com 11 neurônios na camada escondida com 300 épocas de treinamento constitui um modelo representativo para a previsão das concentrações de etanol utilizando variáveis que requerem determinação *off-line*.

7.4.5.2. MODELO 2: modelo empírico com variáveis que não requerem determinação *off-line* (Figura 7.3).

Tabela 7.4. MODELO 2: Número de neurônios na camada oculta, erro quadrático médio (mse), R² e CORR no grupo de teste com 300 épocas de treinamento.

Neurônios	MSE (Ef)	R ² (Ef)	CORR(Ef)	MSE(Ec)	R ² (Ec)	CORR(Ec)
5	0,322	0,973	97,276	2,094	0,962	96,231
6	0,256	0,978	97,831	1,431	0,974	97,423
7	0,152	0,987	98,716	1,347	0,976	97,576
8	0,103	0,991	99,132	1,025	0,982	98,156
9	0,085	0,993	99,285	0,547	0,990	99,016
10	0,045	0,996	99,622	0,521	0,991	99,062
11	0,050	0,996	99,581	0,321	0,994	99,422
12	0,028	0,998	99,765	0,388	0,993	99,302
13	0,024	0,998	99,794	0,311	0,994	99,439
14	0,014	0,999	99,878	0,180	0,997	99,676
15	0,019	0,998	99,838	0,222	0,996	99,600
16	0,012	0,999	99,897	0,172	0,997	99,690
17	0,018	0,998	99,848	0,134	0,998	99,760
18	0,010	0,999	99,917	0,159	0,997	99,714

Tabela 7.5. MODELO 2: Número de neurônios na camada oculta, erro quadrático médio (mse), R² e CORR no grupo de teste com 500 épocas de treinamento.

Neurônios	MSE (Ef)	R ² (Ef)	CORR(Ef)	MSE(Ec)	R ² (Ec)	CORR(Ec)
5	0,276	0,977	97,669	2,537	0,954	95,434
6	0,235	0,980	98,011	1,534	0,972	97,239
7	0,193	0,984	98,366	0,858	0,985	98,457
8	0,130	0,989	98,902	0,892	0,984	98,394
9	0,075	0,994	99,366	0,534	0,990	99,040
10	0,064	0,995	99,461	0,398	0,993	99,283
11	0,048	0,996	99,594	0,515	0,991	99,073
12	0,038	0,997	99,678	0,333	0,994	99,400
13	0,034	0,997	99,709	0,255	0,995	99,542
14	0,020	0,998	99,834	0,166	0,997	99,702
15	0,010	0,999	99,915	0,143	0,997	99,742
16	0,018	0,999	99,851	0,268	0,995	99,518
17	0,014	0,999	99,882	0,174	0,997	99,687
18	0,012	0,998	99,896	0,156	0,997	99,719

Após obter os resultados considerando diferentes números de épocas, as Tabelas 7.4 e 7.5 mostram que, neste caso, se requer um maior número de neurônios e um maior número de épocas de treinamento do que o caso anterior (modelo 1) para atingir resultados similares. A rede neural com 15 neurônios na camada escondida e 300 épocas de treinamento é a rede com menor número de neurônios que apresenta os melhores valores dos indicadores utilizados, portanto se considera que é um modelo adequado para a previsão das concentrações de etanol utilizando variáveis que não requerem determinação *off-line*.

A representação gráfica da topologia dos modelos baseados em redes neurais se mostra na Figura 7.40.

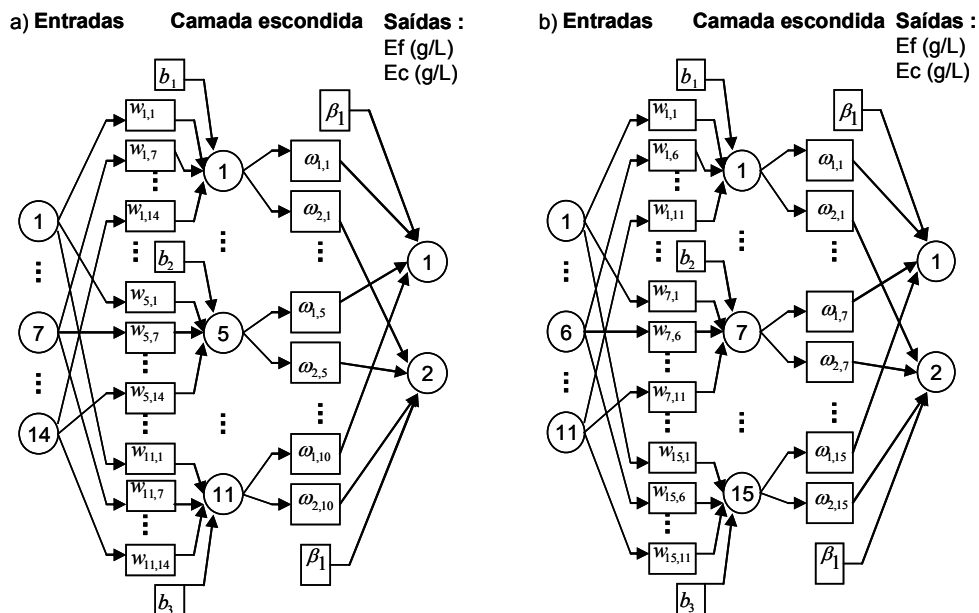


Figura 7.40. Representação gráfica das estruturas de rede neural:
a) Modelo 1 com variáveis que precisam de determinação *off-line*.
b) Modelo 2 com variáveis que não precisam de determinação *off-line*.

7.4.5.3. GRÁFICOS DE PARIDADE COM DADOS EXPERIMENTAIS DE TESTE E ESTIMADOS PELOS MODELOS EMPÍRICOS

As Figuras 7.41, 7.42, 7.43 e 7.44 mostram os resultados da comparação dos valores alvo experimentais das concentrações de etanol no processo, com os valores estimados pelos modelos: 1 (Figuras 7.41 e 7.42) e 2 (Figuras 7.43 e

7.44). Os resultados foram obtidos utilizando integrantes do grupo de teste, esses integrantes não foram utilizados para o treinamento da rede.

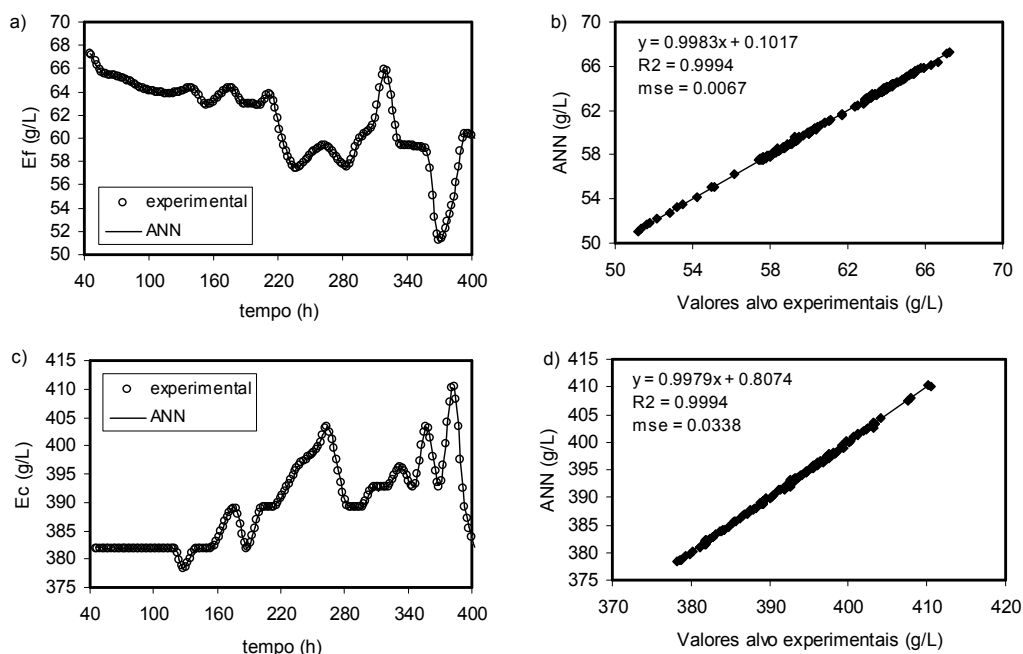


Figura 7.41. MODELO 1: a) e b) Valores da concentração de etanol no fermentador experimentais e estimados pela rede neural. c) e d) Valores da concentração de etanol condensado experimentais e estimados pela rede neural.

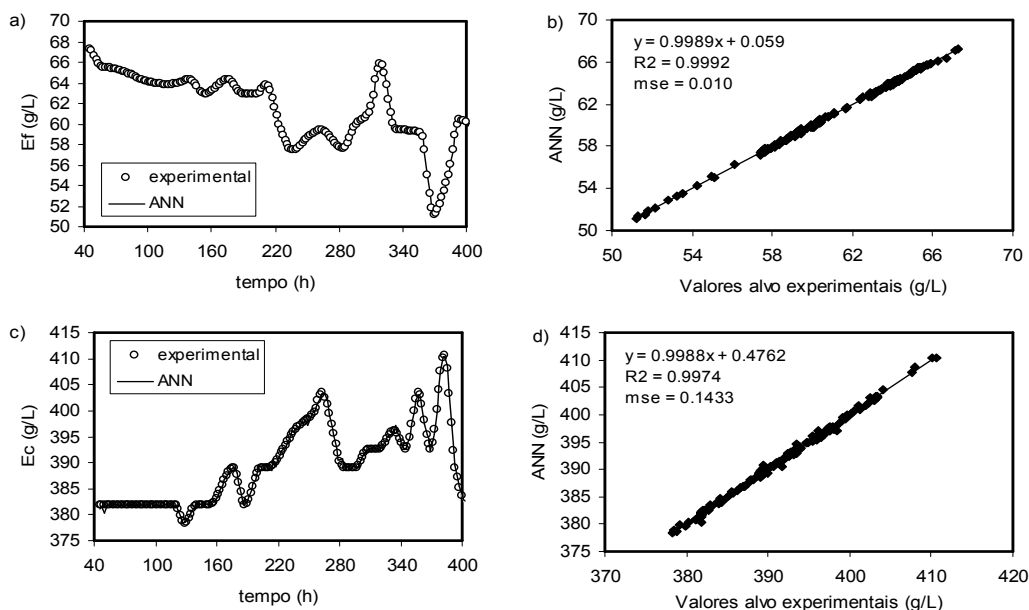


Figura 7.42. MODELO 2: a) e b) Valores da concentração de etanol no fermentador experimentais e estimados pela rede neural. c) e d) Valores da concentração de etanol condensado experimentais e estimados pela rede neural.

As Figuras 7.41, 7.42 mostram que as redes neurais selecionadas têm um bom desempenho ao estimar com alta precisão os valores alvo contidos nos conjuntos de teste dos modelos desenvolvidos. Portanto, o uso de redes neurais torna possível o desenvolvimento de modelos de processo utilizando também variáveis que não requerem determinação *off-line*. Segundo estes resultados, é possível o desenvolvimento de modelos utilizáveis como *software sensors* para monitorar as concentrações de etanol no fermentador E_f e etanol condensado E_c no sistema de separação flash. Rivera et. al. (2010) pesquisaram o desenvolvimento de software sensors para as concentrações de E_f e E_c no processo de fermentação extrativa com flash e relataram em uma etapa preliminar o uso de 7 variáveis de entrada cuja determinação é *on-line*: T_{ferm} , pH, T_b (turbidez), T_{flash} , P_{flash} , F_{in} (Vazão de alimentação para o flash) e D (coeficiente de diluição no fermentador), com coeficientes de determinação $R^2 = 0,90$ para E_f , e $R^2 = 0,93$, para E_c . Em uma etapa posterior, sugeriram o uso de 4 variáveis: T_{ferm} , D , T_{flash} , P_{flash} , com coeficientes de determinação $R^2 = 0,87$ para E_f , e $R^2 = 0,90$, para E_c (caso 12 na Tabela 7.6 página seguinte).

Neste trabalho, foi eliminado o ruído do sinal das variáveis pressão de operação no flash (P_{flash}) e vazão de retorno do filtro (F_r), e foram utilizadas preliminarmente 11 variáveis com a finalidade de melhorar as estimações das concentrações de etanol no fermentador E_f e etanol condensado E_c pelo *software sensor*, como resultado se teve um aumento no coeficiente de determinação até atingir os valores de $R^2 = 0,999$ para a concentração de etanol no fermentador E_f e $R^2 = 0,997$ para a concentração de etanol condensado E_c .

7.4.5.4. ANÁLISE EXPLORATÓRIA DE MODELOS EMPÍRICOS COM DIFERENTES ENTRADAS QUE NÃO REQUEREM DETERMINAÇÃO OFF-LINE

Uma análise exploratória indica que o número de variáveis que não requerem determinação *off-line* no modelo empírico 2 pode ser diminuído com similar eficiência nas estimações. As variáveis de entrada de referência básicas que não requerem determinação *off-line* são: T_{ferm} , F_{pu} , F_p , T_{flash} , P_{flash} , F_r (retorno do filtro), D (coeficiente de diluição no fermentador), e T_b (turbidez).

Na Tabela 7.6 se mostram os valores dos critérios de seleção (mse, R^2 , CORR) para modelos de redes neurais com diferentes entradas que não requerem determinação *off-line* e que têm como saídas as concentrações de etanol Ef e Ec. O número de neurônios da camada oculta que se indica na tabela, é o ótimo em cada caso. Os casos 1 com 11 entradas e 2 com 10 entradas possuem os melhores valores dos critérios de seleção. Portanto, são as melhores alternativas de modelo para estudar seu uso como *software sensor*. No entanto, o caso 6 com 8 entradas é a estrutura mais simples, com resultados ligeiramente menos precisos.

Tabela 7.6. Modelos de redes neurais artificiais (ANN) com diferentes entradas determináveis *on-line* e saídas etanol no fermentador Ef e etanol condensado Ec.

Var modelo	1	2	3	4	5	6	7	8	9	10	11	Nn	Etanol no fermentador, Ef (g/L)			Etanol condensado, Ec (g/L)		
													Mse	R ²	CORR	Mse	R ²	CORR
1	√	√	√	√	√	√	√	√	√	√	√	18	0,010	0,999	99,917	0,159	0,997	99,714
2	√	√	√	-	√	√	√	√	√	√	√	15	0,026	0,998	99,781	0,132	0,998	99,762
3	√	√	√	-	√	√	√	√	√	-	√	16	0,047	0,996	99,606	0,205	0,996	99,630
4	√	√	√	-	√	√	√	-	√	-	√	18	0,030	0,997	99,743	0,225	0,996	99,589
5	-	√	√	-	√	√	√	-	√	-	√	19	0,031	0,997	99,735	0,278	0,995	99,500
6	-	√	√	.	√	√	√	√	√	-	√	18	0,037	0,998	99,684	0,240	0,998	99,567
7	√	-	-	-	√	√	√	√	√	√	√	14	0,047	0,996	99,601	0,533	0,990	99,040
8	√	-	-	-	√	√	√	√	√	-	√	18	0,055	0,995	99,533	0,457	0,992	99,177
9	√	√	√	-	√	√	-	-	√	-	√	17	0,059	0,995	99,501	0,712	0,987	98,712
10	-	√	√	-	√	√	-	-	√	-	√	15	0,062	0,995	99,476	0,700	0,987	98,741
11	√	-	-	-	√	√	-	-	√	-	√	19	0,087	0,993	99,263	1,664	0,970	97,006
12	√	-	-	-	√	√	-	-	√	-	-	13	0,593	0,950	94,982	4,164	0,925	92,506

Nn = número de neurônios na camada escondida.

(√) : Indica presença da variável de entrada no modelo.

Leyenda para variáveis de entrada (Var) nos modelos de ANN da Tabela 7.6

Número	1	2	3	4	5	6	7	8	9	10	11
Var	T _{ferm} °C	F _{pu} mL/h	F _p mL/h	F _o mL/h	P _{flash} mmHg	T _{flash} °C	F _r L/h	F _{in} L/h	D h ⁻¹	pH	T _b %

7.5. CONCLUSÕES

A eliminação de ruído do sinal melhorou notavelmente a qualidade da informação transmitida do comportamento dinâmico das variáveis, aumentando a

precisão das estimações feitas pelos modelos empíricos do processo, e pondo em evidência que é uma estratégia que permite melhorar o uso destes modelos como *software sensors* e como ferramentas para análise por simulação dos processos.

No entanto, para o uso dos modelos empíricos como *software sensors*, se salienta a necessidade de utilizar técnicas *on-line* para eliminação de ruído dos sinais. Prata (2009) mostrou resultados satisfatórios do desenvolvimento destas técnicas, relatando que a pesquisa nesta área está orientada ao desenvolvimento de técnicas robustas que permitem a reconciliação de dados para monitoramento *on-line* dos processos industriais. Morgado (2005) menciona que podem ser usados filtros *anti-aliasing* analógicos e filtros digitais para a filtragem, após amostragem.

As matrizes de correlações constituem ferramentas de fácil implementação que permitem uma análise preliminar das associações existentes entre as variáveis e, portanto um maior conhecimento do processo.

As médias e variâncias móveis mostraram ser uma ferramenta valiosa para a caracterização dinâmica das variáveis do processo. A sensibilidade da variância móvel para perceber modificações nos valores das variáveis é importante para mostrar graficamente as mudanças que experimentam essas variáveis em seu comportamento dinâmico.

A modelagem empírica mostrou que o processo pode ser modelado com variáveis cuja determinação pode ser efectuada *on-line* (11 variáveis), com similar precisão do que quando se considera um maior número de variáveis (14 variáveis), entre as que se incluem variáveis cuja determinação se realiza *off-line*, com uso de equipamento de laboratório.

7.6. REFERÊNCIAS BIBLIOGRÁFICAS

1. Atala, D.I.P. **Montagem, Instrumentação, Controle e Desenvolvimento Experimental de um Processo Fermentativo extrativo de Produção de Etanol**. Tese de Doutorado. Faculdade de Engenharia de Alimentos, Universidade Estadual de Campinas, 2004.
2. Atala, D.I.P. and Maciel, R. (2006). **Experimental development, optimization and control of a Continuous extractive production process for the ethanol production**. Postdoctoral research. Process 151335/2005-2, FAPESP, Brazil.
3. Atarassi, M.M. **Modelagem, simulação e controle de um evaporador flash acoplado a uma unidade de fermentação alcoólica contínua**. Dissertação de Mestrado. Universidade Estadual de Campinas. Brasil, 2005.
4. Cardona, C.A. and Sanchez, O.J. (2007). Fuel ethanol production: Process design trends and integration opportunities. *Bioresource Technology*, v. 98(2007), pp. 2415-2457.
5. Cardona, C.A., Gutiérrez, L.F. and Sánchez, Ó.J. 2008. **Process integration: Base for energy saving**. In: D.M. Bergmann, Editor, *Energy efficiency research advances*, Nova Science Publishers, Hauppauge, NY, USA, pp. 173-212.
6. Cinar, A., Parulekar, S.J., Undey, C. and Birol, G. (2005). **Batch Fermentation, Modeling, Monitoring and Control**. Markel Dekker: New York, NY, USA.
7. Cho, C.W., Hwang, S.T. (1991). Continuous membrane fermentor separator for ethanol fermentation. *Journal of Membrane Science*, v. 57 (1), pp. 21-42.
8. Chovau, S., Gaykawad, S., Straathof, A.J.J., Van der Bruggen, B. (2011). Influence of fermentation by-products on the purification of ethanol from water using pervaporation. *Bioresource Technology*, v. 102. pp. 1669–1674.
9. Christen, P, Minier, M, and Renon, H.(1990). Ethanol extraction by supported liquid membrane during fermentation. *Biotechnology and Bioengineering*, v. 36, pp. 116-123.
10. Cohen P.L.M. **Modelagem e simulação do processo de fermentação extrativa a vácuo com uma câmara de flash e separação do CO2 utilizando uma coluna de absorção**. Dissertação de Mestrado. Universidade Estadual de Campinas. Brasil, 2011.

11. Costa, A. C.; Atala, D. I. P.; Maciel filho, R.; Maugeri filho, F. (2001). Factorial design and simulation for the optimization and determination of control structure for an extractive alcoholic fermentation. *Process Biochemistry*, v. 37, p. 125-137.
12. Dias, F.M., Antunes, A. and Mota, A.M. (2003). **Regularization versus Early Stopping: a case study with a real system**. 2nd IFAC Conference Control Systems Design (CSD'03), Bratislava, República Eslovaca.
13. Duarte, E.R. **Avaliação e desenvolvimento de algoritmos de controle aplicado a um processo extrativo de fermentação alcoólica contínua**. Tese de Doutorado. Universidade Estadual de Campinas. Brasil, 2007.
14. Foresee, F.D. and Hagan, M. T. (1997). Gauss - Newton Approximation to Bayesian Learning. *Proceedings of the 1997 International Joint Conference on Neural Networks*, v. 3, 1930–1935.
15. Freitas, K.P. **Ferramentas estatísticas para monitoramento de modelos inferenciais**. Dissertação de Mestrado UFRJ, Brasil, 2009.
16. Jarzebski, A. B., Malinowski, J. J., and Goma, G. (1989). Modeling of ethanol fermentation at high yeast concentrations. *Biotechnology and Bioengineering*, v. 34, pp. 1225–1230.
17. Ljung, L. (1996). **System Identification**. In W. S. Levine (Ed.). The Control Handbook, CRC Press Inc., 1996, pp. 1033-1054.
18. MacKay, D.J.C. (1992). A Practical Bayesian Framework for Backpropagation Networks. *Neural Computation*, v. 4(3), pp. 448-472.
19. Milton, J.S. and Arnold, J.C. (1990). **Introduction to probability and statistics**, McGraw Hill: New York, USA.
20. Morgado, F.M.R. **Técnicas de controlo não-linear baseadas em Redes Neurais: do algoritmo à implementação**. Tese de Doutor. Universidade de Aveiro, Portugal. 2005.
21. Oppenheim, A.V. and Schafer. R.W.(1989). **Discrete-Time Signal Processing**. Prentice Hall: Englewood Cliffs, NJ, USA.
22. PRATA, D.M. (2009). **Reconciliação robusta de dados para monitoramento em tempo real**. PhD Tese. COPPE/UFRJ. Brasil.
23. Raghavarao, D. and Padgett, L. (2005). **Bloks designs, analysis, combinatorics and applications**. Singapore: World Scientific Publishing.

24. Rivera, E.C., Atala, D.I.P., Maugeri Filho, Costa, AC. Maciel Filho, R. 2010. Development of real-time state estimators for reaction-separation processes: A continuous flash fermentation as a study case. *Chemical Engineering and Processing*, v. 49 (2010), pp. 402-409.
25. Silva, F.L.H. **Modelagem, simulação e controle de fermentação alcoólica contínua extrativa**. Campinas. Tese Doutorado. Faculdade de Engenharia de Alimentos, Universidade Estadual de Campinas, Brasil, 1998.

CAPÍTULO 8. ANÁLISE DAS RELAÇÕES CAUSA- EFEITO EM MODELOS EMPÍRICOS DE REDES NEURAIS: INTERPRETAÇÃO DA INFORMAÇÃO CONTIDA NA CAIXA PRETA

O TEMA DEU ORIGEM A DOIS ARTIGOS:

- 1) A Method of Transfer Functions and Block Diagrams to Study the Contribution of Variables in Artificial Neural Network Process Models.
- 2) A New Framework to Study Cause and Effect Relationships with Neural Networks and its Application to Bioprocess Models.

Um artigo já foi publicado:

Boza-Condorena, E.G. and Costa, A.C. (2011). A method of transfer functions and block diagram to study the contribution of variables in artificial neural network process models. Lecture Notes in Engineering and Computer Science.

Proceedings of the World Congress on Engineering and Computer Science 2011,
v.II, pp. 708-713. San Francisco, CA. USA.

Prefácio

Dando sequência ao estudo do processo extrativo de fermentação alcoólica contínua é preciso conhecer a influência das variáveis de entrada na formação do produto obtido. No capítulo anterior, foram identificadas com base em estudos anteriores, mencionados nas referências bibliográficas, 14 variáveis de entrada para descrever o comportamento do processo de fermentação extrativa com flash. Neste caso, o uso do planejamento fatorial é dificultado porque o estudo é feito utilizando como entradas e saídas dados experimentais reais do processo sequenciais no tempo e porque no caso de conseguir uma adaptação para o uso de planejamento fatorial, um planejamento completo 2^{14} corresponderia a 16 384 ensaios.

No caso em estudo, os modelos utilizados para descrever o comportamento do processo são modelos de redes neurais, então uma alternativa adequada é usar a mesma rede neural para determinar o efeito das variáveis de entrada nas saídas. Neste capítulo, é apresentada uma metodologia desenvolvida para determinar a influência das entradas de uma rede neural nas saídas desta rede. Esta metodologia usa a teoria de funções de transferência e diagrama de blocos e pode ser aplicada a redes neurais com qualquer número de camadas e neurônios. Na literatura, metodologias similares encontradas precisam de maior desenvolvimento, nos casos relatados: 1) os pontos de vista são diversos, não existe tratamento sistemático, 2) fazem referência a poucas variáveis, 3) não utilizam a informação completa dos parâmetros da rede: pesos e bias e 4) as propostas se aplicam apenas a uma camada oculta.

A metodologia desenvolvida foi comparada com outras propostas na literatura para fazer o mesmo tipo de análise, tendo bom desempenho.

A metodologia desenvolvida foi usada para avaliação de dois casos de estudo relacionados à produção de etanol. No primeiro, compararam-se os resultados da análise com a metodologia proposta aos de um artigo publicado na literatura, onde a influência de entradas nas saídas de um processo foi avaliada

usando planejamentos de Plackett-Burman (PB). O uso de um modelo fenomenológico do processo para avaliar a real influência das entradas nas saídas mostrou que a metodologia proposta neste trabalho levou aos resultados corretos, enquanto o planejamento de PB resultou em análises incorretas para algumas das entradas.

Uma vez que, ficou demonstrado que a metodologia proposta leva a resultados corretos, ela foi aplicada para determinar a influência das entradas do modelo de redes neurais proposto para o processo de fermentação alcoólica extrativa nas saídas de interesse. O resultado foi utilizado posteriormente para reduzir o número de variáveis no modelo de processo com resultados satisfatórios, quando os modelos foram usados em esquemas de controle de processos.

8.1. PART 1: A METHOD OF TRANSFER FUNCTIONS AND BLOCK DIAGRAMS TO STUDY THE CONTRIBUTION OF VARIABLES IN ARTIFICIAL NEURAL NETWORK PROCESS MODELS

Abstract

In this paper it is proposed the use of transfer functions and block diagram algebra to describe cause and effect relationships in artificial neural network process models. Explicit formulae are derived for feedforward neural networks with an arbitrary number of inputs, outputs and hidden layers. This novel approach provides a general interpretation of the information existing in neuron interconnections. Numerical applications for the particular case of transfer functions expressed in terms of relative contributions of weights showed that the method was applicable in the evaluation of the effect of inputs on a given output, with good results.

Index Terms: Variables contribution; Backpropagation; Artificial neural networks; Process modeling.

8.1.1. INTRODUCTION

Knowledge of the relative importance of input variables in a neural network model provides two basic advantages [1] - [4]: a) The information can be used to build the optimal neural network model via the selection of inputs, which can improve the generalization capability of the model and allow for faster training of the neural network, with economic savings if measurement of the variables is expensive; b) It provides a better understanding of the process model since the irrelevant variables are identified.

Another point to be considered is that when a great number of input variables are available but the size of the training set is limited, the likelihood of overfitting is increased [5] [6]. In this case, the selection of variables that have the strongest influence on the output is of extreme importance.

Many researchers have tried to express the importance of input variables in neural network models and proposed the use of various methods [3][4][7] -[10].

In the present paper it is described a method that uses transfer functions and block diagrams to express input-output dependencies in ANNs, then it is shown that by using this method the concept of analyzing weights to estimate the importance of the inputs can be extended to topological net structures with multiple inputs, multiple outputs and multiple hidden layers.

8.1.2. THE TRANSFER FUNCTION METHOD

8.1.2.1. BASIC CONCEPTS

8.1.2.1.1. Transfer functions

A transfer function **G** is a mathematical statement that relates an input, **x**, to an output, **y**, of a system [11][12] .

$$G = \frac{y}{x} \quad (8.1.1)$$

The transfer function **G** transforms the input **x** (cause) into an output **y** (effect) as shown in (8.1.2).

$$y = G * x \quad (8.1.2)$$

8.1.2.1.2. Block diagrams

The block diagram can be used to describe cause and effect relationships throughout a dynamic system [13]. Figure 8.1.1 shows the block diagram for (8.1.2).

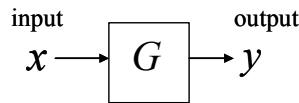


Figure 8.1.1. Transfer function

The transfer functions in block diagrams are represented by blocks. The lines that interconnect blocks represent signals that flow interconnecting the system elements. Other components are the pickoff point and the summing point.

The four components of a block diagram are shown in Figure 8.1.2. A signal from input x_1 flows into the pickoff point (**P**) and two signals flow out of the point. In the summing point (**S**) two signal flows are added, the result is one signal that flows out of the point.

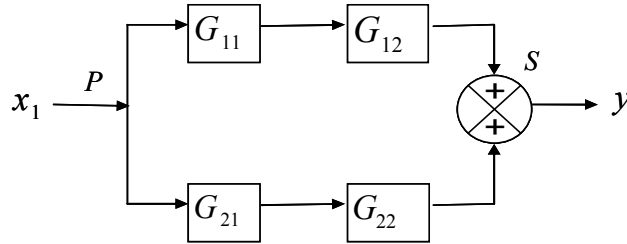
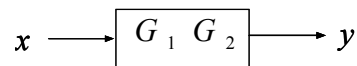
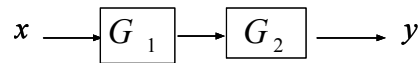


Figure 8.1.2. Blok diagram for transfer functions between input x_1 and output y .

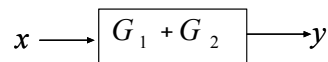
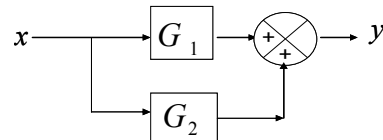
8.1.2.1.3. Basic rules of block diagram algebra

The cause and effect relationship between input x_1 and output y from the block diagram in Figure 8.1.2 is obtained using two basic rules of block diagram algebra, which are: blocks in series combine with each other by multiplication and blocks in parallel combine with each other by algebraic addition. Figure 8.1.3 and Figure 8.1.4 show these two basic rules and the mathematical representation.



$$y = G_1 * G_2 * x = G_2 * G_1 * x$$

Figure 8.1.3. Rule 1: blocks in series



$$y = (G_1 + G_2)x = G_1 * x + G_2 * x$$

Figure 8.1.4. Rule 2: blocks in parallel

Equation (8.1.3) shows the result of applying the basic rules of block diagram algebra first to the blocks in series and then for the resulting blocks in parallel of Figure 8.1.2. Equation (8.1.4) is the result of applying the associative property and Equation (8.1.5) is the mathematical statement of the transfer function between input x_1 and output y . Figure 8.1.5 shows the reduced block diagram obtained from Figure 8.1.2.

$$y = G_{11} * G_{12} * x_1 + G_{21} * G_{22} * x_1 \quad (8.1.3)$$

$$y = (G_{11} * G_{12} + G_{21} * G_{22}) * x_1 \quad (8.1.4)$$

$$\frac{y}{x_1} = G_1 = (G_{11} * G_{12} + G_{21} * G_{22}) \quad (8.1.5)$$

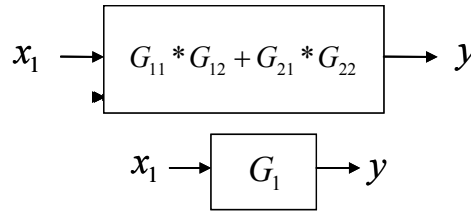


Figure 8.1.5. Reduced block diagram

Figure 8.1.5 shows that the effect of the input x_1 on the output y is expressed by a transfer function G_1 .

8.1.2.2. GENERALIZED EQUATIONS FOR THE EFFECT OF INPUTS ON THE OUTPUTS (E_{OI})

Notation used

Figure 8.1.6 shows the notation used for transfer function in neural networks structure.

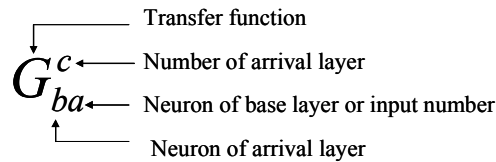


Figure 8.1.6. Notation used for transfer function the effect of the input on the output (E_{OI})

Figure 8.1.7 shows a portion of a neural network structure with one input, two neurons in the hidden layer and one neuron in the output layer. It also presents the transfer function associated to each connection between neurons of adjacent layers.

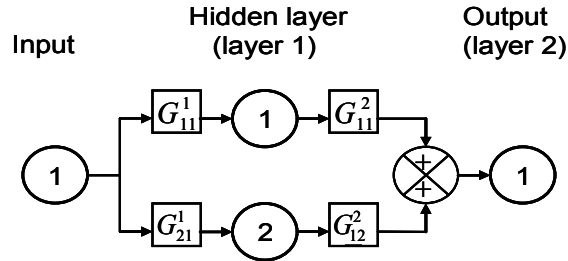


Figure 8.1.7. Block diagram to show transfer functions in the neural network structure.

From Figure 8.1.7, we obtain the effect of the input on the output (E_{OI}), using the basic rules of block diagram algebra.

$$E_{OI} = G_{11}^1 * G_{11}^2 + G_{21}^1 * G_{12}^2 \quad (8.1.6)$$

Figure 8.1.8 shows a neural network with two layers and a topology 3:2: 2.

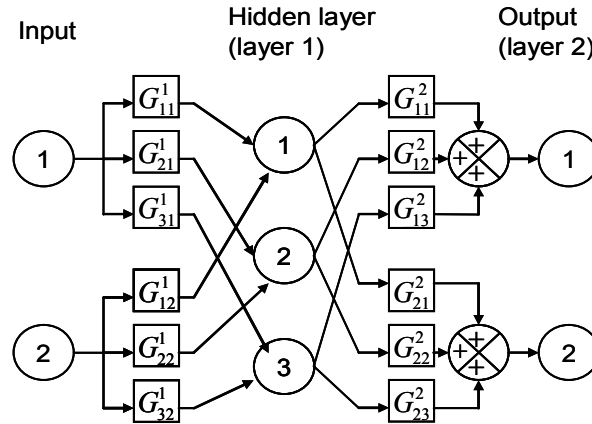


Figure 8.1.8. Transfer functions in neural network with two inputs, three neurons in the hidden layer and two neurons in the output layer

From Figure 8.1.8, the transfer functions for the effects of each input on the output are the following:

$$E_{11} = G_{11}^1 * G_{11}^2 + G_{21}^1 * G_{12}^2 + G_{31}^1 * G_{13}^2 \quad (8.1.7)$$

$$E_{12} = G_{12}^1 * G_{11}^2 + G_{22}^1 * G_{12}^2 + G_{32}^1 * G_{13}^2 \quad (8.1.8)$$

$$E_{21} = G_{11}^1 * G_{21}^2 + G_{21}^1 * G_{22}^2 + G_{31}^1 * G_{23}^2 \quad (8.1.9)$$

$$E_{22} = G_{12}^1 * G_{21}^2 + G_{22}^1 * G_{22}^2 + G_{32}^1 * G_{23}^2 \quad (8.1.10)$$

From (8.1.7) to (8.1.10) we can infer that the general equation for effects from input **I** on the output **O**, when the feedforward neural network has two layers, with **n1** neurons in the hidden layer is,

$$E_{OI} = \sum_{j=1}^{n1} G_{jI}^1 * G_{Oj}^2 \quad (8.1.11)$$

With a similar procedure general equations for more complex topologies were obtained.

The general equation for effects from input **I** on the output **O**, when the feedforward neural network has three layers, with **n1** neurons in hidden layer 1 and **n2** neurons in layer 2 is:

$$E_{OI} = \sum_{i=1}^{n1} \sum_{j=1}^{n2} G_{iI}^1 G_{ji}^2 G_{Oj}^3 \quad (8.1.12)$$

The general equation for the effect of input **I** on the output **O**, when the neural network has **r** layers, with **n(1)** neurons in hidden layer **1**, **n(2)** neurons in the hidden layer **2**, **n(r-1)** neurons in the **r-1** hidden layer, is:

$$E_{OI} = \sum_{i(1)=1}^{n(1)} \dots \sum_{i(r-2)=1}^{n(r-2)} \sum_{i(r-1)=1}^{n(r-1)} G_{i(1)I}^1 \dots G_{i(r-1)i(r-2)}^{r-1} G_{Oi(r-1)}^r \quad (8.1.13)$$

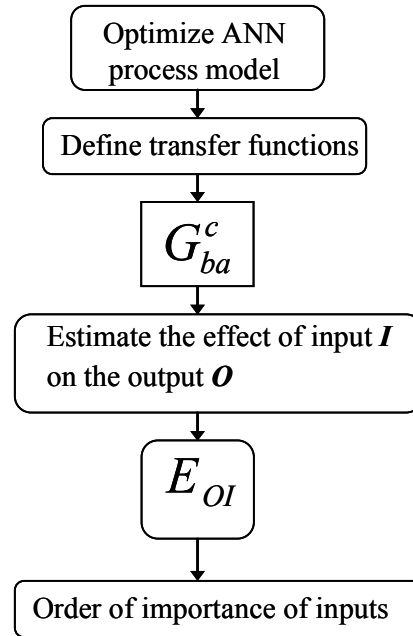


Figure 8.1.9. Algorithm to apply the method of transfer functions

8.1.3. NUMERICAL APPLICATION

8.1.3.1. PARTICULAR CASE: Relative contributions expressed in terms of absolute values of the weights as transfer function.

The weights associated to each connection between neurons of adjacent layers were used in transfer functions. Figure 8.1.10 shows the notation used.

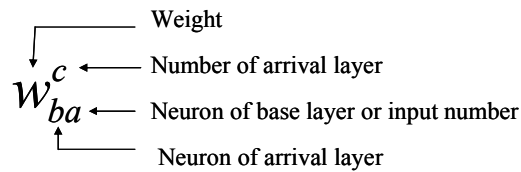


Figure 8.1.10. Notation used for weights in neural networks

Equation (8.1.14) shows the transfer function expressed as a relative contribution of the absolute values of the weights.

$$G_{ba}^c = \frac{|w_{ba}^c|}{\sum_{a=1}^{na} |w_{ba}^c|} \quad (8.1.14)$$

Where:

G is a transfer function, **c** is the number of the arrival layer, **b** is a neuron of the arrival layer, **a** is a neuron of the base layer or an input number, **na** is the number of neurons in the base layer or the inputs, $|w_{ba}^c|$ is an absolute value of the connection weight, $\sum_{a=1}^{na} |w_{ba}^c|$ is the sum of the weights that correspond to the connections between the neuron **b** of the arrival layer and each one of the **na** neurons of the base layer.

The relative contribution of an input **x** to an output **y**, when compared to the relative contributions of the other inputs to the same output, expresses the relative effect of input **x** on output **y**, and hence its importance in the set of inputs, as a factor to produce a response **y**.

Expressing (8.1.11) – (8.1.13) in terms of relative contributions, we obtain (8.1.15) – (8.1.17) to estimate the effects of the input variables on the output variables (E_{OI}). From (8.1.11) the general equation in terms of relative contributions as transfer functions for the effects of input **I** on the output **O**, when the feedforward neural network has 2 layers, with **n1** neurons in the hidden layer (layer 1), and **ni** inputs, is,

$$E_{OI} = \sum_{j=1}^{n1} \frac{|w_{jI}^1|}{\sum_{i=1}^{ni} |w_{ji}^1|} * \frac{|w_{Oj}^2|}{\sum_{j=1}^{n1} |w_{Oj}^2|} \quad (8.1.15)$$

From (8.1.12) the general equation in terms of relative contributions as transfer functions for the effects of input **I** on the output **O**, when the feedforward neural network has 3 layers, with **n1** neurons in the hidden layer 1, **n2** neurons in the hidden layer 2, and **ni** inputs, can be written:

$$E_{OI} = \sum_{i=1}^{n1} \sum_{j=1}^{n2} \frac{|w_{il}^1|}{\sum_{h=1}^{ni} |w_{ih}^1|} * \frac{|w_{ji}^2|}{\sum_{i=1}^{n1} |w_{ji}^1|} * \frac{|w_{Oj}^3|}{\sum_{j=1}^{n2} |w_{Oj}^2|} \quad (8.1.16)$$

Finally, from (8.1.13), the general equation in terms of relative contributions as transfer functions for the effect of input **I** on the output **O**, when the neural network has **r** layers, **ni** inputs, **n(1)** neurons in layer 1, **n(2)** neurons in layer 2, ... **n(r-1)** neurons in layer **r-1** is:

$$E_{OI} = \sum_{i(1)=1}^{n(1)} \dots \sum_{i(r-2)=1}^{n(r-2)} \sum_{i(r-1)=1}^{n(r-1)} \frac{|w_{i(1)I}^1|}{\sum_{g=1}^{ni} |w_{i(1)g}^1|} * \dots$$

$$\dots * \frac{|w_{i(r-1)i(r-2)}^{r-1}|}{\sum_{i(r-2)=1}^{n(r-2)} |w_{i(r-1)i(r-2)}^{r-1}|} * \frac{|w_{Oi(r-1)}^r|}{\sum_{i(r-1)=1}^{n(r-1)} |w_{Oi(r-1)}^r|} \quad (8.1.17)$$

8.1.3.1.1. Example 1: Neural network with 5: 2: 1 topology, with five inputs, two hidden neurons and one output neuron

In order to test the ability of the proposed method to determine the order of importance of the effects of inputs on outputs and to compare with results obtained by using other methods the following procedure was used:

- 1) To use an experimental data set of a process. In this example the experimental data set was obtained from the literature and corresponds to the process described by Isar *et al.* (2006). The experimental data set of the process contain 5 input variables and 1 output variable.

- 2) To use the deterministic mathematical model of the process in order to obtain the reference order for the importance of the effects of inputs on the output.

In this example the reference order was established using the coefficients of the mathematical model (expressed in terms of coded factors) obtained with the experimental design methodology by Isar *et al.* (2006) [14]. The mathematical relationship of the input variables and the output variables is a second order polynomial equation where the linear coefficients are the following:

Input variables	A	B	C	D	E
Linear coefficients	-0.93	-0.32	-1.31	-0.35	+1.38

The model was appropriately validated and verified by the authors, who reported a coefficient of determination (R^2) equal to 0.9851, a high F-value, a non significant lack of fit. They reported also that the adequate precision value, which measured the 'signal to noise ratio' was 53.247, which indicated an adequate signal, because a ratio greater than 4 was desirable, so is ensured a good adjustment of the model to the experimental data and also that this model could be used to navigate the design space.

- 3) To select the best neural network model to the experimental data set.

In this case 360 simulations were performed with different network architectures. The network architecture 5:2:1, with five input neurons, two neurons in the intermediate layer and one neuron in the output layer, was the most appropriate working with the function 'tansig' in the hidden layer and 'purelin' in the output layer. Training was performed using the Levenberg-Marquardt algorithm. Artificial neural networks with the highest values of determination coefficient R^2 and minimum mean square error (mse) have shown more consistency in results. The training of the neural network was performed using 75% of the experimental design data (24 of 32 experiments). In the validation of the networks the 25% remaining input data were used, which was unknown to the network (8 experiments). The accuracy of the

selected ANN model to describe the experimental data was also tested by determining the predicted optimum for the process using the selected neural network model and different optimization methods (genetic algorithms and others) and comparing the results with the experimental value reported by Isar et al. (2006). The results of these tests were excellent and informed that the selected model was able to describe the experimental data accurately.

- 4) To use the connection weights of the selected ANN model and to determine the order of importance of the effects of all the inputs on the output using the proposed transfer function method. In this example the appropriate equation is Equation 8.1.15.
- 5) To use the connection weights of the selected model and to determine the order of importance of the effects of all the inputs on the output using other methods based on connection weights. i.e., the algorithm proposed by Garson (1991) for a two layers ANN takes the distribution of weights in the neural network into account, with the following equation:

$$I_j = \frac{\sum_{m=1}^{N^h} ((|w_{jm}^{ih}| / \sum_{K=i}^{N^i} |w_{km}^{ih}|) * |w_{mn}^{ho}|)}{\sum_{K=1}^{K=N^i} \left\{ \sum_{m=1}^{m=N^h} ((|w_{Km}^{ih}| / \sum_{K=i}^{N^i} |w_{km}^{ih}|) * |w_{mn}^{ho}|) \right\}}$$

In the equation, I_j is the j-th input variable relative importance, N^i and N^h are the input and hidden neuron numbers respectively, and w stands for the neural network weights, and i , h and o superscripts refer to the input, hidden and output layers. The subscripts k , m and n refer, respectively, to the input, hidden and output layers.

The importance of the inputs in the algorithm proposed by Olden (2004) for a two layers ANN is represented by the following equation:

$$I_{i/o} = \sum_{hn=1}^{Hn} w_{hni} * w_{ohn}$$

Where w is the neural network raw connection weight value. Hn is the number of neurons in the hidden layer. The subscripts hn , i , and o refer, respectively, to a neuron in the hidden layer, input layer and output layer.

- 6) To compare the obtained results in steps 4) and 5) with the reference values from step 2). In this example for comparison purposes all the results were normalized in the range [0.1, 0.9].

The connection weights of the selected neural network are shown in Table 8.1.1.

Table 8.1.1. Example 1: ANN weights associated to each connection between neurons

Weights (W_{ij})	Hidden layer		Output layer	
	Destination		Weights (W_{ij})	Destination
	Neuron 1	Neuron 2		Neuron 1
Source Inputs	j =1	J =2	Source	O =1
$I = 1$	4.2452	0.2094	j = 1	0.0436
$I = 2$	3.2627	0.0638	j = 2	-1.4816
$I = 3$	14.7474	0.2515		
$I = 4$	9.5223	0.0907		
$I = 5$	0.9688	-0.2702		
$\sum_{i=1}^{k=5} w_{ji}^1 $	32.7464	0.8856	$\sum_{j=1}^{n=2} w_{oj}^2 $	1.5252

The effects of the inputs on the output, E_{OI} , are obtained using (8.1.15). In this case: $n_i = 5$ (number of inputs), $n_1 = 2$ (number of neurons in the hidden layer), $O = 1$ (number of outputs).

Figure 8.1.11 shows the order of importance of the effects of all the inputs on the output using the transfer function method (Equation 8.1.15), the Garson method [7] and the method that uses raw connection weight values [10].

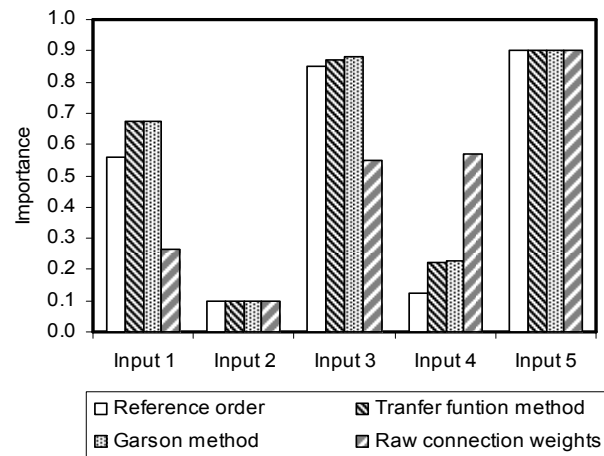


Figure 8.1.11. Comparison of the importance of the inputs' effects using different methods.

In Figure 8.1.11, in order to compare the obtained results with the reference values, all the data set was normalized in the range [0.1, 0.9]. The reference order was established using the absolute value of the coefficients of the mathematical model as was mentioned in step 2) of the explained procedure.

In this example, the transfer function method and the Garson method [7] predicted the same order of importance of the inputs as the reference mathematical model [14]. The use of the raw connection weight values proposed by Olden *et al.* (2004) [10] predicted the order of importance of the inputs erroneously. Two factors that probably contributed to the wrong result were the big difference in magnitude between the highest and the lowest weight values in Table 8.1.1 and the great dispersion of weight values around the mean (variance = 23.33). The use of relative contributions expressed in terms of absolute values of connection weights, as transfer functions, as proposed in this work, led to better results in this case.

8.1.3.1.2. Example 2: Neural network with 11:5:5:1 topology, to model a batch fermentation process for bioethanol production

8.1.3.1.2.1. Introduction

The second case study is a process from the work of Andrade *et al.* (2009) [15]. These authors proposed a mathematical model to describe a batch ethanol fermentation process, and used Plackett-Burman designs [16] to estimate the order of importance of the effects of the kinetic parameters of the mathematical model on the concentrations of biomass (X) and ethanol (P), as well as on the substrate consumption time (t).

In this numerical application, the order of importance of the kinetic parameters on one of the outputs of the process (the substrate consumption time) was determined using the transfer function method, and the results were compared with those obtained by Andrade *et al.* (2007) [15] using Plackett-Burman (PB) designs. Both results were compared with the reference order, obtained by applying the 'perturb' method to the mathematical model proposed by Andrade *et al.* (2007) [15]. The reference values for the kinetic parameters were taken from Andrade *et al.* (2007) [15].

The 'perturb' method was applied to build the profiles of the effects of the following inputs on the response: fermentation time for total substrate consumption:

- 1) $\mu_{\max}$ = maximum specific growth rate (h^{-1})
- 2) $X_{\max}$ = biomass concentration when cell growth ceases (kg/m^3)
- 3) P_m = product concentration when cell growth ceases (kg/m^3)
- 4) Y_x = limit cellular yield (kg/kg)
- 5) Y_{px} = yield of product based on cell growth (kg/kg)
- 6) K_s = substrate saturation parameter (m^3/kg)
- 7) K_i = substrate inhibition coefficient (m^3/kg)
- 8) m_p = ethanol production associated with growth ($\text{kg}/[\text{kg}.\text{h}]$)
- 9) m_x = maintenance parameter ($\text{kg}/[\text{kg}.\text{h}]$)

10) m = parameter used to describe cellular inhibition

11) n = parameter used to describe the inhibition by product.

The changes used in the 'perturb' method consisted of variations of -20%, -10%, +10% and +20%, which were produced in the selected input variable around the reference values of the kinetic parameters, while keeping all the other inputs constant. The results are shown in Figure 8.1.12.

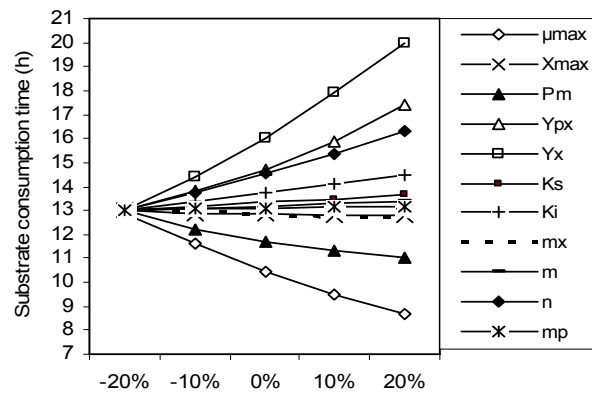


Figure 8.1.12. Profiles of the variation of the output variable according to the changes in the input variables

Figure 8.1.12 shows that there are six inputs with greater influence on the output: Y_x , Y_{px} , μ_{max} , n , P_m , and K_i . The most important input affecting the output (fermentation time for total substrate consumption) was Y_x , the limit in cell yield (kg/kg).

In order to quantitatively express the information shown in Figure 8.1.13, the S (index) was defined, which measures the cumulative sum of changes in the slope value. The slope is the ratio of the output change to the difference between two consecutive scaled values of inputs, and is calculated by the following equation:

$$S(index) = \sum_{i=1}^{ni} |O_{i+1} - O_i| / |sI_{i+1} - sI_i| \quad (8.1.18)$$

Where: sI = scaled value of input, O = output value,
 i = interval number between two consecutive scaled values of inputs.

8.1.3.1.2.2. ANN model topology

With the purpose of applying the transfer functions method, a feedforward neural network with 11:5:5:1 topology, with 11 inputs, 5 neurons in the first hidden layer, 5 neurons in the second hidden layer and 1 output, was selected to model the relationships between the above mentioned factors and the response: fermentation time for total substrate consumption.

The number of neurons in the hidden layers was determined using the cross-validation technique, in order to avoid model overfitting and to achieve good generalization from the training data set. This technique splits the data sample into a training data set and a validation data set. The performance of the trained neural network was evaluated by the ability to predict the elements of the validation data set, which was expressed in terms of the mean square error (mse).

$$mse = \frac{1}{N} \sum_{k=1}^N (t_k - ev_k^O)^2 \quad (8.1.19)$$

In (8.1.19), k is the number of data points in the validation data set, which varies between 1 and N , t_k is the k th *target value*, and ev_k^O is the k th net estimated output value.

8.1.3.1.2.3. Comparison of the results for the importance of the effects of input variables on the outputs

The relative importance of all the input variables on a given output using the weight values for this particular ANN structure was obtained using (8.1.12).

When the transfer functions are fractional contributions expressed in terms of absolute values of weights, (8.1.12) adopts the form of (8.1.16). In this case: ni = 11 (number of inputs); $n1$ = 5 (number of neurons in the first hidden layer); $n2$ =

5 (number of neurons in the second hidden layer); $O = 1$ (output); and $I = 1, 2, 3, \dots, 11$ (output number).

The orders of importance of the six inputs with greater influence on the output, obtained using the transfer function method and the Plackett-Burman design [15], are shown in Figure 8.1.13.

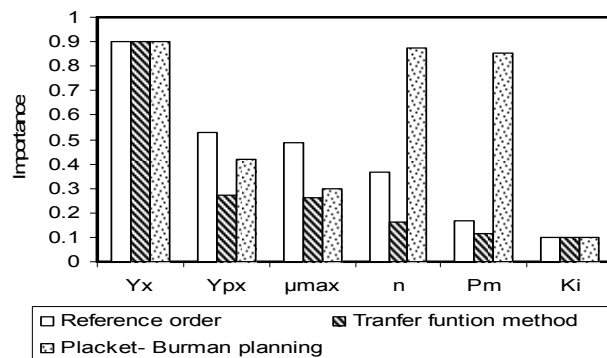


Figure 8.1.13 Comparison of the importance of the inputs' effects on the fermentation time for total substrate consumption, using different methods.

Figure 8.1.13 shows the order of importance of the kinetic parameters on substrate consumption time calculated by the perturb method, by the transfer function method and by PB design methodology. As the perturb method performs disturbances of different magnitudes in all of the parameters by simulation of the deterministic mathematical model developed by Andrade et al. (2009) for the process, it can be considered as giving the reference order of importance of the parameters: $Yx > Ypx > \mu_{max} > n > Pm > Ki$. The order of importance determined using PB design is $Yx > n > Pm > Ypx > \mu_{max} > Ki$, which is not correct. It should be emphasized that the Plackett Burman methodology was applied by Andrade et al. (2009) using the same deterministic model used in the present work to simulate the process. It can be seen from the figure that the transfer function method with the absolute weight values proposed in this work determined the order of importance correctly.

8.1.4. CONCLUSION

The transfer function method provided a general framework to study the information contained in the weights of ANN based models, which was useful for estimating the importance of inputs on outputs, and facilitated the derivation of applicable equations for models based on neural networks with two or more layers of neurons, extending the possibilities of analyzing these cases with respect to the weight based methods found in the literature.

Transfer functions between adjacent neurons in ANNs could be used to relate inputs and outputs. In numerical applications, good results were obtained when using relative contributions expressed in terms of absolute values of weights as transfer functions, in order to estimate the effects of inputs on outputs.

A good neural network model of a process gives secure information about the relative importance of the input variables, which highlights the importance of the availability of good models to describe the dynamic behavior of chemical and biotechnological processes.

The results of the case study showed that the proposed method, which uses transfer functions and the rules of block diagram algebra to estimate the importance of the effects of inputs on outputs in ANN based models leads to better results than the application of Plackett Burman designs.

8.1.5. REFERENCES

- [1] A. Engelbrecht, and I. Cloete, "Feature extraction from feedforward neural networks using sensitivity analysis," In *Proceedings of the International Conference on Systems, Signals, Control, Computers*, Durban, South Africa, 1998, pp. 221-225.
- [2] Z. Reitermanová, "Feedforward Neural Networks – Architecture Optimization and Knowledge Extraction," In *WDS'08 Proceedings of Contributed Papers*,

Part I: Mathematics and Computer Sciences, *Prague, Czech Republic, 2008*, pp. 159–164.

- [3] Z. Huang, H. Chen, C. J. Hsu, W. H. Chen and S. Wu, “Credit rating analysis with support vector machines and neural networks: a market comparative study,” *Decis Support Syst.*, vol. 37, no. 4 , pp. 543–558, Sept. 2004.
- [4] J. Tikka, “***Simultaneous input variable and basis function selection for RBF networks***,” *Neurocomputing*, vol. 72 no. 10-12, pp. 2649-2658, Jun. 2009.
- [5] N. Chapados, and Y. Bengio, “Input decay: Simple and effective soft variable selection,” In *Proceedings of the 2001 IEEE/INNS International Joint Conference on Neural Networks*, Washington D.C., 2001, pp.1233 – 1237.
- [6] T. Similä and J. Tikka, “Combined input variable selection and model complexity control for nonlinear regression,” *Pattern Recognit Lett*, vol. 30, no. 3, pp. 231-236, Feb. 2009.
- [7] G. D. Garson, “Interpreting neural-network connection weights,” *AI Expert*, vol. 6, no. 7, pp. 47-51, Apr. 1991.
- [8] P. Hajela and Z. P. Szewczyk, “Neurocomputing strategies in structural design on analyzing weights of feedforward neural networks,” *Struct Multidisc Optim.*, vol. 8, no. 4, pp. 236-241, Dec. 1994.
- [9] Y. Dimopoulos, P. Bourret and S. Lek, “Use of some sensitivity criteria for choosing networks with good generalization ability,” *Neural Process Lett*, vol. 2, no. 6, pp. 1-4. 1995.
- [10] J. D. Olden, M. K. Joy and R. G. Death, “An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data,” *Ecol Modell*, vol. 178, no. 389-397, Nov. 2004.
- [11] W. Luyben, **Process modeling, simulation and control for chemical engineers**. New York, NY: McGraw-Hill, 1990.

- [12] D. E. Seborg, T.F. Edgar, and D.A Mellichamp, **Process Dynamics and Control**. Hoboken, New Jersey, NJ: John Wiley & Sons, Inc., 2003.
- [13] C. Mei, "On Teaching the Simplification of Block Diagrams," *Int J Eng Ed*, vol. 18, no. 6, pp. 697-703, My. 2002.
- [14] J. Isar, A. Lata, S. Saurabh and S.A. Rajendra, "Statistical method for enhancing the production of succinic acid from *Escherichia coli* under anaerobic conditions," *Bioresour Technol*, vol. 97, pp. 1443–1448, Sept. 2006.
- [15] R. R. Andrade, "**Procedure for development of a robust mathematical model for alcoholic fermentation process**," M.S. thesis, State University of Campinas, Sao Paulo, Brazil, 2007.
- [16] R.L. Plackett and J.P. Burman, "The design optimum multifactorial experiments," *Biometrika*, vol. 33, no.4, pp. 305-325, 1946.

8.2. PART 2: A NEW FRAMEWORK TO STUDY CAUSE AND EFFECT RELATIONSHIPS WITH NEURAL NETWORKS AND ITS APPLICATION TO BIOPROCESS MODELS

Abstract

Knowledge of the effects of the factors on the responses in chemical and biotechnological processes is important in the modeling, optimization and construction of process control architectures. The present paper proposes the use of transfer functions and block diagram algebra to study the cause and effect relationships in artificial neural network process models. A shortcut procedure is presented where matrix equations are derived for feedforward neural networks with two and three layers, and generalized for any number of inputs, outputs and hidden layers. This novel approach provides a broader interpretation of the information existing in neuron interconnections. Numerical applications with the particular case of using relative contributions of weights and biases in transfer functions and comparisons of the results with component principal analysis showed that the method was also applicable to the dynamic behavior modeling of a continuous extractive fermentation process with good results.

Keywords: Process Systems Engineering, Process Modeling, Artificial neural networks, Input variable effect

8.2.1. INTRODUCTION

Modeling with neural networks is a powerful method to represent nonlinear relationships amongst inputs and outputs. However, they are often considered as black boxes because the mechanism by which they predict the output variable is not well explained (GEVREY *et al.*, 2003). For this reason, it is also difficult to interpret the relative importance of the input variables in neural network models. This limitation has been a frequent complaint with respect to neural networks in the literature (HUANG *et al.*, 2004).

Knowledge of the relative importance of input variables in a neural network model provides two basic advantages (ENGELBRECHT & CLOETE, 1998; HUANG *et al.*, 2004; TIKKA, 2009) : a) The information can be used to build the optimal neural network model via the selection of inputs, which can improve the generalization capability of the model and allow for faster training of the neural network, with economic savings if measurement of the variables is expensive; b) It provides a better understanding of the process model since the irrelevant variables are identified.

Another point to be considered is that when a great number of input variables are available, but the size of the training set is limited, the likelihood of overfitting is increased (CHAPADOS & BENGIO, 2001; SIMILÄ & TIKKA, 2009). In this case, the selection of variables that have the strongest influence on the output is of extreme importance.

However, selecting relevant variables from a given set of possibilities and estimating their relative importance is one of the great challenges in modeling non linear systems. Given its importance in many real world applications, effective methods to achieve this goal are required and this has become the focus of research in multivariable systems.

Research activity in ANNs has shown that the incoming and outgoing weights of the neuron contain embedded information which is valuable: a) to optimize the ANN topology by evaluating the relevance of single network elements in 'pruning' methods (HAGIWARA, 1994; REITERMANOVÁ, 2008); b) to extract symbolic rules by using appropriate ANN rule extraction algorithms (TAHA & GHOSH, 1999; GHAI, 2006) and c) to assess the effect of all input quantities on a given output in ANN based models (GARSON, 1991; OLIVEIRA-ESQUERRE *et al.*, 2002; GEVREY *et al.*, 2003; SHARMA *et al.*, 2004; OLDEN *et al.*, 2004, KAUR *et al.*, 2006; CHEN, J.C.H. & CHEN, W.H., 2008; GUIMARÃES *et al.*, 2008; SHI *et al.*, 2009). However, in the literature these methods do not make use of the information

contained in the ANN biases and only consider ANN topologies with one hidden layer.

In this work we propose a novel approach that uses transfer functions and block diagram algebra to extract information about the cause and effect relationship amongst input and output variables from process models based on feedforward neural networks. The proposed approach extends easily the analysis and interpretation of input-output dependencies in ANN process models to ANN topologies with multiple inputs, multiple outputs and multiple hidden layers. Also, the bias neurons information is taken into account.

The article is organized as follows: 1) Basic concepts of transfer functions and rules of block diagram algebra are briefly reviewed. 2) The proposed method, termed method of transfer functions, is presented, with an application example of the rules of block diagram algebra in the structure of an artificial neural network. 3) An algorithm to apply a shortcut procedure for the method of transfer functions is showed. 4) Numerical applications with ANN bioprocess models are presented to illustrate the performance of the proposed method in the particular case of using relative contributions of weights and biases expressed in terms of absolute values as transfer functions 5) Conclusions are given at the end of the article.

8.2.1.1. Transfer functions

A transfer function **E** is a mathematical statement that relates an input, **x**, to an output, **y**, of a system (LUYBEN, 1990; SEBORG *et al.*, 2003). The transfer function **E** transforms the input **x** (cause) into an output **y** (effect) as shown in Eq. (8.2.1).

$$y = E * x \quad (8.2.1)$$

8.2.1.2. Block diagrams

In control engineering, the block diagram is a primary tool that can be used with transfer functions to describe cause and effect relationships throughout a dynamic system (MEI, 2002). Figure 8.2.1 shows the block diagram for Eq. (8.2.1).

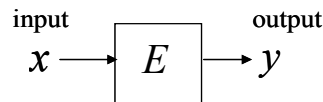


Figure 8.2.1. Transfer function.

Two basic rules of block diagram algebra, are: blocks in series combine with each other by multiplication, and blocks in parallel combine with each other by algebraic addition. Figure 8.2.2 and Figure 8.2.3 show these two basic rules and the mathematical representation.

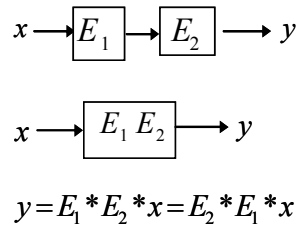


Figure 8.2.2 Rule 1: Blocks in series.

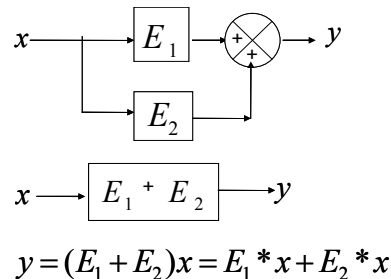


Figure 8.2.3 Rule 2: Blocks in parallel.

The transfer functions in block diagrams are represented by blocks. The lines that interconnect blocks represent signals that flow and interconnect the system elements. Other components are the pickoff point (**P**) and the summing point (**S**). The four components of a block diagram are shown in Figure 8.2.4: a signal from the input x flows into the pickoff point (**P**) and three signals flow out of the point. At the summing point (**S**) three signal flows are added and the result is one signal that flows out of the point.

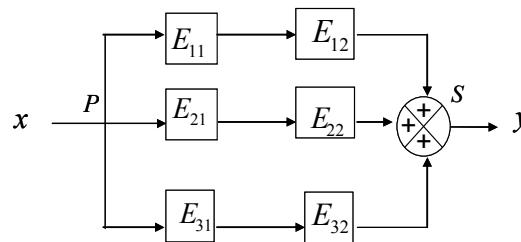


Figure 8.2.4. Blok diagram for transfer functions between input x and output y

The cause and effect relationship between the input x and the output y from the block diagram in Figure 8.2.4 are obtained using the above mentioned basic rules of block diagram algebra. Figure 8.2.5 shows the reduced block diagram.

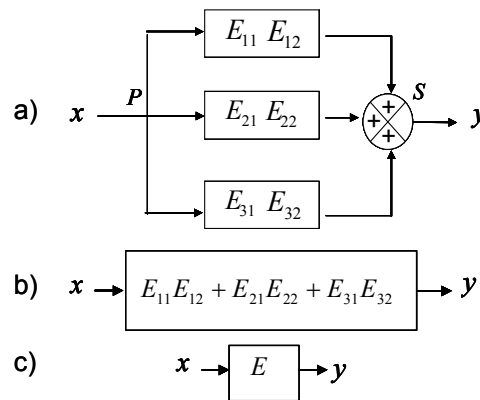


Figure 8.2.5. Block diagram reduction for Figure 8.2.4

Figure 8.2.5a shows the result of applying the first rule to the blocks in series in Figure 8.2.4, Figure 8.2.5b shows the result of applying the second rule to the remaining blocks in parallel, Figure 8.2.5c shows that the effect of the input x on the output y in the reduced block diagram is expressed by a transfer function **E**.

The block diagram reduction can be expressed by the following equations (8.2.2) – (8.2.4). Eq. (8.2.2) is a mathematical representation for Figure 8.2.5a and Eq. (8.2.3) is the one for Figure 8.2.5b. Eq. (8.2.4) is the mathematical statement of the transfer function between the input $\mathbf{x}$ and the output $\mathbf{y}$.

$$y = E_{11}E_{12}x + E_{21}E_{22}x + E_{31}E_{32}x \quad (8.2.2)$$

$$y = (E_{11}E_{12} + E_{21}E_{22} + E_{31}E_{32})x \quad (8.2.3)$$

$$\frac{y}{x} = E = E_{11}E_{12} + E_{21}E_{22} + E_{31}E_{32} \quad (8.2.4)$$

8.2.2. THE METHOD OF TRANSFER FUNCTIONS TO STUDY THE EFFECTS OF INPUTS ON THE OUTPUTS IN ARTIFICIAL NEURAL NETWORKS: SHORTCUT PROCEDURE

The method is based on the following assumptions. First, it assumes that in the ANN neuron interconnections, there is valuable information that can be used to estimate the effect of inputs on the outputs. Secondly, it assumes that the information existing between two neurons of adjacent layers expresses a cause and effect relationship between the input neuron and the output neuron. Thirdly, it assumes that between neurons of adjacent layers exists a SISO system with linear time-invariant components. Fourthly, it assumes that the MIMO ANN process models are systems that have linear and non linear time-invariant components and that the linear time-invariant components can be used to build a method based on transfer functions and block diagram algebra to assess the cause and effect relationships amongst inputs and outputs. Figure 8.2.6 shows the notation used in this work for transfer functions in neural networks structures.

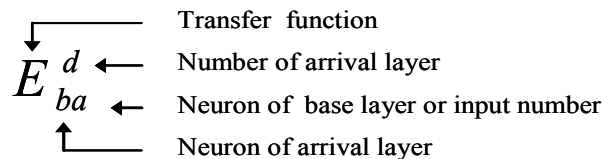


Figure 8.2.6. Notation used for transfer functions.

Figure 8.2.7 shows a block diagram reduction in the SISO subsystem existing between neurons **a** and **b** of adjacent layers.

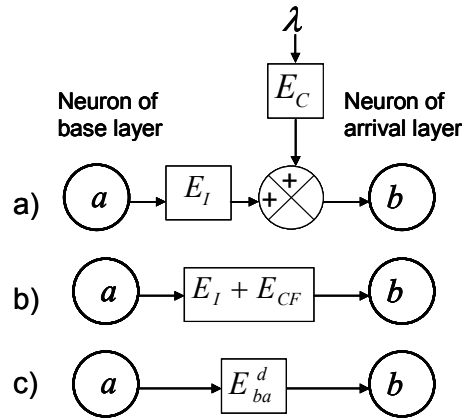


Figure 8.2.7. Block diagram reduction in the SISO subsystem between the neurons **a** and **b**

In Figure 8.2.7, the cause and effect relationship between neuron **a** and neuron **b**, has a causal relationship modifier represented by the transfer function E_I , and a correction factor represented by $E_{CF} = \lambda E_C$. Figure 8.2.7a and 8.2.7b shows that both transfer functions are added at the summing point applying the rule 2 for blocks in parallel. Figure 8.2.7c presents the final result represented by the transfer function E^d_{ba} , which expresses the causal relationship between the neuron **a** and the neuron **b**.

Figure 8.2.8 shows a portion of a neural network structure with one input, three neurons in the hidden layer and one neuron in the output layer. It also presents the transfer function associated with each connection between the neurons of adjacent layers.

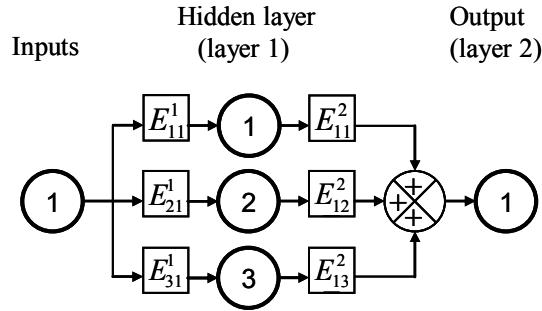


Figure 8.2.8. Block diagram to show transfer functions in the neural network structure.

In Figure 8.2.8 the effect of input 1 on output 1 in terms of transfer functions is obtained applying first rule 1 to the blocks in series and then rule 2 for blocks in parallel, to add the partial results at the summing point

$$\text{Effect (output 1 from input 1)} = E_{11}^1 E_{11}^2 + E_{21}^1 E_{12}^2 + E_{31}^1 E_{13}^2 \quad (8.2.5)$$

This result has two useful representations:

a) A matrix equation expressed as a matrix multiplication

$$\text{Effect (output 1 from input 1)} = E_{1 \times 3}^1 * E_{3 \times 1}^2 \quad (8.2.6)$$

$$\text{Where: } E_{1 \times 3}^1 = \begin{bmatrix} E_{11}^1 & E_{21}^1 & E_{31}^1 \end{bmatrix} \text{ and } E_{3 \times 1}^2 = \begin{bmatrix} E_{11}^2 \\ E_{12}^2 \\ E_{13}^2 \end{bmatrix}$$

In the matrix equation, the dimensions of each matrix are obtained in a straightforward way from the ANN topology, which in this case is 1: 3 : 1, with 1 input, three neurons in the hidden layer and 1 output. Then the dimensions of the first matrix are 1 x 3, and of the second matrix are 3 x 1.

The procedure that uses the matrix equations to obtain the effect of inputs on the outputs on process models based on ANN is proposed in this work and is termed as shortcut procedure of the method of transfer functions.

b) A summatorial equation. Eq. (8.2.5), which expresses the addition of three terms and can be represented by the following equation:

$$\text{Effect (output 1 from input 1)} = \sum_{j=1}^3 E_{j1}^1 E_{1j}^2 \quad (8.2.7)$$

Where j is the neuron number in the hidden layer ($j = 1, 2, 3$).

The procedure to solve the summatorial equations in the case of complex ANN topologies is time consuming and for this reason it is termed as long procedure of the method of transfer functions.

The following are examples of the application of the notion of transfer functions in ANNs and the rules of block diagram algebra, in order to obtain the effect of the inputs on the outputs.

Example 1: neural network with two layers and a 3: 2: 2 topology.

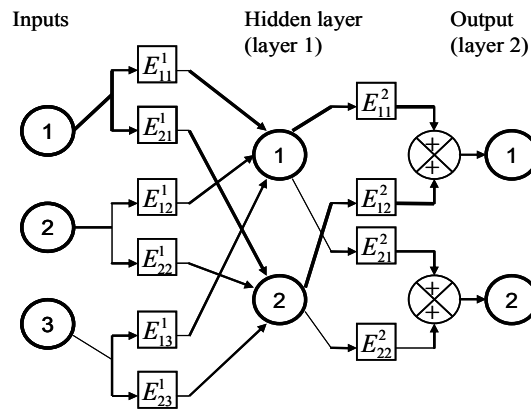


Figure 8.2.9 Transfer functions in neural network with three inputs, two neurons in the hidden layer and two neurons in the output layer.

To obtain from Figure 8.2.9 in a straightforward manner the effect of the selected input on the desired output by using the block diagram algebra we can: 1) select an input number and an output number in the final layer, 2) follow each flow of the signal from the selected input until to the same summing point that is next to the desired output multiplying the transfer functions of the blocks in séries, 3) At

the summing point add the obtained result from each signal flow. The addition result expresses the effect of the selected input on the desired output.

From Figure 8.2.9 the transfer functions for the effects of each input (**I**) on the output (**O**), E_{OI} , are the following:

a) Effects of inputs 1, 2 and 3 on the output 1:

$$E_{11} = E_{11}^1 E_{11}^2 + E_{21}^1 E_{12}^2 \quad (8.2.8)$$

$$E_{12} = E_{12}^1 E_{11}^2 + E_{22}^1 E_{12}^2 \quad (8.2.9)$$

$$E_{13} = E_{13}^1 E_{11}^2 + E_{23}^1 E_{12}^2 \quad (8.2.10)$$

b) Effects of inputs 1, 2 and 3 on the output 2:

$$E_{21} = E_{11}^1 E_{21}^2 + E_{21}^1 E_{22}^2 \quad (8.2.11)$$

$$E_{22} = E_{12}^1 E_{21}^2 + E_{22}^1 E_{22}^2 \quad (8.2.12)$$

$$E_{23} = E_{13}^1 E_{21}^2 + E_{23}^1 E_{22}^2 \quad (8.2.13)$$

The matrix representation of the transfer functions (Eqs. 8.2.8 – 8.2.13) is obtained with the following important steps:

- 1) Determine the ANN topology. From Figure 8.2.9 it is 3: 2 : 2
- 2) Determine the dimensions of the matrices, 3 x 2 for the first matrix and 2 x 2 for the second matrix
- 3) Write the matrix equation that represents the effect of inputs (**I**) on the outputs (**O**). The matrix equation is:

$$E_{OI} = E_{3 \times 2}^1 E_{2 \times 2}^2 \quad (8.2.14)$$

$$\text{Where: } E_{3 \times 2}^1 = \begin{bmatrix} E_{11}^1 & E_{21}^1 \\ E_{12}^1 & E_{22}^1 \\ E_{13}^1 & E_{23}^1 \end{bmatrix} \text{ and } E_{2 \times 2}^2 = \begin{bmatrix} E_{11}^2 & E_{12}^2 \\ E_{21}^2 & E_{22}^2 \end{bmatrix}$$

The result of the matrix multiplication reproduces Eqs. (8.2.8) –(8.2.13).

Example 2: neural network with a 2:3: 3: 2 topology

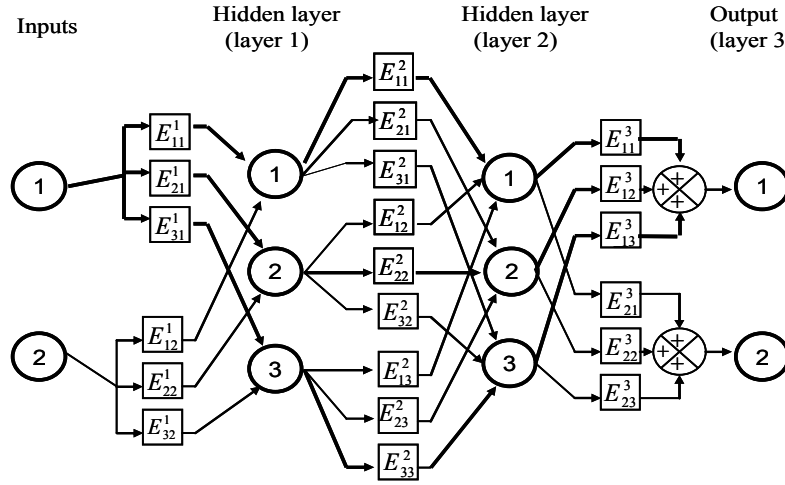


Figure 8.2.10. Transfer functions in neural network with two inputs, three neurons in the hidden layer 1, three neurons in the hidden layer 2 and two neurons in the output layer.

From Figure 8.2.10 the shortcut procedure of the method of transfer functions provides easily the matrix equation that represents the effect of inputs on the outputs:

- 1) The ANN topology from Figure 8.2.10 is 2: 3 : 3 : 2.
- 2) There are three layers of neurons, thus there are three matrices. The dimensions of the matrices are: first matrix 2 x 3, second matrix 3 x 3 and third matrix 3x2.
- 3) The matrix equation that represents the effects of the inputs on the outputs is:

$$\mathbf{E}_{OI} = \mathbf{E}^1_{2 \times 3} \mathbf{E}^2_{3 \times 3} \mathbf{E}^3_{3 \times 2} \quad (8.2.15)$$

$$\text{Where: } \mathbf{E}^1_{2 \times 3} = \begin{bmatrix} E^1_{11} & E^1_{21} & E^1_{31} \\ E^1_{12} & E^1_{22} & E^1_{32} \end{bmatrix}; \mathbf{E}^2_{3 \times 3} = \begin{bmatrix} E^2_{11} & E^2_{21} & E^2_{31} \\ E^2_{12} & E^2_{22} & E^2_{32} \\ E^2_{13} & E^2_{23} & E^2_{33} \end{bmatrix}; \mathbf{E}^3_{3 \times 2} = \begin{bmatrix} E^3_{11} & E^3_{21} \\ E^3_{12} & E^3_{22} \\ E^3_{13} & E^3_{23} \end{bmatrix}$$

Table 8.2.1 summarizes the matrix equations for ANN models with 2, 3..., m neuron layers.

Table 8.2.1: General matrix equations for ANN models with 2, 3..., m neuron layers.

Number of neuron layers	Matrix equations
2	$E_{OI} = E_{nI \times n(1)}^1 E_{n(1) \times nO}^2$
3	$E_{OI} = E_{nI \times n(1)}^1 E_{n(1) \times n(2)}^2 E_{n(2) \times nO}^3$
M	$E_{OI} = E_{nI \times n(1)}^1 \dots E_{n(m-2) \times n(m-1)}^{m-1} E_{n(m-1) \times nO}^m$

In Table 8.2.1, ***nI*** is the number of inputs, ***nO*** is the number of outputs, ***m*** is the number of layers; and *n*(1), *n*(2),...*n*(*m*-1) are the number of neurons in the hidden layers 1,2,..., *m*-1, of the ANN model topology; ***k*** is the neuron number in the third hidden layer; ***i*** (*m* -1) and ***i*** (*m* -2) are the neuron numbers in the *m*-1 and *m*-2 layer, respectively. An useful algorithm to apply the shortcut procedure for the method of transfer functions in order to estimate the cause and effect relationships in ANN process models is shown in Figure 8.2.11.

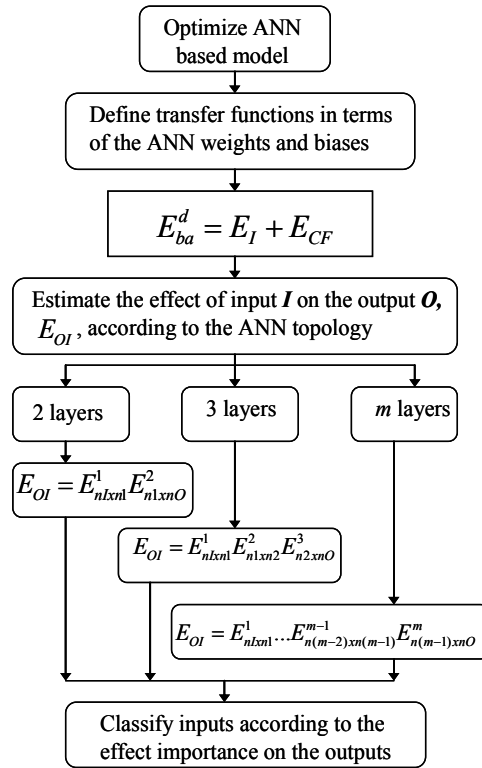


Figure 8.2.11. Algorithm to apply the shortcut procedure for the method of transfer functions.

8.2.3. NUMERICAL APLICATIONS

8.2.3.1. PARTICULAR CASE: Relative contributions expressed in terms of absolute values of the weights and biases as transfer functions

Figure 8.2.12 shows the notation used in this work for weights on ANN structures.

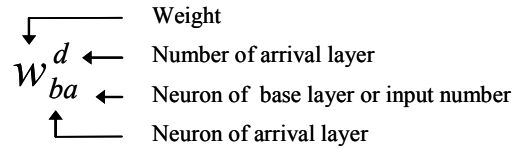


Figure 8.2.12. Notation used for ANN weights.

In this case the weight and biases associated with each connection between neurons of adjacent layers were used as the transfer functions. As was shown in Figure 8.2.7: $E_{ba}^d = E_I + E_{CF}$

Where the transfer function E_I is the causal relationship modifier and E_{CF} is the correction factor.

In this work, after testing the use of different mathematical statements by simulation, the following equations for each component of the transfer function E_{ba}^d are applied.

$$E_I = \frac{|w_{ba}^d|}{\sum_{i=1}^{na} |w_{bi}^d|} \quad \text{and} \quad E_{CF} = \frac{1}{na} \left[\frac{|b_b^d|}{\sum_{j=1}^{nb} (|b_j^d|)^\beta} \right]$$

Where:

d is the number of the arrival layer, b is a neuron in the arrival layer, a is a neuron in the base layer or an input number, na is the number of neurons in the base layer or of inputs, $|w_{ba}^d|$ is the absolute value of the connection weight, and $\left| \sum_{a=1}^{na} w_{ba}^d \right|$ is the sum of the weights that correspond to the connections between the neuron b of the arrival layer and each of the na neurons in the base layer, nb is the number of

neurons in the arrival layer, $|b_b^d|$ is the bias absolute value of the neuron b in the arrival layer d , $\sum_{b=1}^{nb} |b_b^d|$ is the sum of the absolute values of the neuron biases in the arrival layer. $\beta(d) = 1$ if $nb > 1$, $\beta(d) = 0$ if there is only one neuron in the arrival layer d ($nb = 1$).

Then, the transfer function E_{ba}^d that expresses the strength of the cause and effect relationship between neurons is represented by the following equation:

$$E_{ba}^d = \frac{|w_{ba}^d|}{\sum_{i=1}^{na} |w_{bi}^d|} + \frac{1}{na} \left[\frac{|b_b^d|}{\sum_{j=1}^{nb} (|b_j^d|)^{\beta(d)}} \right] \quad (8.2.16)$$

8.2.3.1.1. Case 1: Neural network with 5: 2: 1 topology, with five inputs, two hidden neurons and one output neuron

In the present example, in order to test the ability of the proposed method to determine the order of importance of the influence of the inputs on the outputs, a neural network was trained to describe the process of Isar *et al.* (2006): succinic acid production using *Escherichia Coli* under anaerobic conditions. The data set contains 32 experiments with five input factors and one output. The connection weights and biases of the neural network are shown in Table 8.2.2. In the discussion of the results for this example the effect of the ANN bias values in the estimation of the importance of the input variables is also discussed.

Table 8.2.2. Case 1: Weights associated to each connection between neurons

Weights (w_{ji})	Hidden layer Arrival layer 1		Weights (w_{oi})	Output layer Arrival layer 2
	Neuron 1	Neuron 2		
Inputs	j = 1	J = 2	Base layer	O = 1
I = 1	4.2452	0.2094	j = 1	0.0436
I = 2	3.2627	0.0638	j = 2	-1.4816
I = 3	14.7474	0.2515		
I = 4	9.5223	0.0907		
I = 5	0.9688	-0.2702		
Bias (b_j^1)	-12.9209	0.0049	Bias (b_o^2)	-0.0022

From the ANN topology, 5: 2 : 1, the matrix equation that represents the effects of the inputs on the outputs is: $E_{OI} = E_{5 \times 2}^1 E_{2 \times 1}^2$

$$\text{Where: } E_{5 \times 2}^1 = \begin{bmatrix} E_{11}^1 & E_{21}^1 \\ E_{12}^1 & E_{22}^1 \\ E_{13}^1 & E_{23}^1 \\ E_{14}^1 & E_{24}^1 \\ E_{15}^1 & E_{25}^1 \end{bmatrix} \text{ and } E_{2 \times 1}^2 = \begin{bmatrix} E_{11}^2 \\ E_{12}^2 \end{bmatrix}$$

The value of each transfer function, E_{ba}^d , in the matrices is obtained by solving the general Eq. (8.2.16) :

$$E_{ba}^d = \frac{|w_{ba}^d|}{\sum_{i=1}^{na} |w_{bi}^d|} + \frac{1}{na} \left[\frac{|b_b^d|}{\sum_{j=1}^{nb} (|b_j^d|)^{\beta(d)}} \right]$$

The values of the ANN weights and biases are obtained from Table 8.2.2. Finally, the effects of the inputs on the output, E_{OI} , are obtained by performing the substitution of numerical values and the matrix multiplication:

$$E_{5 \times 2}^1 = \begin{bmatrix} \frac{4.2452}{32.7464} + \frac{1}{5} \left(\frac{12.9209}{12.9258} \right) & \frac{0.2094}{0.8856} + \frac{1}{5} \left(\frac{0.0049}{12.9258} \right) \\ \frac{3.2627}{32.7464} + \frac{1}{5} \left(\frac{12.9209}{12.9258} \right) & \frac{0.0638}{0.8856} + \frac{1}{5} \left(\frac{0.0049}{12.9258} \right) \\ \frac{14.7474}{32.7464} + \frac{1}{5} \left(\frac{12.9209}{12.9258} \right) & \frac{0.2515}{0.8856} + \frac{1}{5} \left(\frac{0.0049}{12.9258} \right) \\ \frac{9.5223}{32.7464} + \frac{1}{5} \left(\frac{12.9209}{12.9258} \right) & \frac{0.0907}{0.8856} + \frac{1}{5} \left(\frac{0.0049}{12.9258} \right) \\ \frac{0.9688}{32.7464} + \frac{1}{5} \left(\frac{12.9209}{12.9258} \right) & \frac{0.2702}{0.8856} + \frac{1}{5} \left(\frac{0.0049}{12.9258} \right) \end{bmatrix}$$

$$E_{2 \times 1}^2 = \begin{bmatrix} \frac{0.0436}{1.5252} + \frac{1}{2} (0.0022) \\ \frac{1.4816}{1.5252} + \frac{1}{2} (0.0022) \end{bmatrix}, \text{ and}$$

$$E_{OI} = \begin{bmatrix} 0.3296 & 0.2365 \\ 0.2996 & 0.0721 \\ 0.6503 & 0.2841 \\ 0.4907 & 0.1025 \\ 0.2295 & 0.3052 \end{bmatrix} \begin{bmatrix} 0.0297 \\ 0.9725 \end{bmatrix} = \begin{bmatrix} 0.2398 \\ 0.0790 \\ 0.2956 \\ 0.1143 \\ 0.3036 \end{bmatrix}$$

Table 8.2.3 shows the order of importance of the effects of all the inputs on the output using the shortcut procedure of the method of transfer functions, the Garson method (GARSON, 1991) and the method that uses raw connection weight values (OLDEN *et al.*, 2004). The reference order was established using the coefficients of the mathematical model developed using experimental design methodology by Isar *et al.* (2006).

Table 8.2.3. Case 1: Comparison of the importance of the inputs' effects using different methods.

Input	Reference order ^(a)		Transfer function method.		Garson method ^(b)		Raw connection weights ^(c)	
	Model coefficients ^(a)	Order	Importance	Order	Importance	Order	Importance	Order
Input 1	-0.93	3	0.2398	3	0.2326	3	-0.13	4
Input 2	-0.32	5	0.0790	5	0.0730	5	0.05	5
Input 3	-1.31	2	0.2956	2	0.2900	2	0.27	3
Input 4	-0.35	4	0.1143	4	0.1092	4	0.28	2
Input 5	1.38	1	0.3036	1	0.2951	1	0.44	1

^(a) Coefficients of the mathematical model obtained using experimental design methodology, expressed in terms of coded factors (Isar *et al.*, 2006). ^(b) Garson, 1991. ^(c) Olden *et al.*, 2004.

Table 8.2.3 shows that the method of transfer functions and the Garson method predicted the same order of importance of the inputs as the reference mathematical model (ISAR *et al.*, 2006). The 'Perturb' method (GEVREY *et al.*, 2003; KAUR *et al.*, 2006) was also tested by changing 10% of each selected input while maintaining the others unchanged and the same order of importance of the inputs of the reference was obtained.

The use of the raw connection weight values proposed by Olden *et al.* (2004) predicted the order of importance of the inputs erroneously. Two factors that probably contributed to the wrong result were the big difference in magnitude

between the highest and the lowest weight values in Table 8.2.2 and the great dispersion of the weights values around the mean (variance = 23.33).

The method of transfer functions uses a correction factor expressed in terms of the bias, which is important in the determination of the importance of the inputs. Figure 8.2.13 shows the input (I) effects on the output when different values of biases (b) are used in the correction factor for the output neuron. It can be seen that when values of $b = 0.5$ or $b = 1$ are used as bias values in the output layer, the order of importance of the input effects on the output is different from the one obtained when smaller values are used. The influence of the biases is an important factor that is not considered in the Garson method, in the method that uses the raw connection weight values and in the methods based on weights that are mentioned in the literature cited. Therefore, the use of relative contributions expressed in terms of absolute values of weights and biases, in transfer functions, as proposed in this work, led to good results in this case.

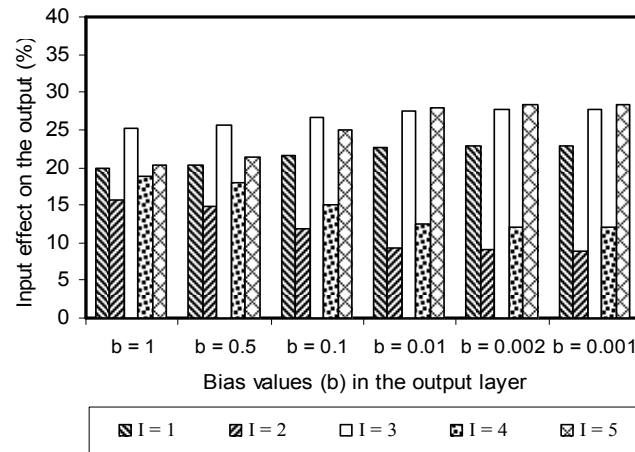


Figure 8.2.13. Influence of the bias value on the estimation of the importance of the input (I) effects on the output.

8.2.3.1.2. Case 2: Neural network with 13:7:1 topology, to model a continuous extractive fermentation process for bioethanol production from sugarcane molasses.

The second case is the study of the effects of 13 input factors on two output responses of an extractive fermentation process for bioethanol production.

The data and process information were obtained from previous works of Atala & Maciel (2006) and Boza-Condorena *et al.* (2011a and b).

The continuous extractive fermentation process produces ethanol from sugarcane molasses. In this process the yeast cells (*Saccharomyces cerevisiae*) are inside the fermentor, the medium (sugarcane molasses) is added continuously and the fermentation product (ethanol) is also continuously removed at a rate equal to the feed flow rate, so that the volume is held constant and there is a continuous flow of product.

A schematic diagram of the fermentation pilot plant is presented in Figure 8.2.14.

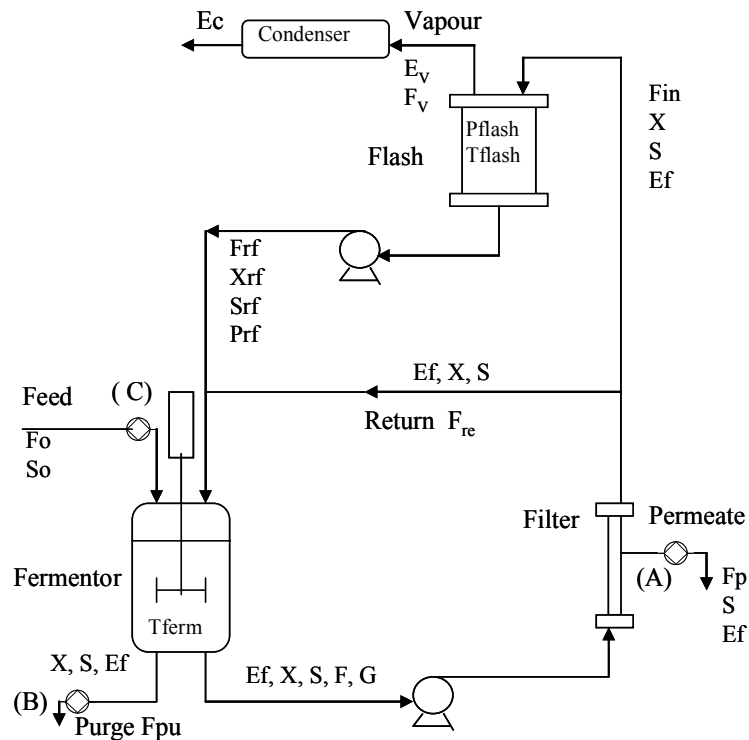


Figure 8.2.14. Schematic diagram of the fermentation pilot plant

The pilot plant operation can be briefly explained as follows:

- 1) At the beginning of the continuous extractive fermentation the permeate pump (A) of the filtration system was turned on. The flow of this pump was adjusted to remove permeate and to maintain different dilution rates: 0.03 h^{-1} (33.33 h residence time), 0.10 h^{-1} (10 h residence time), 0.15 h^{-1} (6.67 h residence time), 0.20 h^{-1} (5 h residence time) and 0.35 h^{-1} (2.85 h residence time).
- 2) When the cell concentration in the fermentor reached a value of 10% v/v, the purge bomb (B) was activated without changes in the residence time of the process.
- 3) The removal of fluid with the permeate and the purge pumps decreased the fermentation broth level into the fermentor activating the feeding pump (C) (connected to an on-off level controller) that supply fresh medium (sugarcane molasses), by this action the level was kept constant throughout the fermentation.
- 4) The temperature in the fermentor was maintained constant at $33.25 \pm 0.25 \text{ }^{\circ}\text{C}$, throughout the fermentation process.
- 5) The substrate concentration (total reducing sugars) in the feed flow to the fermentor was also maintained constant at 180 g/L throughout the fermentation process.
- 6) The flash tank was used to separate ethanol from the fermentation broth and was operated with a feed flow rate of 200 L/h, a vacuum pressure at $150 \pm 40 \text{ mmHg}$, and a temperature at $33.8 \pm 0.4 \text{ }^{\circ}\text{C}$.
- 7) The liquid remaining in the flash tank containing the fermentation broth with lower concentration of ethanol was returned to fermentor using a helical pump.

Therefore, from the plant operation description, the permeate pump (A), the purge pump (B) and the feeding pump (C) were very important in this process, they operated continuously and were responsible for maintaining the dilution rate or residence time of the process constant and for supplying a continuous feed flow to the fermentor. Additionally the purge pump was used to keep the fermentor cell concentration constant.

The input variables considered to develop the plant model based on ANN were the following : 1) Permeate flow rate (F_p), 2) Fermentor purge flow rate (F_{pu}), 3) Feed flow rate to the fermentor (F_0) , 4) Fermentor biomass concentration (X), 5) Fermentor substrate concentration (S), 6) Residence time (Tr), 7) Fermentor glycerol concentration (G), 8) Cell viability (C_v), 9) Fermentor outlet flow rate (F), 10) Fermentor temperature (T_{ferm}), 11) Flash tank inlet flow rate (F_{in}), 12) Flash tank temperature (T_{flash}), 13) Flash tank pressure (P_{flash}). The output variable was the fermentor ethanol concentration (E_f).

An artificial neural network with 13:7:1 topology was trained to model the process using experimental data from 400 h of plant operation. 75% of the input/output patterns were used in training and 25% of input/output patterns were used in the validation stage. The weight and biases are showed in Table 8.2.4.

Table 8.2.4. Case 2: Weights associated to each connection between neurons

Weights (w_{ji})	Hidden layer							Weights (w_{Oj})	Output Layer Arrival layer 2
	Arrival layer 1								
	Neuron 1	Neuron 2	Neuron 3	Neuron 4	Neuron 5	Neuron 6	Neuron 7		
Inputs	j =1	j =2	j =3	J =4	j =5	j = 6	J = 7	Source	Neuron 1
$I = 1$	-0.0913	2.0547	-3.9030	-0.0481	-0.0736	-3.2635	-0.0001	J = 1	-1.9439
$I = 2$	0.7877	-0.7963	6.1231	1.9009	0.5390	-24.2780	3.1185	j = 2	-0.9963
$I = 3$	8.9926	2.2582	3.4959	1.3161	6.1685	-17.0353	2.3592	j = 3	-0.2579
$I = 4$	-6.5591	10.1845	-1.1485	-1.2029	-4.4437	8.3930	-2.2318	j = 4	11.6378
$I = 5$	-0.0070	-0.0220	0.0438	0.0044	-0.0022	-0.0394	0.0041	j = 5	2.6585
$I = 6$	-0.0694	0.0167	0.0334	-0.0076	-0.0280	0.2861	-0.0082	j = 6	-0.1424
$I = 7$	-0.0587	0.1831	-0.7368	-0.1219	-0.0538	-0.2952	-0.1166	j = 7	-9.0520
$I = 8$	0.1794	0.1211	-0.0485	-0.0282	0.1341	-0.9811	-0.0248		
$I = 9$	1.2092	-1.4351	4.6711	0.4832	1.3804	34.7078	0.4723		
$I = 10$	-2.2002	-1.9251	2.6039	-0.7884	-1.7662	-13.6756	0.2015		
$I = 11$	0.4770	-0.9505	0.7165	0.5692	0.0869	28.1482	-0.4089		
$I = 12$	-2.6314	-5.0165	-2.0039	1.0058	-2.3469	-2.9263	1.2292		
$I = 13$	-0.5250	-0.6712	3.4306	-0.1294	-0.5487	3.0585	-2.3148		
Bias (b_j^1)	-1.9036	1.5279	6.9926	3.8386	-1.5669	-4.5269	2.9000	Bias (b_O^2)	-1.9316

The performance of the trained neural network was evaluated by its ability to predict the elements of the validation data set, these elements are experimental target values that were unknown by the neural network during training. The results

are shown in Figure 8.2.15, and show that the ANN model has a very good generalization capacity.

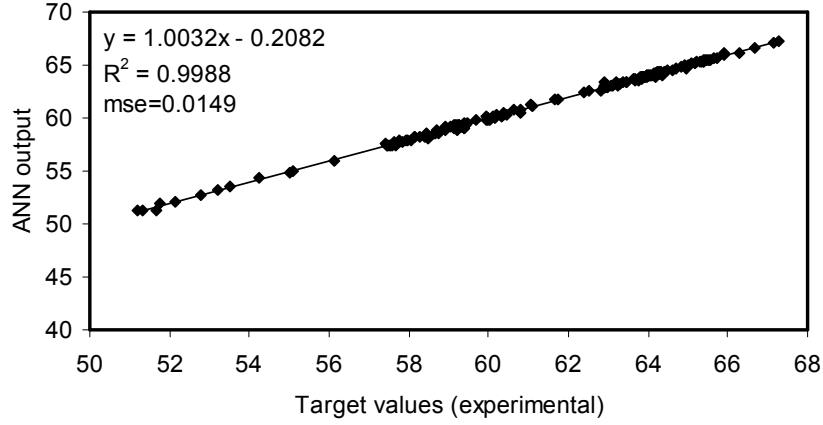


Figure 8.2.15 Experimental target values of the validation data set and network outputs.

From the ANN topology, 13: 7: 1, the matrix equation that represents the effect of the inputs on the outputs is: $E_{OI} = E_{13 \times 7}^1 E_{7 \times 1}^2$

$$\text{Where: } E_{13 \times 7}^1 = \begin{bmatrix} E_{11}^1 & E_{21}^1 & \dots & E_{71}^1 \\ E_{12}^1 & E_{22}^1 & \dots & E_{72}^1 \\ \vdots & \vdots & \ddots & \vdots \\ E_{1,13}^1 & E_{2,13}^1 & \dots & E_{7,13}^1 \end{bmatrix} \text{ and } E_{7 \times 1}^2 = \begin{bmatrix} E_{11}^2 \\ E_{12}^2 \\ \vdots \\ E_{17}^2 \end{bmatrix}$$

As in example 1, each transfer function E_{ba}^d , is obtained by using the ANN weights and biases, in this case from Table 5 to solve the general Eq. (8.2.16):

$$E_{ba}^d = \frac{|w_{ba}^d|}{\sum_{i=1}^{na} |w_{bi}^d|} + \frac{1}{na} \left[\frac{|b_b^d|}{\sum_{j=1}^{nb} (|b_j^d|)^{\beta(d)}} \right]$$

Figure 8.2.16 shows the order of importance of the effects of all the inputs on the output fermentor ethanol concentration (E_f).

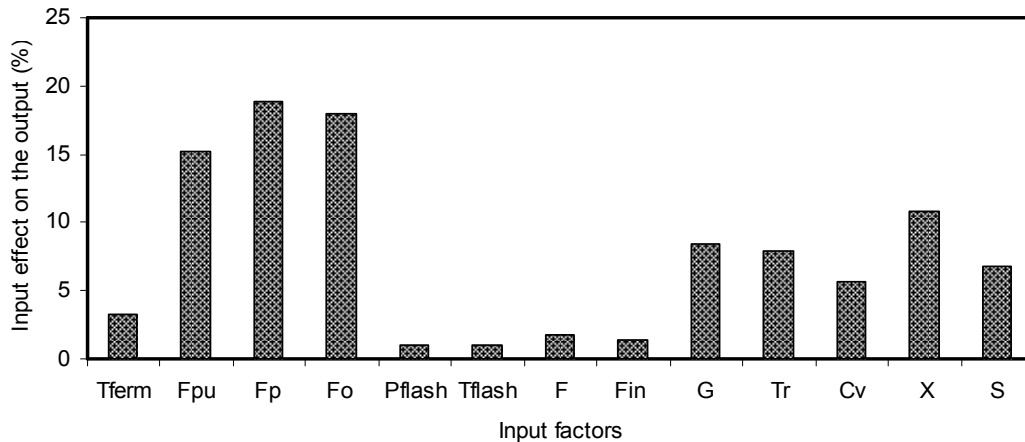


Figure 8.2.16. Input effects on the output (E_f) with the method of transfer functions

Figure 8.2.16 shows that the three input factors with the highest effect on the output are: permeate flow rate F_p , feed flow rate to the fermentor F_o , and purge flow rate F_{pu} .

The input factors with the lowest effect on the output are fermentor temperature, T_{ferm} , flash tank temperature, T_{flash} , feed flow rate to the flash tank, F_{in} , flash tank pressure, P_{flash} , and fermentor outlet flowrate F . The results can be explained because the input variables: T_{ferm} , T_{flash} , F_{in} and P_{flash} were maintained almost constant during all the experiment. This accounts for the low value of their effects on the output. In PART 2 of this chapter are showed results of additional studies performed with P_{flash} and F , to verify their effects using procedures of signal filtering.

Therefore Figure 8.2.16 shows that the information obtained with the use of Eq. (8.2.16) fits well with the plant dynamics and with the dynamic behavior of the process variables during the plant operation. The permeate flow rate, the feed flow rate and the purge flow rate had important roles in the operation of the plant and

the group of variables whose values were kept almost constant during the experiment show the lowest effect values.

8.2.3.1.2.1. Analyzing the results using Principal Component Analysis

In order to confirm the results obtained by applying the method of transfer functions on the neural network model of the process, an effective procedure for performing an analysis of the input vector was used, the principal component analysis. This technique has three effects: 1) it orthogonalizes the components of the input vectors (so that they are uncorrelated with each other), 2) it orders the resulting orthogonal components (principal components) so that those with the largest variation come first, and 3) it eliminates those components that contribute the least to the variation in the data set, therefore it builds an importance order based on the information of the variances.

Principal components are transformed vectors which express contributions to the total variation in the input data set in terms of percentage.

In this case five principal components were extracted from the data set of thirteen input variables. Table 8.2.5 shows the contributions of the principal components Cp1, Cp2, Cp3, Cp4 and Cp5 to the total variance in the component space.

The principal component Cp1 is the most important and explains 51.48 % of the variation in the data set. The total variation explained by the five principal components is 93.14%.

Table 8.2.5. Principal Component contributions to the total variation in the input data set.

Principal Component	Contributions (%)
Cp1	51.48
Cp2	24.72
Cp3	9.50
Cp4	4.45
Cp5	2.98
Total variation explained (%)	93.14

Table 8.2.5 shows that Cp1 and Cp2 explain more than 75% of the variation in the input data set, therefore correlation coefficients of the inputs variables with these principal components could be useful to inform about the importance of each variable to explain the process behaviour.

Figure 8.2.17 is a graphic representation of the component matrix. It shows the value of the correlation coefficients r between each variable and the principal components Cp1 and Cp2.

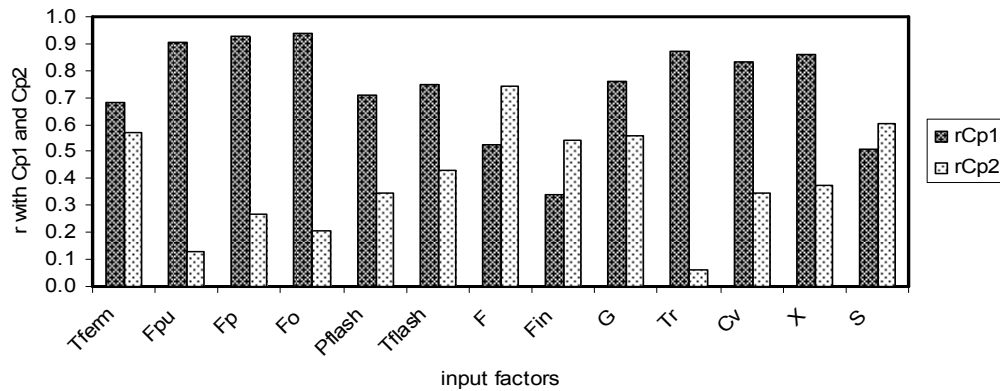


Figure 8.2.17. Correlation coefficients r between each input factor and the principal components Cp1 and Cp2.

Figure 8.2.17 shows that the following three variables have the highest correlation coefficients r with the principal components Cp1: permeate flow rate F_p , feed flow rate to the fermentor F_0 , and fermentor purge flow rate F_{pu} , therefore they could be said to have the more important effect on the output. This is the same result that was obtained by using the transfer function method with the ANN model of the fermentation process. The fermentor biomass concentration X and residence time Tr have also high association with Cp1 but its r coefficient is lower.

In Figure 8.2.17 T_{ferm} and T_{flash} have a value of correlation coefficient of $r = 0.68$ and $r = 0.75$, respectively, with the principal component Cp1, however the variation coefficient of these variables in the data base are 0.52%, and 0.56%

respectively; therefore it is estimated that values of r coefficient lower than 0.8 with the principal component Cp1 in this case identify variables with non significant contribution in the data set.

Figure 8.2.17 shows that Pflash and F have a lower r coefficient with the principal component Cp1, in both cases $r < 0.8$, F has a bigger r coefficient with the component Cp2 than with Cp1 but also with a lower value $r < 0.8$.

8.2.4. CONCLUSIONS

The transfer function method provided a general framework to study the information contained in the weights and biases of ANN based models. The application of block diagram algebra using transfer functions to express the cause and effect relationship between adjacent neurons in ANNs was useful to estimate the importance of the effects of inputs on outputs and facilitated the derivation of applicable equations for MIMO process models based on neural networks with two or more layers of neurons, extending the possibilities of analyzing these cases with respect to the weight based methods found in the literature.

In general the relative contribution of an input x to an output y , when compared to the relative contributions of the other inputs to the same output, expresses the relative effect of the input x on the output y , and hence its importance in the set of inputs.

In numerical applications, good results were obtained when using relative contributions expressed in terms of absolute values of weights and biases in transfer functions in order to estimate the effects of inputs on outputs.

A good neural network model of a process gives secure information about the relative importance of the input variables, which highlights the importance of the availability of good models to describe the dynamic behavior of chemical and biotechnological processes.

The approach used in this work has been applied with successful results to identification of variables in designing linear and non linear control structures based on Model Predictive Control (MPC) with ANN internal models for the process described in example 3 (BOZA-CONDORENA *et al.*, 2011; BOZA-CONDORENA & COSTA, 2011a; BOZA-CONDORENA & COSTA, 2011b). The proposed method could be a valuable tool to determine the importance of the causal effects of inputs on outputs and to analyze process plant dynamic behaviors.

Acknowledgments

The authors acknowledge CAPES and FAPESP for their financial support.

Nomenclature

ANN	artificial neural network
G	transfer function
x	input
y	output
P	pickoff point
S	summing point
E_{ij}	Effect of input i on the output j
w	weight in an artificial neural network
B	bias in an artificial neural network
mse	mean square error
N	number of data points in the validation data set
k	the k th <i>target value</i> of the data point in the validation data set
t_k	the k th <i>target value</i>
ev_k^O	the k th <i>net estimated output value</i>

8.2.5. LITERATURE CITED

1. Atala, D.I.P., & Maciel, R. (2006). **Experimental development, optimization and control of a Continuous extractive production process for the ethanol production**. Postdoctoral research. Process 151335/2005-2, FAPESP, Brazil.
2. Boza-Condorena, E.G., Atala, D.I.P. & Costa, A.C.(2011). Dynamic Matrix Control with Internal ModelBased on ANN of a Continuous Extractive Process for Bioethanol Production. In *Industrial Control Systems*. Robert C. Gilbert (Ed). Nova Science Publishers, Inc. Hauppauge, NY, USA.
3. Boza-Condorena, E.G. and Costa, A.C. (2011a). Non-linear predictive control of a fermentor in a continuous reaction-separation process. *Lecture Notes in Engineering and Computer Science. Proceedings of the World Congress on Engineering and Computer Science 2011*. Vol. II, pp.663 - 668. San Francisco, CA. USA.
4. Boza-Condorena, E.G. and Costa, A.C. (2011b). A method of transfer functions and block diagram to study the contribution of variables in artificial neural network process models. *Lecture Notes in Engineering and Computer Science. Proceedings of the World Congress on Engineering and Computer Science 2011*. Vol.II. pp.708-713. San Francisco, CA. USA.
5. Chapados, N., & Bengio, Y. (2001). Input decay: Simple and effective soft variable selection. In *Proceedings of the 2001 IEEE/INNS International Joint Conference on Neural Networks* (pp.1233 – 1237).
6. Chen, J. CH., & Chen, W. H. (2008). Diagnostic analysis of a small-scale incinerator by the Garson index. *Inform Sci*, 178 (23), 4560 - 4570.
7. Engelbrecht, A., & Cloete, I. (1998). Feature extraction from feedforward neural networks using sensitivity analysis. In V.B. Bajic (Ed.), *Proceedings of the International Conference on Systems, Signals, Control, Computers* (pp. 221-225). Durban, South Africa: IAAMSAD and the South African branch of the Academy of Nonlinear Sciences.
8. Ghai, V.V. (2006). **Knowledge based approach using neural networks for predicting corrosion rate**. MSc Thesis, College of Engineering and Technology, Ohio University, USA.
9. Garson, G.D. (1991). Interpreting neural-network connection weights. *AI Expert*, 6 (7), 47-51.

10. Gevrey, M., Dimopoulos, I., & Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural networks models. *Ecol Modell*, 160, 249-264.
11. Guimarães, O.L., Reis, M.H., Villela, D.N., Siquiera, A.F., Izário, H.J., Aquino, H.O., *et al.* (2008). Discoloration process modeling by neural network. *Chem Eng J*, 140, 71- 76.
12. Hagiwara, M. (1994). Simple and effective method for removal of hidden units and weights, *Neurocomputing*, 6 (2), 207–218.
13. Huang, Z., Chen, H., Hsu, C.J., Chen, W.H., & Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decis Support Syst*, 37(4), 543–558.
14. Isar, J., Lata, A., Saurabh, S., & Rajendra, S.A. (2006). Statistical method for enhancing the production of succinic acid from *Escherichia coli* under anaerobic conditions. *Bioresour Technol*, 97, 1443–1448.
15. Kaur, K., Kaur, A., & Malhotra, R. (2006). Alternative methods to rank the impact of object oriented metrics in fault prediction modeling using neural networks. *World Acad Sci Eng Tech*, 19, 143-148.
16. Luyben, W. (1990). **Process modeling, simulation and control for chemical engineers**, (2nd ed.). New York: McGraw-Hill.
17. Mei, C. (2002). On Teaching the Simplification of Block Diagrams. *Int J Eng Ed*, 18 (6), 697-703.
18. Olden, J.D., Joy, M.K., & Death, R.G. (2004). An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecol Modell*, 178, 389-397.
19. Oliveira-Esquerre, K.P., Mori, M., & Bruns, R.E. (2002). Simulation of an industrial wastewater treatment plant using artificial neural networks and principal components analysis. *Braz J Chem Eng*, 19(4), 365-370.
20. Reitermanová, Z. (2008). *Feedforward Neural Networks – Architecture Optimization and Knowledge Extraction*. In J. Safrankova & J. Pavlu (Eds.), WDS'08 Proceedings of Contributed Papers, *Part I: Mathematics and Computer Sciences* (pp. 159–164). Prague, Czech Republic: Matfyzpress.
21. Seborg, D.E., Edgar, T.F., & Mellichamp, D.A. (2003). **Process Dynamics and Control**, (2nd ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

22. Sharma, R., Singh, K., Diwakar, S., & Ghosh, R. (2004). Neural network applications for detecting process faults in packed towers. *Chem Eng Process*, 43 (7), 841-847.
23. Shi, Y., Gai, G.S., Zhao, X.T., Hu, Y.Y. (2009). Influence and Simulation Model of Operational Parameters on Hydrogen Bio-production Through Anaerobic Microorganism Fermentation Using Two Kinds of Wastes. In S.I. Ao, C. Douglas, W.S. Grundfest, J. Burgstone (Eds.), *Proceedings of the World Congress on Engineering and Computer Science 2009*. Vol.II, pp. 938-942. San Francisco, CA: Newswood Limited.
24. Similä, T., & Tikka, J. (2009). Combined input variable selection and model complexity control for nonlinear regression. *Pattern Recognit Lett*, 30(3), 231-236.
25. Tikka, J. (2009). ***Simultaneous input variable and basis function selection for RBF networks***. *Neurocomputing*, 72(10-12), 2649-2658.
26. Taha, I.A., & Ghosh, J. (1999). Symbolic interpretation of artificial neural networks. *IEEE Transactions on Knowledge and Data Engineering*, 11 (3), 448-463.

8.3. PART 3: USE OF THE TRANSFER FUNCTION METHOD TO REDUCE THE NUMBER OF INPUTS IN ANN PLANT MODELS FOR EXTRACTIVE FERMENTATION WITH FLASH UNIT

8.3.1. INTRODUCTION

The modeling process is a complex task in the development of advanced computer-based applications. According to Seborg (2004) it is important to remember that a process model is nothing more than a mathematical abstraction the equations are at best an approximation to the real process. In modeling it is important to obtain a solution taking into account complexity, accuracy, cost an effort needed to develop the model. In the present case as was explained before there is no available a theoretical model because the complex properties of the fermentation broth that gives difficulty to the prediction of the vapor – liquid equilibrium in the flash separation unit, then in the present research are used for some variables the experimental data stored by the data acquisition system and for the other the data from experimental determinations at laboratory obtained by sampling the process.

Fonseca (1998), mentions that the networks often yield satisfactory results after training but do not perform well in control applications. Therefore it is necessary not only to train empirical models but also to test the predictive control technology with different empirical models of the process in order to asses its performance in control applications.

The purpose of this section is to select two models with different topology and training procedure to be used in the next chapter as process plant models, these models in the controlled system receive control actions from the controller, which has an internal plant model, both in the present research are ANN based models, therefore the models from this section will be used with the models of the next chapter to test the performance of the controlled systems.

This section of the article is organized as follows:

- 1) Two alternatives for modeling using different number of variables are identified in both the output variable is the fermentor ethanol concentration and the second one is a model with reduction in the number of input variables.
- 2) The procedures that were used to reduce the number of inputs in modeling are explained.
- 3) The results of training the ANN plant models for the two alternatives before mentioned are presented.
- 4) The possible manipulated variables are analyzed.
- 5) Conclusions are given at the end of the section.

8.3.2. ALTERNATIVES FOR PLANT MODELING USING DIFFERENT NUMBER OF VARIABLES

A) ALTERNATIVE I.- The plant model based on ANN (M13) was the same that was analyzed in Part 2 (item 8.2.3.1.2) without reduction in the number of input variables. In this case the input variables of the model were the following:

1. Permeate flow rate (F_P)
2. Fermentor purge flow rate (F_{PU})
3. Feed flow rate to the fermentor (F_0)
4. Fermentor biomass concentration (X)
5. Fermentor substrate concentration (S)
6. Residence time (Tr)
7. Fermentor Glycerol concentration (G)
8. Cell viability (Cv)
9. Fermentor outlet flow rate (F)
10. Fermentor temperature (T_{ferm})
11. Flash tank inlet flow rate (F_{in})
12. Flash tank temperature (T_{flash})
13. Flash tank pressure (P_{flash})

B) ALTERNATIVE II.- A plant model based on ANN (M8) with reduction in the number of input variables was used. In this case the input variables to develop the model were the following:

1. Permeate flow rate (F_P)
2. Fermentor purge flow rate (F_{PU})
3. Feed flow rate to the fermentor (F_0)
4. Fermentor biomass concentration (X)
5. Fermentor substrate concentration (S)
6. Residence time (τ)
7. Fermentor outlet flow rate (F)
8. Flash tank pressure (P_{flash})

In the following items procedures that were used to reduce the number of inputs by estimating their importance in the input data set are explained.

8.3.3. USE OF VARIABLE CONTRIBUTIONS TO THE ANN OUTPUT TO IDENTIFY THE IMPORTANCE OF VARIABLES

To develop an empirical model, the first step before training the ANN is to define the number of input variables to be used, for doing this it is necessary to identify the importance of each input variable usually it is made by assessing the effect that the input has on the output behavior.

In the present chapter the method of transfer functions (BOZA and COSTA, 2011) was used in order to determine the importance of the effect of each input variable on the output variable. This method uses the interconnection weights of multilayer feedforward neural networks to extract information about the cause and effect relationship between input and output variables.

The input-output dependencies are expressed in terms of transfer functions, which are used to estimate the importance of the influence of the inputs on the outputs.

The procedure to determine the importance of the inputs by using the transfer function method was explained in Part 1 and 2 of this Chapter:

The results of using two models: 1) Model I without reduction in the number of inputs with neural network topology 13: 7: 1, with thirteen inputs, seven neurons in the hidden layer and one output and 2) Model II with reduction in the number of inputs with topology 8: 7: 1 with eighth inputs, seven neurons in the hidden layer and one output, are shown in Table 8.3.1.

Table 8.3.1. Input variable contribution to the output variable: ethanol concentration, E_f ; using the transfer function method.

Order	Variable	Importance	Importance
		Topology 13:7:1	Reduced Model Topology 8:7:1 (*)
1	Permeate flow rate (F_P)	0.1933	0.1439
2	Fermentor purge flow rate (F_{PU})	0.1825	0.1107
3	Feed flow rate to the fermentor (F_0)	0.1792	0.1379
4	Fermentor biomass concentration (X)	0.1149	0.1627
5	Fermentor substrate concentration (S)	0.0764	0.0765
6	Residence time (τ_r)	0.0727	0.1623
7	Fermentor glycerol concentration (G)	0.0616	--- -----
8	Cell viability (C_v)	0.0515	---
9	Fermentor outlet flow rate (F)	0.0186	0.0640 (**)
10	Fermentor temperature (T_{ferm})	0.0175	--- -----
11	Flash tank inlet flow rate (F_{in})	0.0125	--- ----
12	Flash tank temperature (T_{flash})	0.0100	--- -----
13	Flash tank pressure (P_{flash})	0.0094	0.1419 (**)

(*) Average value of three determinations. (**) Filtered data sets

The fermentor temperature (T_{ferm}), flash tank inlet flow rate (F_{in}), and flash tank temperature (T_{flash}), have small values in Table 8.3.1 because they were maintained almost constant during the operation time, therefore its coefficient of

variation (standard deviation/mean) were the smaller ones in the input data set. So, in a first step without these variables the number of variables were reduced from 13 to 10.

In a second step by using graphic representations of the data of the flash tank pressure (Pflash) and the fermentor outlet flow rate (F) it was verified that these variables present signal noise. In all ANN models for the process that were obtained without signal filtering and using different number of ANN parameter values, these variables showed to have minor importance in the input data set, however when signal filtering was used, their importance estimated by the same method grew up considerably. The ANNs models used have shown to be very sensitive to the signal noise, they gave less importance to variables with unpredictable variations in the measured signal from moment to moment, possibly this is because in these cases the true underlying trajectory is more difficult to be identified than in the case of filtered signals with clearly defined trajectories. Therefore Pflash and F, were considered for the plant model.

In a third step was taken the decision to not consider for the model the variables fermentor glycerol concentration (G) and cell viability (Cv) because of their importance values in Table 8.3.1 and because their variable values are obtained by off line analysis in much more time that the ones of the variables with values obtained on line, this fact makes necessary the additional use of an interpolation procedure in order to solve the difference.

The first six variables in Table 8.3.1 Fp, Fpu, Fo, X, S and tr are considered important variables to explain the process behavior according to their importance values and to the literature cited in references (ATALA, 2001; ATALA, 2004; CCOPA RIVERA, 2010; COSTA,2001).

In Table 8.3.1 we have: variables that cannot be manipulated or controlled in control loops because their values are obtained by off-line analysis: X, S, G and Cv; controlled variables with non significant variation in the operation time: Tferm, Fin, and Tflash, and possible manipulated variables: Fp, Fpu, Fo, Pflash, F and tr.

The first four manipulated variables have the higher influence on the variable to be controlled, the concentration of ethanol in the fermentor. From the results in Table 8.3.1 the selected inputs for a model with reduced number of inputs are the following: F_p , F_{pu} , F_o , X , S , tr , $Pflash$ and F . Therefore in this way the number of variables for the plant model was reduced from thirteen to eight variables

8.3.4. USE OF PRINCIPAL COMPONENT ANALYSIS TO IDENTIFY THE IMPORTANCE OF VARIABLES AND TO REDUCE THE NUMBER OF INPUTS

Principal component analysis was used in order to confirm the results obtained by using the transfer function method, as mentioned before principal component analysis is an effective procedure for performing an analysis of the input vector, it builds an importance order for the principal components based on the information of the variances and eliminates those components that contribute the least to the variation in the data set. Principal components are transformed vectors which are used to express different contributions to the total variation in the input data set in terms of percentage. The greater the contribution to the total variation, the more important the principal component.

In this case principal component analysis was applied to the input vector in different operation times: 50 - 400 h, 100 - 400 h, 150 - 400 h and 200 - 400 h; in order to reduce the dimension of the input vector. In each case five principal components were extracted from the data set of thirteen variables. Table 8.3.2 and Table 8.3.3 show the contributions of the principal components Cp_1 , Cp_2 , Cp_3 , Cp_4 and Cp_5 to the total variance in the component space, in the operation times of 50 - 400 h and 150 - 400 h respectively.

A) Case 1: Operation time 40 - 400 h

Two data sets were considered to show the effect of signal filtering in case of flash tank pressure ($Pflash$) and fermentor outlet flow rate (F): data set 1A with filtered signals, and data set 2A without filtered signals.

The principal component Cp1, the most important component, explains 53.27 % of the variation in the data set 1A and 51.48 % in the data set 2A. The first three principal components contain 88.28 % of the variation in the data set 1A and 85.7 % of the variation in the data set 2A respectively.

\
Table 8.3.2. Principal Component contributions to the total variation in the input data set : Operation time 40 - 400 h

Principal Component	Contributions (%) data set 1A	Contributions (%) Data set 2A
Cp1	53.27	51.48
Cp2	25.36	24.72
Cp3	9.65	9.50
Cp4	4.06	4.45
Cp5	2.61	2.98
Total variation explained (%)	94.95	93.14

B) Case 2 Operation time 150 – 400 h

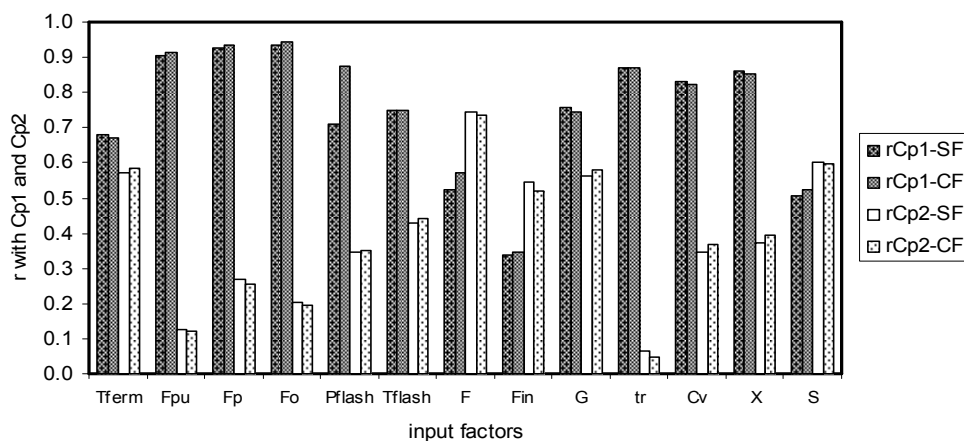
Two data sets were considered to show the effect of signal filtering in case of the flash tank pressure (Pflash) and the fermentor outlet flow rate (F): data set 1A with filtered signals and data set 2A without filtered signals.

The principal component Cp1, the most important component, explains 47.89 % of the variation in the data set 1A and 45.47 % in the data set 2A. The first 3 principal components contain 78.91% of the variation in the data set 1A and 76.22 % of the variation in the data set 2A respectively.

Table 8.3.3. Principal Component contributions to the total variation in the input data set : Operation time 150 - 400 h

Principal Component	Contributions (%) data set 1A	Contributions (%) Data set 2A
Cp1	47.89	45.47
Cp2	21.09	20.84
Cp3	9.93	9.91
Cp4	6.33	6.22
Cp5	5.99	6.05
Total variation explained (%)	91.23	88.49

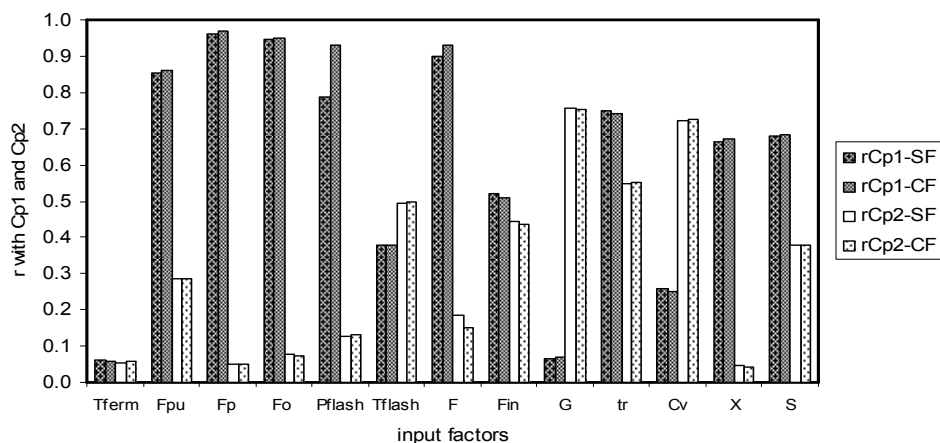
Figures 8.3.1 and 8.3.2 show the r correlation coefficients between each variable in the data set with each principal component and therefore describe that the variable which is more related to the more important principal component in the data set is more important than the variable which are less related with the same principal component.



CF: data set with filtered signals of Pflash and F.

SF: data set without filtered signals of Pflash and F.

Figure 8.3.1. Correlation coefficients r between each input factor and the principal components Cp1 and Cp2. Operation time: 50 – 400 h.



CF: data set with filtered signals of Pflash and F.

SF: data set without filtered signals of Pflash and F.

Figure 8.3.2. Correlation coefficients r between each input factor and the principal components Cp1 and Cp2. Operation time: 150 – 400 h.

The coefficients of variation (standard deviation/mean) of the variables: T_{ferm}, T_{flash} and F_{in}; in the data set are the smaller ones because they were maintained almost constant during the process, they are lower than 0.6%, therefore the values of r coefficients of these variables with the principal component Cp1 were used as a reference to interpret the results from Figures 8.3.1 and 8.3.2. In Figure 8.3.1 the r coefficient of T_{ferm}, T_{flash} and F_{in} are $r \leq 0.75$ and in Figure 8.3.2 for the same variables their values are $r \leq 0.6$, then less importance was attributed to the variables with those r values in the data set.

Using the value of $r \leq 0.75$ with the principal component Cp1 in Figure 8.3.1 may be inferred that the signal filtering was important in case of pressure in the flash tank (P_{flash}) to have better information of the importance of this variable in the process. P_{flash} without signal filtering have a value of $r = 0.71$ however with signal filtering the value obtained was $r = 0.87$, then it is an important variable in the data set.

The same criteria to interpret the value of r coefficient with the principal component Cp1 were used in a first step to select the variables for a process model with reduction in the number of inputs, in this first step were considered inputs variables with $r > 0.75$ in Figure 8.3.1 and variables with $r > 0.6$ in Figure 8.3.2. A second step was to consider that in a greater operation period the relationships between variables are similar to the ones showed in Figure 8.3.2. Therefore with these criteria the selected variables were the following: F_{pu}, F_p, F_o, P_{flash}, F, t_r , X and S; in this way was reduced the number of input variables from thirteen to eight variables. The same result was obtained by applying the transfer function method on ANN process models in item 8.3.2.

In the final model with reduced number of inputs was used the dilution coefficient $D = 1/t_r$ instead of t_r because D showed similar r coefficient with the principal component Cp1 than t_r in the data set of the operation time 50 – 400 h, however in the data set of the operation time 150 – 400h, the r coefficient of D was greater than that of t_r .

8.3.5. DESCRIPTION OF THE ANN PLANT MODELS

8.3.5.1. Alternative I

The plant's model of alternative I is identified as M13 in the next chapter to indicate an ANN model with 13 input variables the model topology is 13: 7 : 1. It has thirteen input variables and one output variable, the fermentor ethanol concentration (Ef) (see item 8.3.2).

According to Kwong (2005) and Morari, M., Lee, J., García (2002), the ANN model of the plant to be controled was trained with a sequence of inputs $\{u(0), u(T), u(2T), \dots, u(kT), \dots\}$ and a sequence of outputs $\{y(0), y(T), y(2T), \dots, y(kT), \dots\}$, which allows us to describe the output behavior during the time operation because the sequence of ouputs is the result of the sequence of inputs, the arguments 0, T, 2T, ..., kT, ... denote the given time as a multiple of the sampling interval T. The sampling interval was 30 minutes.

The selected neural network has 7 neurons in the hidden layer because this configuration presented the lowest values of *mse* in the validation test. Figure 8.3.3 shows the ANN selected model topology.

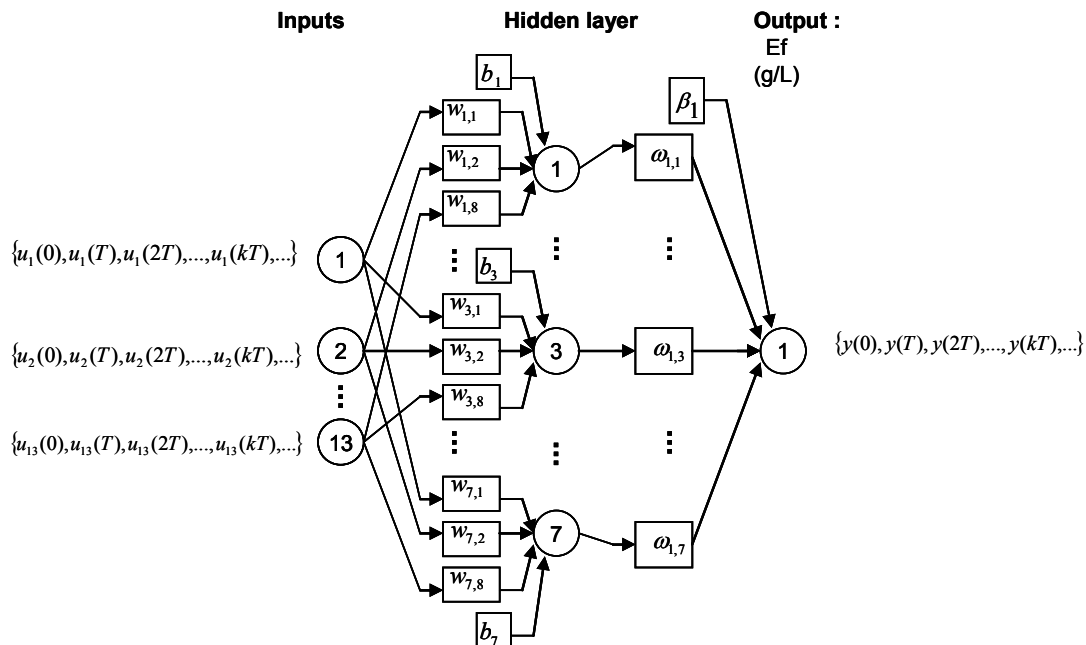


Figure 8.3.3. Alternative 1: ANN selected model topology.

Table 8.3.4 presents the mse and R^2 values obtained after the validation test with a different number of neurons in the ANN hidden layer.

Table 8.3.4. Alternative I: ANN model selection, mean squared error (mse) and R^2 values in the validation test

Neurons	Mse	R^2
4	0.76560	0.9935
5	0.04548	0.9962
6	0.03274	0.9973
7	0.01489	0.9988
8	0.05223	0.9978
9	0.06690	0.9945
10	0.02585	0.9979
11	0.05760	0.9953

Figure 8.3.4 shows the sequence of outputs experimental and predicted (E_f) by the ANN and that the nonlinear behavior of the ethanol concentration in the fermentor is modeled accurately by the selected ANN.

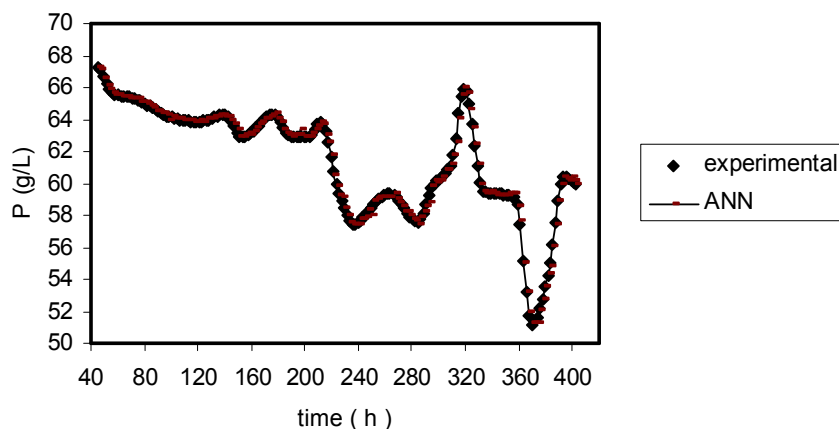


Figure 8.3.4. Experimental and ANN predicted values of the fermentor ethanol concentration in the pilot plant with extractive fermentation process.

8.3.5.2. Alternative II

The plant model based on ANN was obtained by reduction in the number of input variables from the model of the alternative I from thirteen to eight input

variables; the output variable is the fermentor ethanol concentration (E_f) (see item 8.3.2).

In the selection of the neural network model were used as criteria the neural network with low number of parameters, good performance and low number of training epochs. The better values of R^2 , mse , parameter number and number of epochs of training were considered. The values of R^2 and mse were obtained in the validation test. Table 8.3.5 presents the results.

According to Table 8.3.5 different alternatives of ANN plant model were trained, in models 1, 2 and 3 the sequence of outputs was $\{y(T), y(2T), y(3T), \dots, y((k+1)T), \dots\}$, and the sequence of inputs was $\{u(0), u(T), u(2T), \dots, u(kT), \dots\}$, the arguments 0, T, 2T, ..., kT, ... denote the given time as a multiple of the sampling interval T. The models give a predictive description of the output behavior at each sampling time, because, using the values of the inputs which were obtained in the current sampling time they gives values of outputs that correspond to the next sampling time. In Table 8.3.5 the sequence of inputs is represented using t_i , where $i = 0, 1, 2, 3, \dots, N$ (N is the last sampling time); instead of the sequence of times: 0, T, 2T, ..., NT.

In models 4, 5 and 6 were used at each sampling time past values of the inputs resulting the following sequence of outputs $\{y(2T), y(3T), y(4T), \dots, y((k+1)T), \dots\}$, and the following sequence of inputs: $\{[u(0), u(T)], [u(T), u(2T)], [u(2T), u(3T)], \dots, [u((k-1)T), u(kT)], \dots\}$. In this case the models give a predictive description of the output behavior at each sampling time, because, using the values of the inputs which were obtained in the current sampling time (i.e. $t_i = 2T$) and the past sampling time (i.e. $t_{i-1} = T$) they give the outputs that correspond to the next sampling time (i.e. $t_{i+1} = 3T$).

In models 7, 8, 9 the models also gives a predictive description of the output behavior at each sampling time, because using the values of the inputs which were obtained in the current and three past sampling times they give values of the outputs that correspond to the next sampling time.

In models 10 to 18 the output (controlled variable) is also used as input, i.e. in models 10, 11 and 12 the sequences of used inputs is: $\{y(0), y(T), y(2T), \dots, y((kT), \dots\}$ and $\{u(0), u(T), u(2T), \dots, u(kT), \dots\}$, while the sequence of outputs is $\{y(T), y(2T), y(3T), \dots, y((k+1)T), \dots\}$.

Table 8.3.5. Alternative 2: ANN selection, mean squared error (mse) and R^2 values in the validation test.

Model	ANN Topology	Parameter Number	Inputs (u)	Output (y)	R^2	Mse	Epochs
1	8:7:1	71	$u(t_i), \dots, u(t_{N-1})$	$Y(t_{i+1}), \dots, y(t_N)$	0.9987	0.0159	140
2	8:7:1	71			0.9987	0.0155	500
3	8:7:1	71			0.9984	0.0194	800
4	8:7:1	127	$u(t_{i-1}), u(t_i), \dots, u(t_{N-2}), u(t_{N-1})$	$Y(t_{i+1}), \dots, y(t_N)$	0.9958	0.0498	200
5	8:7:1	127			0.9973	0.0323	500
6	8:7:1	127			0.9984	0.0188	800
7	8:7:1	239	$u(t_{i-3}), u(t_{i-2}), u(t_{i-1}), u(t_i), \dots, u(t_{N-4}), u(t_{N-3}), u(t_{N-2}), u(t_{N-1})$	$y(t_{i+1}), \dots, y(t_N)$	0.9705	0.3659	200
8	8:7:1	239			0.9757	0.2893	500
9	8:7:1	239			0.9571	0.6099	800
10	9:5:1	56	$u(t_i), \dots, u(t_{N-1}); y(t_i), \dots, y(t_{N-1})$	$Y(t_{i+1}), \dots, y(t_N)$	0.99997	0.0003	200
11	9:5:1	56			0.99998	0.0003	500
12	9:5:1	56			0.99998	0.0003	800
13	9:5:1	101	$u(t_{i-1}), u(t_i), \dots, u(t_{N-2}), u(t_{N-1}); y(t_{i-1}), y(t_i), \dots, y(t_{N-2}), y(t_{N-1})$	$Y(t_{i+1}), \dots, y(t_N)$	0.9999	0.0012	200
14	9:5:1	101			0.9999	0.0014	500
15	9:5:1	101			0.9997	0.0040	800
16	9:5:1	191	$u(t_{i-3}), u(t_{i-2}), u(t_{i-1}), u(t_i), \dots, u(t_{N-4}), u(t_{N-3}), u(t_{N-2}), u(t_{N-1}); y(t_{i-3}), y(t_{i-2}), y(t_{i-1}), y(t_i), \dots, y(t_{N-4}), y(t_{N-3}), y(t_{N-2}), y(t_{N-1});$	$y(t_{i+1}), \dots, y(t_N)$	0.9988	0.0147	200
17	9:5:1	191			0.9982	0.0245	500
18	9:5:1	191			0.9974	0.0342	800

From Table 8.3.5 models 1, 2 and 3 with 8 input variables and models 10, 11 and 12 with 9 inputs using in this case the output as other input variable are the best alternatives. They present high values of R^2 , lower values of mse and parameter numbers. In the present study case the introduction of one delay time in the inputs increases the parameter number in near of 80% giving more complexity to the neural net without an increase in its performance.

The following schemes show the structure of the selected alternatives for ANN modeling.

1) Structure A: models 1, 2, 3 and 4 with 8 input variables.

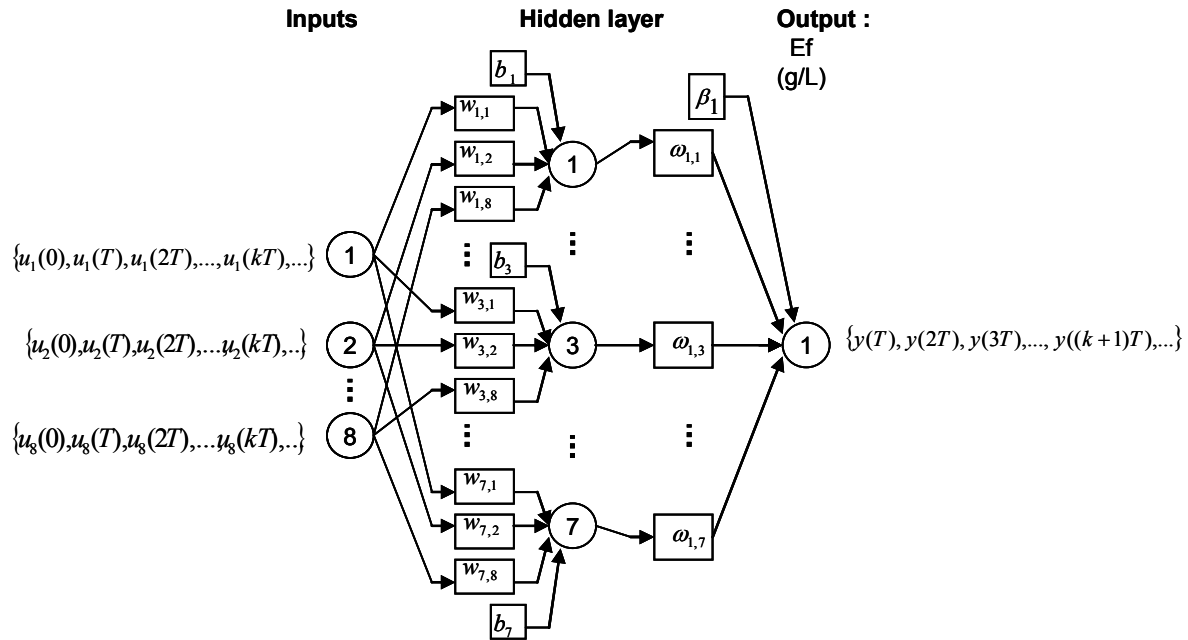


Figure 8.3.5. Topology of the ANN plant model to be controlled. Alternative II, structure A

Structure A was selected for the ANN plant model of the process to be controlled in alternative II, the representative model is the model 1 in Table 8.3.5. This model is identified as model M8 in the next chapter doing reference to a model with 8 input variables. The selected model has the lowest overfitting

because of the training epochs (epochs = 140) used to reach the best training point indicated by the MATLAB software when training the neural networks. The model was trained with a sequence of inputs $\{u(0), u(T), u(2T), \dots, u(kT), \dots\}$ and a sequence of outputs $\{y(T), y(2T), y(3T), \dots, y((k+1)T), \dots\}$, in this way the sequence of predicted outputs is obtained as a result of the sequence of inputs delayed one sampling time at each case. The given time using the arguments $T, 2T, \dots, kT, \dots$ represents multiples of the sampling interval T .

2) Structure B: models 10,11 and 12 with the measured output in the set of 9 input variables

In the control approach developed in this work the measured output $\{y(0), y(T), y(2T), \dots, y(kT), \dots\}$ (controlled variable) is one input of the MPC controller (see Figure 9.2, Chapter 9) then structure B is used in the next chapters for the ANN plant model of the controller.

8.3.6. SELECTION OF MANIPULATED VARIABLES FROM THE INPUT DATA SET

From the use transfer function method and the principal component analysis the possible manipulated variables are also: F_P , F_{PU} , F_0 , P_{flash} , F and D ; the first four have the higher influence on the variable to be controlled, the concentration of ethanol in the fermentor.

During the operation time were used dilution rates D of 0.03, 0.1, 0.15, 0.2, and 0.35 h, which were maintained constant for more than 3 residence times in each case to measure its effects on the formation of the product.

Tests were performed to verify if the dilution rate D could be used as a manipulated variable in a control scheme and the results were not successful, with

the use of normalized values (0.1 - 0.9) the results improved a bit but in general they were not successful.

In case of the fermentor outlet flow rate (F), the outflow of the fermentor contains fermentation broth and is divided after filtering into two main streams. Approximately 75% is recycled back to the fermentor, this diminishes the real effect of the variable, and 25% is sent to the flash tank. In the filtration unit the pump used should perform the circulation of the fermentation broth through the filter with sufficient pressure to force the solvent permeation and to send the fermentation broth to the flash tank. In this unit, the controlled variable is the pressure of the system and the manipulated variables are: the outlet flow rate from the fermentor (F), which is driven by the rotation of the circulation pump, and the valve opening of return to the fermentor. Therefore the fermentor outlet flow rate F cannot be used as manipulated variable in two different control loops.

The vacuum pressure of the flash tank (P_{flash}) is a controlled variable in the plant. The control system of the vacuum pressure is explained in detail in Atala (2004). The measurement of the vacuum is provided by the pressure transmitter of the flash tank and according to it the vacuum pump is driven. The operator adjusts the desired level of vacuum and the oscillation (upper and lower limits) in which the vacuum can operate. According to the reports of the plant operation, the vacuum pressure is affected by the cell concentration in the fermentation broth, when the cell concentration increases, the fermentation broth decreases its boiling point, it evaporates more easily and the phenomenon of boiling takes place. In order to avoid this phenomenon that destabilizes the control of the liquid level in the flash tank, the pressure was increased by the operator.

Therefore in this case it is not the best choice to use the vacuum pressure of the flash tank as a manipulated variable, because it depends on the cell concentration in the fermentation broth and it is necessary to take care of the operational conditions that may be affected when the vacuum level is not

appropriate. On the other hand, the cell retention that leads to modify the vacuum pressure is a characteristic tendency in the system operation, since the cells are recycled with the filtration system.

Thus, in order to develop alternatives of nonlinear predictive control in this and the following chapters the following manipulated variables were tested: the fermentor purge flow rate (F_{PU}), the permeate flow rate (F_P), and the feed flow rate to the fermentor (F_0). According to the analysis of the experimental data recorded through the data acquisition system and the reports of the plant operation these were the only options due to the characteristics of the process in bench scale.

8.3.7. CONCLUSIONS

In this section were presented two alternatives for modeling the process the first model has 13 input variables and the second was obtained by reducing the number of inputs to 8 inputs variables, in both cases the model has good values of the criteria used in the selection process and they were obtained by using different training procedures. The transfer function method was used to reduce the number of inputs and its results were confirmed by using the principal component analysis. The possible manipulated variables for control loops that were identified in this section are the fermentor purge flow rate (F_{PU}), the permeate flow rate (F_P), and the feed flow rate to the fermentor (F_0).

8.3.8. REFERENCES

Atala, D.I.P. **Assembly, instrumentation, control and experimental development of an extractive fermentative process of ethanol production**, PhD dissertation, UNICAMP, SP, Brazil. 2004.

Atala, D. I. P.; Costa, A. C.; Maciel Filho, R.; Maugeri, F.(2001). Kinetics of ethanol fermentation with high biomass concentration considering the effect of temperature. *Appl. Biochem. Biotechnol.* 91-93, 353 - 365.

Beale, M.H.; Hagan, M.T. and Demuth, H. 2012. **Neural Network Toolbox**. User's Guide. Natick, MA, USA: The MathWorks, Inc.

Boza-Condorena, E.G. and Costa, A.C. (2011). A method of transfer functions and block diagram to study the contribution of variables in artificial neural network process models. *Lecture Notes in Engineering and Computer Science. Proceedings of the World Congress on Engineering and Computer Science 2011*. Vol.II. pp.708-713. San Francisco, CA. USA.

Ccopa Rivera, E.; Atala, D. I. P.; Maugeri, F.; Costa, A. C.; Maciel Filho, R. Development of real-time state estimators for reaction-separation processes: A continuous flash fermentation as a study case. *Chem. Eng. Process.* 2010, 49, 402-409.

Costa, A. C.; Atala, D. I. P.; Maugeri, F.; Maciel Filho, R. (2001). Factorial design and simulation for the optimization and determination of control structures for an extractive alcoholic fermentation, *Process. Biochem.*, vol. 37 (2), pp.125-137.

Fonseca, E.F. (1998). **Predictive control based in neural networks of continuous alcoholic fermentation**; (M.S. Thesis) UFPE, Recife - PE. 1998.

Jackson, J.E. (1991). **A user Guide to Principal Component**, New York, John Wiley and Sons.

Kwong, W.H. **Introdução ao Controle Preditivo com Matlab**. São Carlos, SP. Brasil: EdUFSCar, 2005.

Morari, M., Lee, J., García, C. (2002). Model Predictive Control.
http://scholar.google.com/scholar?start=0&hl=en&as_sdt=0,5&cluster=8739634585205070294.

Seborg, D.E., Edgar, T.F., & Mellichamp, D.A. (2003). **Process Dynamics and Control**, (2nd ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

CAPÍTULO 9. DYNAMIC MATRIX CONTROL WITH INTERNAL MODEL BASED ON ANN OF A CONTINUOUS EXTRACTIVE PROCESS FOR BIOETHANOL PRODUCTION

CAPÍTULO DE LIVRO PUBLICADO:

UMA VERSÃO INICIAL DESTE CAPÍTULO FOI PUBLICADA EM :BOZA-CONDORENA, E.G., ATALA, D.I.P., & COSTA, A.C. (2011). *DYNAMIC MATRIX CONTROL WITH INTERNAL MODEL BASED ON ANN OF A CONTINUOUS EXTRACTIVE PROCESS FOR BIOETHANOL PRODUCTION*. IN INDUSTRIAL CONTROL SYSTEMS. ROBERT C. GILBERT (ED). NOVA SCIENCE PUBLISHERS, INC. HAUPPAUGE, NY, USA.

Abstract

The MPC (Model Predictive Control) control algorithms have been well accepted for industrial applications (DECHECHI, 1998; MORARI and LEE, 1999; PICHÉ *et al.*, 2000; TATJEWSKI, 2006). Among the advantages of these algorithms is the fact that they deal directly with interactions between inputs and outputs and are easily applied to a multivariable process (MIMO - Multiple Input Multiple Output) (SANTOS and BIEGLER, 1999). In addition, restrictions on inputs and outputs due to physical constraints and security, common in process industries, can be directly integrated into the MPC algorithm. One of the most popular predictive control algorithms is the DMC or dynamic matrix control (QIN and BADGWELL, 1997; GAWKOWSKI, 2007), which is based on a linear model.

One difficulty in using linear control algorithms is that the majority of the chemical processes exhibit a nonlinear behavior. Thus, controllers based on linear models must have their parameters adjusted in a conservative manner to ensure stability in a wide range of operating conditions, which can result in a very slow control (CHU *et al.*, 1998). In some cases, even with this conservative adjustment of parameters, the linear controller does not provide satisfactory performance. Some examples where this can occur are non-stationary processes in which an optimal profile is implemented (GATTU and ZAFIRIOU, 1999), systems that exhibit complex interactions with constraints and highly nonlinear phenomena (SANTOS and BIEGLER, 1999), and processes subject to large disturbances (RAWLINGS, 2000).

Another option for improving the performance of the control scheme is the use of predictive control algorithms based on nonlinear models (LE LANN *et al.*, 1999). Several types of models have been used for this purpose, such as differential equations, neural networks, etc. (MORARI and LEE, 1999).

Artificial neural networks have been increasingly applied to describe chemical processes due to their ability to approximate highly nonlinear systems involving a large number of state variables. Different structures of neural networks were considered as non-linear models for advanced control algorithms (FONSECA, 1998; HENRIQUE, 1999; ROHANI *et al.*, 1999; KAMBHAMPATI *et al.*, 2000; MELEIRO *et al.*, 2001; NAGY, 2007).

In this chapter, the goal is to study the nonlinear control of a continuous extractive process of bioethanol production. An algorithm based on DMC is proposed, and are used different internal models based on artificial neural networks to represent the dynamic behavior of the manipulated and controlled variables, in substitution of the linear convolution model of the DMC controller.

The extractive alcoholic fermentation process combines the fermentor and a flash vessel to continuously extract part of the ethanol produced. Experimental data from 350 h of operation of a pilot plant is used to develop the ANN based model for the control system.

Nomenclature

Cv	Cell viability (%)
D	Dilution rate (h^{-1})
F ₀	Feed flow rate to the fermentor (L/h)
F	Fermentor outlet flow rate (L/h)
F _p	Permeate flow rate (L/h)
F _{pu}	Fermentor purge flow rate (L/h)
F _{in}	Flash tank inlet flow rate (L/h)
F _{rf}	Flash tank liquid outlet flow rate (L/h)
F _{re}	Return stream flow rate (L/h)
F _v	Flash tank vapour outlet flow rate (L/h)
G	Fermentor glycerol concentration (g/L)
E _f	Fermentor ethanol concentration (g/L)
Pr _f	Ethanol concentration in flash tank return flow rate (g/L)
P _{flash}	Flash tank pressure (mmHg)
T _{flash}	Flash tank temperature (°C)
S ₀	Inlet substrate concentration (g/L)
S	Fermentor substrate concentration (g/L)
S _{rf}	Substrate concentration in flash tank return flow rate (g/L)
T _{ferm}	Fermentor temperature (°C)

t	Time (h)
t_r	Residence time (h)
V	Total volume of the system (L)
X	Fermentor biomass concentration ((%v/v)
X _{rf}	Biomass concentration in flash tank return flow rate (%v/v)

9.1. INTRODUCTION

The control algorithms MPC (Model Predictive Control) have been well accepted for industrial applications (DECHECHI, 1998, MORARI and LEE, 1999). According to Piché *et al.* (2000), these algorithms have been used in multiple applications in the refining, petrochemical, chemical, paper and food industries. Among the advantages of these algorithms is the fact that they deal directly with interaction between inputs and outputs and are easily applied to multivariable processes. In addition, restrictions on inputs and outputs due to physical constraints and security, common in processing industries, can be directly integrated into the MPC algorithm.

In the center of an MPC algorithm is a dynamic model, in the case of the DMC algorithm, a linear convolution model. Until recently, most industrial applications used linear dynamic models. These models were preferred to the nonlinear ones due to the difficulty of developing nonlinear models from data of the plant and because of computational requirements involved in using nonlinear models in MPC formulation (PICHÉ *et al.*, 2000). The rapid development in computer power, however, led to the emergence of various control algorithms involving nonlinear models (COSTA and MACIEL FILHO, 2003).

The main problem with the linear controllers is that they are not robust, i.e. they do not maintain performance in different regions of operation or changes in operating conditions (COSTA *et al.*, 2002). According to Piché *et al.* (2000), MPC algorithms based on linear models are acceptable when the process operates at a single setpoint and the use of the controller is mainly for disturbance rejection. However, many chemical processes do not operate in a single setpoint; they are operated under different conditions depending on market requirements. Thus, the

control system must be designed to keep the process operating properly at different operating points. Due to the nonlinearity of chemical processes, the changes of the operating condition usually make a linear model which well represents the process in a given operating point to present large errors in another operating point.

One possibility in dealing with this problem is to allow the linear model to adapt to changes in the process, i.e. using adaptive control algorithms (DECHECHI, 1998). However, adaptive schemes may present stability problems, especially when frequent disturbances affect the process.

Another option for improving the performance of the control system is the use of predictive control algorithms based on nonlinear models (LE LANN *et al.*, 1999). The neural networks are being increasingly applied to describe chemical processes due to their ability to approximate highly nonlinear systems with a large number of state variables. Hussain (1999) presents a review of the applications of neural networks in control processes. A review of studies in the literature (FONSECA, 1998; SANTOS *et al.* (2000); PICHÉ *et al.* (2000); Kambhampati *et al.* 2000; MELEIRO *et al.*, 2001), shows that the networks are able to describe the system in a large number of cases, and are quite versatile, since they can be incorporated with relative facility in several known strategies for nonlinear control.

Taking this into consideration, in this chapter a DMC control algorithm are proposed to control an extractive alcoholic fermentation process. In the controllers a plant model based on ANN is used to calculate the control actions. The control algorithms were programmed in MATLAB.

The article is organized as follows: 1) Basic concepts of artificial neural network modeling are briefly reviewed. 2) The extractive fermentation process for bioethanol production is described. 3) a) the dynamic matrix control with internal model based on ANN is explained, b) the inputs and output of the internal plant model of the controller are identified and, c) two alternatives to train these ANN models are explained. The purpose is to test the performance of the DMC control

algorithm when are used different training methods for the internal plant models of the controllers. 4) The results and discussion of working with the alternatives of modeling and training are presented. 5) Conclusions are given at the end of the article.

9.2. ARTIFICIAL NEURAL NETWORK MODELING

Modeling of biotechnological processes is complex, mainly due to the difficulties in determining accurate kinetic models. Neural networks can be used as a rapid and accurate modeling technique for these processes, avoiding the choice among many possible kinetic models (CCOPA RIVERA *et al.*, 2010 a and b). The alcoholic fermentation, nevertheless, is a well-studied process and its kinetics has been described by many authors (ATALA *et al.*, 2001; ANDRADE *et al.*, 2007). The difficulty in describing the process studied in this chapter with a phenomenological mathematical model is the modeling of the flash tank. Most of the substances found in the culture broth to be separated in the flash tank are polar components. Besides, the very low concentration of many other components (apart from ethanol and water), called congeners, lead to difficulties in correlating and predicting the concentration of the distilled product in the continuous extractive fermentation (ÁLVAREZ *et al.*, 2008).

In this study multilayer feedforward networks with one hidden layer and bayesian regularization were proposed to model the extractive alcoholic fermentation process, with a tan–sigmoid transfer function on the hidden layer and a linear transfer function on the output layer. The data set of inputs and targets was normalized before training to make it more efficient, in the range $[-1,1]$ as follows.

$$x_{iN} = \frac{2 * (x_i - x_{\min})}{x_{\max} - x_{\min}} - 1, \quad (9.1)$$

where: $x_{\min}$, $x_{\max}$ and x_{iN} are minimum, maximum and normalized values of the variable x_i .

The input, x_i , weights, W_{ji} , bias, b_j^1 , transfer functions, f , and the value of the output variable ge_k^O are related by the following equation:

$$ge_k^O = f^2 \left(\sum_{j=1}^N W_{kj} f^1 \left(\sum_{i=1}^M W_{ji} x_i + b_j^1 \right) + b_k^2 \right) \quad (9.2)$$

where: $j = 1, \dots, N$; $i = 1, \dots, M$; $k = 1, \dots, K$

The weight W_{ji} , have a source on the i th input (x_i), and a destination on the j th neurons on the hidden layer (layer 1). The weight W_{kj} , have a source on the j th neuron of the hidden layer, and a destination on the k th neurons of the output layer (layer 2). Superscript 1 indicates that the bias b , or the activation function f is associated with the first layer. Superscript 2 indicates layer 2. The glucose estimated in the k th neuron of the output (o) layer is represented by ge_k^O .

The training procedure requires a set of inputs and outputs, so that a particular input leads to a specific target output. During the training process, weights and biases are iteratively adjusted to minimize a performance function (F), which is expressed in terms of the mean of squared error (mse) that exists between the neural network estimated output and the specified targets.

The number of neurons in the hidden layers was determined by using the cross-validation technique in order to avoid model overfitting and to achieve good generalization from the training data set. This technique splits the data sample into a training data set and a testing data set. The performance of the trained neural network is evaluated by the ability to predict the elements of the testing data set, which is expressed in terms of mean square error (mse).

$$mse = \frac{1}{N} \sum_{k=1}^N (t_k - ge_k^O)^2 \quad (9.3)$$

In equation (9.3), k is the number of data points in the testing data set which vary between 1 and N , t_k is the k th *target value*, ge_k^O is the k th *net estimated output value*.

The information used to train the ANN based models of the fermentation process corresponds to an extractive fermentation experiment that had a total operating time of about 350 hours. The data were stored by the monitoring program of the pilot plant. 75% of the data were used for training and 25% of the data which was unknown by the network was used in the testing data set.

9.3. EXTRACTIVE FERMENTATION PROCESS FOR BIOETHANOL PRODUCTION

Figure 9.1 shows the flow sheet of the extractive fermentation process for ethanol production using sugar cane molasses as substrate and the yeast *Saccharomyces cerevisiae* in the fermentation process for conversion of sugar into ethanol.

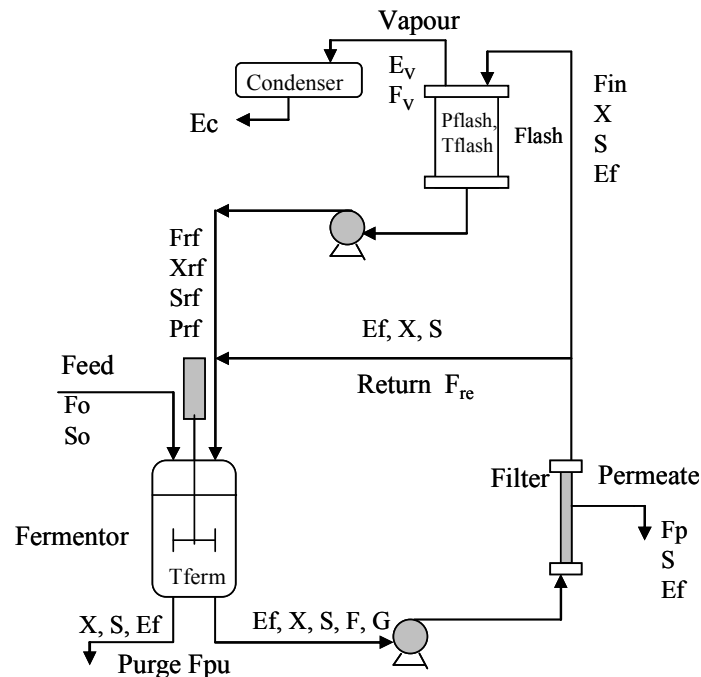


Figure 9.1. Flow diagram of the ethanol production pilot plant with extractive fermentation process.

The scheme couples the fermentor to a flash distillation system under vacuum. The flash evaporator is equivalent to a stage of a distillation column. The system operates at reduced pressures (vacuum) facilitating the separation of ethanol- water mixture at low temperatures.

The extractive fermentation system has the following equipment: fermentor, filtration system, flash tank, helical pump, peristaltic pump, vacuum system and condensing system. The process output variable (controlled variable) to the proposed control system is the fermentor ethanol concentration (E_f).

Two ANN models were used to test the model predictive control technology M13 and M8 described in Chapter 8, Part 3.

9.4. DYNAMIC MATRIX CONTROL WITH INTERNAL MODEL BASED ON ANN

9.4.1. DYNAMIC MATRIX CONTROL (DMC)

The basic concepts of the DMC control algorithm were originally presented by Cutler and Ramaker (1979). The linear DMC algorithm uses time-domain step response models (*convolution* models) to estimate future changes in the manipulated variables that lead the controlled variables to follow an optimal path (set point). This optimal path is determined by optimizing an objective function.

The controller objective is to determine a set of control actions in the future to minimize the objective function, which is composed of two parts:

- 1) The difference between the setpoint and the model predictions of the process within a prediction horizon (NP).
- 2) The sum of the squared changes in the manipulated variable multiplied by the square of the suppression factor in a control horizon (NC)

The second part is used to restrict the movements of the manipulated variable. The objective function is the following (LUYBEN, 1990):

$$J = \sum_{i=1}^{NP} [ysp - \hat{y}_{CL,i}]^2 + f^2 \sum_{i=1}^{NC} [(\Delta u_k)^{new}]^2 \quad (9.4)$$

where:

ysp = set point, f = weighting factor (suppression factor) which prevent large swings in the manipulated inputs, Δu = increment in the manipulated variable, $(\Delta u)^{new}$ = value of the future changes in the manipulated variable that minimize the performance index J .

$\hat{y}_{CL}$ = prediction made by the model on a prediction horizon called “closed loop response” so $\hat{y}_{CL,i}$, is the value at the i th step into the future.

$$\hat{y}_{CL,i} = \hat{y}_{OL,i} + \sum_{k=1}^{NC} a_{ik} (\Delta u_k)^{new} \quad (9.5)$$

$\hat{y}_{OL}$ = prediction of the open loop response.

The best values of the future changes in the manipulated variables $(\Delta u)^{new}$, such that the performance index, J , is minimized, are obtained by the equation:

$$(\Delta u)^{new} = [A^T A + f^2 I]^{-1} A^T E \quad (9.6)$$

where: $[A^T A + f^2 I]^{-1} A^T$ is a matrix of constant elements. A is the matrix of coefficients of the step response. E is the matrix of elements $e_i = ysp - y_{OL,i}$, where $i=1,2,\dots, NP$.

Minimization of the objective function, Eq. (9.1), is subject to certain restrictions on manipulated and controlled variables.

$$y_{\min} \leq \hat{y}_{k+i} \leq y_{\max} \quad (i = 1, \dots, NP) \quad (9.7)$$

$$u_{\min} \leq u_{k+i} \leq u_{\max} \quad (i = 1, \dots, NC) \quad (9.8)$$

$$|u_{k+i} - u_{k+i-1}| \leq \Delta u_{\max} \quad (i = 1, \dots, NC) \quad (9.9)$$

The DMC uses a corrected value for the prediction $\hat{y}$ in equation (9.1), incorporating a feedback strategy. At instant k , the predicted value of the output is compared to the measured value; the difference is used to correct the predicted value $\hat{y}$ in the NP moments ahead. For example, for the moment $k+1$:

$$y_{k+1}^c = \hat{y}_{k+1} + (y_k - \hat{y}_k) \quad (9.10)$$

Although the optimization is based on a control horizon, only the first control action is implemented and the optimization problem is solved again at the next sampling time. The DMC algorithm has the following steps at each point in time (LUYBEN, 1990):

- 1) Calculate the NP values of $\hat{y}_{OL}$, using the past values of the manipulated variables and the present measured value of the controlled variable.
- 2) Calculate the NC values of the future changes in the manipulated variables, $(\Delta u_k)^{new}$ from Eq. (9.6).
- 3) Implement the first change $(\Delta u_1)^{new}$.
- 4) At the next sampling period, measure the controlled variable to get a new value and repeat the steps above.

9.4.2. INTERNAL MODEL BASED ON ANN

The following diagram illustrates the model predictive control. The controller consists of the neural network plant model (internal model) and the optimization block. The optimization block determines the values of the manipulated variable u^* that minimizes the performance index, J , and then the optimal first control action u is input to the plant. In Figure 9.2 y_{sp} is the desired output, and ym is the network model output.

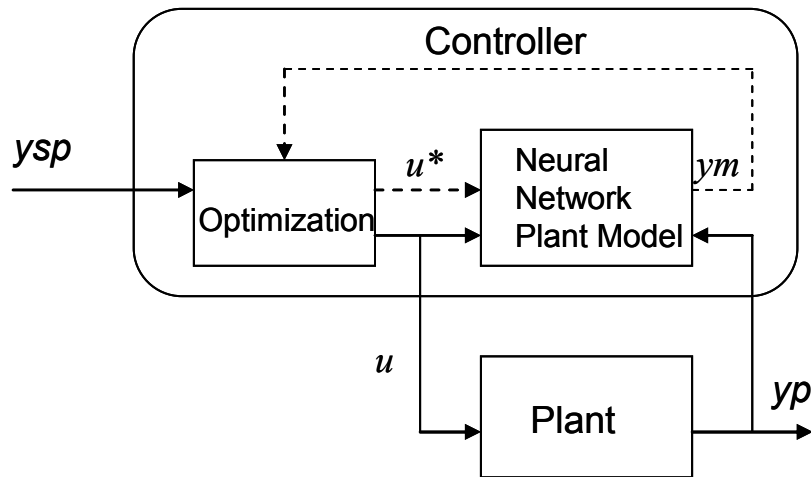


Figure 9.2. Model predictive control process

Fuente: Neural network toolbox, MATLAB.

The model predictive control is a control technique that uses the internal plant model of the controller to calculate the control actions, this model can be a model based on the conservation principles or an empirical model, and the empirical model can be linear or non linear.

Kwong (2005) states that the controller plant model usually should describe the dynamic characteristics of the plant to be controlled and that is searched a plant model which allows us to determine the sequence of outputs: $\{\hat{y}(0), \hat{y}(T), \hat{y}(2T), \dots, \hat{y}(kT), \dots\}$ resulting from a sequence of inputs: $\{u(0), u(T), u(2T), \dots, u(kT), \dots\}$. An example of linear model is the step response model for a system of first order: $y(t) = 1 - e^{-t/5}$.

According to Kwong (2005) most industrial applications of MPC has been based on empirical linear models, this was also a reason for performing this research where it is studied the use of internal plant models of the controller based on ANN to calculate the control actions in model predictive control.

The Multilayer perceptron (MLP) with one hidden layer is reported in the literature with good results to be used as dynamic plant model (Werbos,1992),

therefore, it is used in this work as internal model of the DMC controller. As was proposed in several works of nonlinear predictive control in the literature (COSTA *et al.*, 2002, 2003; SANTOS *et al.*, 2000), the internal model uses current values of the controlled variable and the manipulated variable to predict future values of the plant output. Therefore the data necessary to train the ANN internal model is: the manipulated variables in the current time $u(t)$, the controlled variable in the current time $y_p(t)$ and the controlled variable a step forward $y_p(t+1)$;

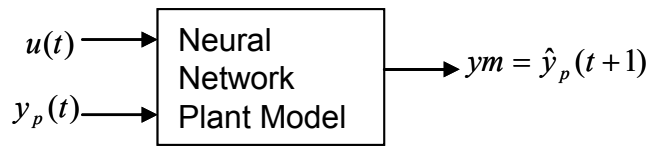


Figure 9.3. Inputs and output of the internal model of the DMC controller

In Figure 9.3 ym is the output of the model and $\hat{y}_p(t+1)$ is the future value of the plant output predicted by the internal model.

Another important step in the network training is the evaluation of their predictive ability. Usually when the network is used to identify processes, their predictive ability is tested using a set of validation data, i.e. a set of input data that have not been used for training. Fonseca (1998), however, mentions that the networks often yield satisfactory results in the validation test but do not perform well in control applications. This author suggests that the network should be evaluated for its ability to predict the steady states of the plant. Other authors used the same approach to evaluate the performance of networks used as models for predictive control schemes (SANTOS *et al.*, 2000; COSTA *et al.*, 2002, 2003).

In order to train the internal ANN model two approaches that were identified from the literature are used in the present work:

- 1) To generate the training data by making random step disturbances in the manipulated variable. In this approach an available plant model (reference model) is used to generate the training data (COSTA *et al.*, 2002, 2003; SANTOS *et al.*,

2000). The authors suggest that the training data: input - output patterns; should be generated by making random step disturbances in the manipulated variable used as input of the model (Alternative A).

2) To train the internal model by using directly the empirical data set that were obtained from the process operation (Werbos,1992). In this approach to train the ANN internal model the input – output data which were obtained by sampling the system during the operation time are used. The data is split into a training set and a testing set. The inputs are the manipulated and controlled variables in the current time and the output is the controlled variable a step forward (Alternative B).

9.5. RESULTS AND DISCUSSION

9.5.1. ALTERNATIVE A.

Plant model: M 13 with topology 13 : 7 : 1, described in Chapter 8, Part 3.

9.5.1.1. ANN internal plant model of the controller

The ANN internal model of the controller in alternative A is identified as IMA, it has a topology 2: 4 : 1, with two inputs: the controlled variable and the manipulated variable at the current time and one output: the controlled variable one time ahead. The controlled variable is the fermentor ethanol concentration (Ef) and in this chapter the manipulated variable is the purge flow rate (Fpu).

To generate data to train the ANN internal model 36 successive random step disturbances were made in the manipulated variable using the ANN plant model M13. The interval between disturbances was chosen to ensure that new steady states were reached, as suggested by Santos *et al.* (2000).

The performance in describing the training data set was measured by the following equation (MILTON and ARNOLD, 1990).

$$CORR = (1 - \frac{SEE}{S\tau\tau})100\% \quad (9.11)$$

where $SEE = \sum_{k=1}^N (y_e(k) - y(k))^2$, $S_{\tau\tau} = \sum_{k=1}^N (y_e(k) - \bar{y}_e)^2$, $y_e(k)$ is an experimental output k , $y(k)$ is the corresponding ANN output, $\bar{y}_e$ is the experimental output average and N is the number of experimental data.

The best value of *CORR* was the criterion used to select the topology of the ANN internal model of the controller.

The network was able to describe the training data with a correlation (*CORR*) of 99.21% which is calculated by equation (9.11). It provided a good description of the steady states of the process for various values of fermentor purge flow rate (F_{pu}). The purpose of this model is to use the values of the manipulated and controlled variables at each sampling time (t) to predict future values of the controlled variable, with these values the optimization block of the controller calculate the future control actions to lead the plant to the desired steady state. The structure of the internal model of the controller (IMA) is shown in Figure 9.4.

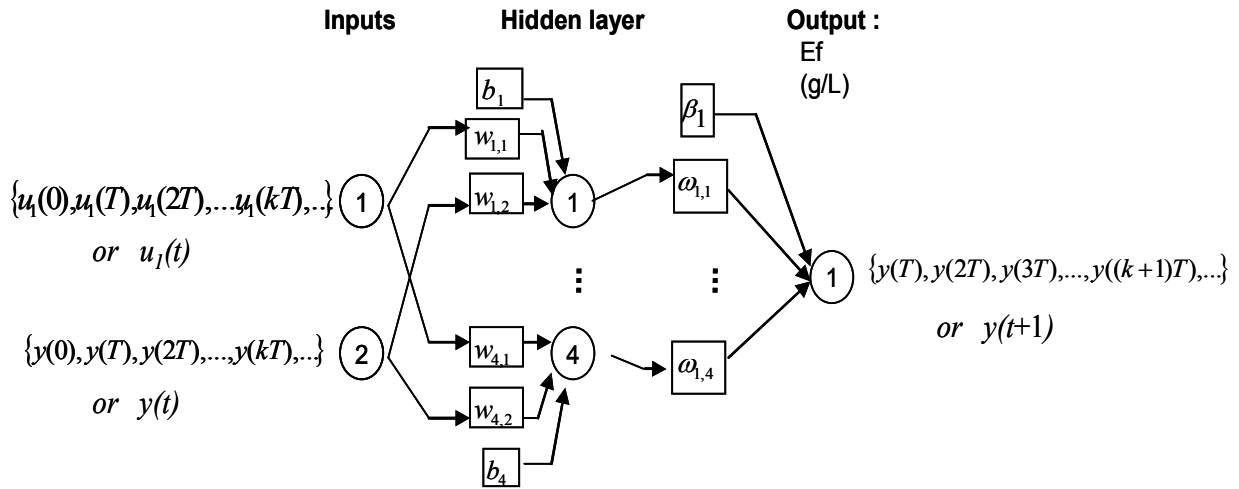


Figure 9.4 Alternative A: Topology of the internal ANN plant model of the controller

9.5.1.2. Non linear predictive control

A SISO nonlinear predictive control scheme was developed to control the ethanol concentration in the fermentor manipulating the purge flow rate. The internal model neural network was trained using the current values of manipulated and controlled variables as input and the controlled variable one step ahead as output. As the network is trained to predict only one step ahead, in the control scheme developed, the network is iterated for a total of NP predictions in the future, where NP is the prediction horizon. Thus, the output of the same network is used as input in the subsequent iterations. The sampling time considered was 30 minutes. Constraints on controlled and manipulated variables were the following:

$$Ef_{k+1} \geq 0 \quad (i = 1, \dots, NP)$$

$$5.7 \leq F_{PU} \leq 180 \quad (i = 1, \dots, NC)$$

$$\left| (F_{PU})_{k+i} - (F_{PU})_{k+i-1} \right| \leq 50 \quad (i = 1, \dots, NC)$$

The controller parameters were chosen after trial and error tests:

$$f = 0.18, NP = 10 \text{ and } NC = 1.$$

To assess the controller performance rigorous tests were performed by changing the setpoint and the permeate flow rate (F_P). This variable was chosen as a disturbance because it is the variable that most influences the controlled variable besides the manipulated variable F_{PU} .

- Regulator performance

In this case in order to test the regulator performance of the controller step disturbances of $\pm 10\%$ were made in the permeate flow rate (F_P). The results are shown in Figures 9.5 and 9.6. The model used to simulate the plant is M13 and the internal model of the controller is IMA.

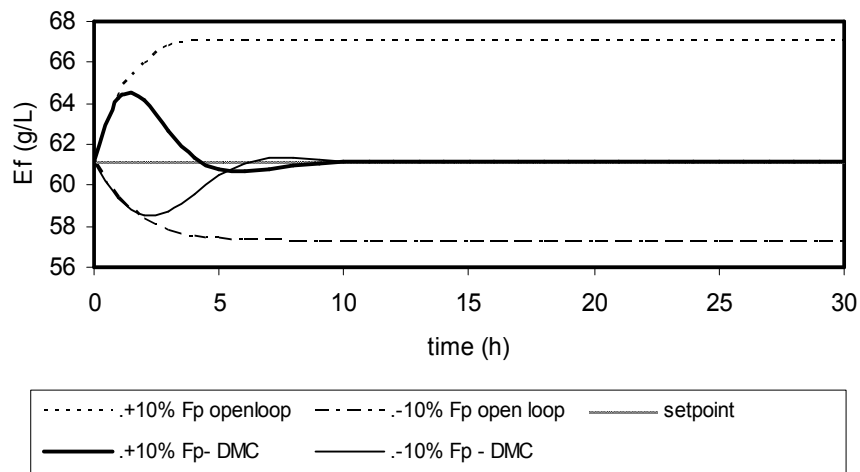


Figure 9.5. Controlled variable responses against time for step disturbance of $\pm 10\%$ in the permeate flow rate (F_p).

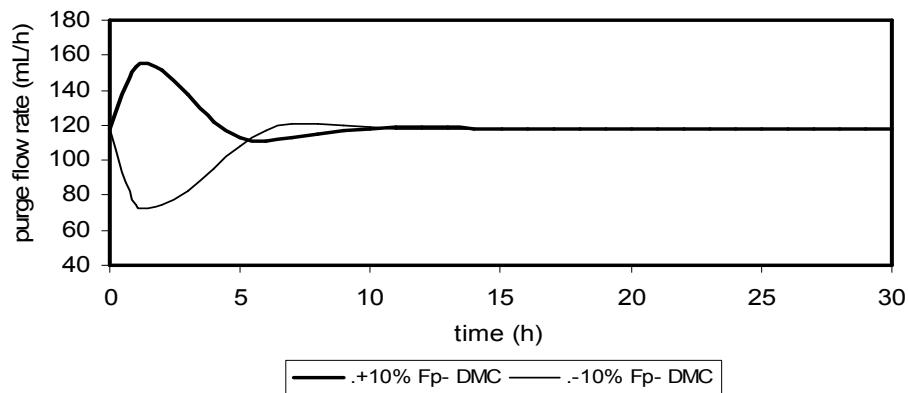


Figure 9.6. DMC controller actions to reach the steady state for an initial step disturbance of $\pm 10\%$ in the permeate flow rate (F_p).

Figures 9.5 and 9.6 show the behavior of controlled and manipulated variables with time. In Figure 9.5 the system response in open loop to the same disturbances is also shown. It may be noted that the controller was able to bring the system back to steady state.

- Servo performance

In this case successive step disturbances were made in the value of the set point using the model M13 to simulate the plant and the internal model of the controller IMA. Figures 9.7a and 9.7b show the behavior of controlled and

manipulated variables, respectively. It can be observed that the controller performed well bringing the system back to a steady state.

The behavior of the purge flow rate (F_{pu}) and the fermentor ethanol concentration (E_f) in Figures 9.7a and 9.7b may be explained in the following way. The extractive alcoholic fermentation process combines the fermentor and a flash vessel to continuously extract part of the ethanol produced. When the ethanol is extracted the condensed ethanol concentration (E_c) increases and the fermentor ethanol concentration (E_f) decreases.

According to the operation reports (ATALA, 2004), with the increase of cell concentration, the fermentation broth decreases its boiling point, it evaporates more readily and therefore, takes place the phenomenon of boiling. Therefore when increasing the cell concentration, more ethanol is extracted in the condensed stream and the fermentor ethanol concentration (E_f) decreases. The purge directly modify the cells concentration in the system, increasing the purge decreases the cell concentration, decreasing the purge increases the cell concentration, therefore when decreasing the purge (F_{pu}), the cell concentration increases, the ethanol extracted increases and the fermentor ethanol concentration decreases (E_f).

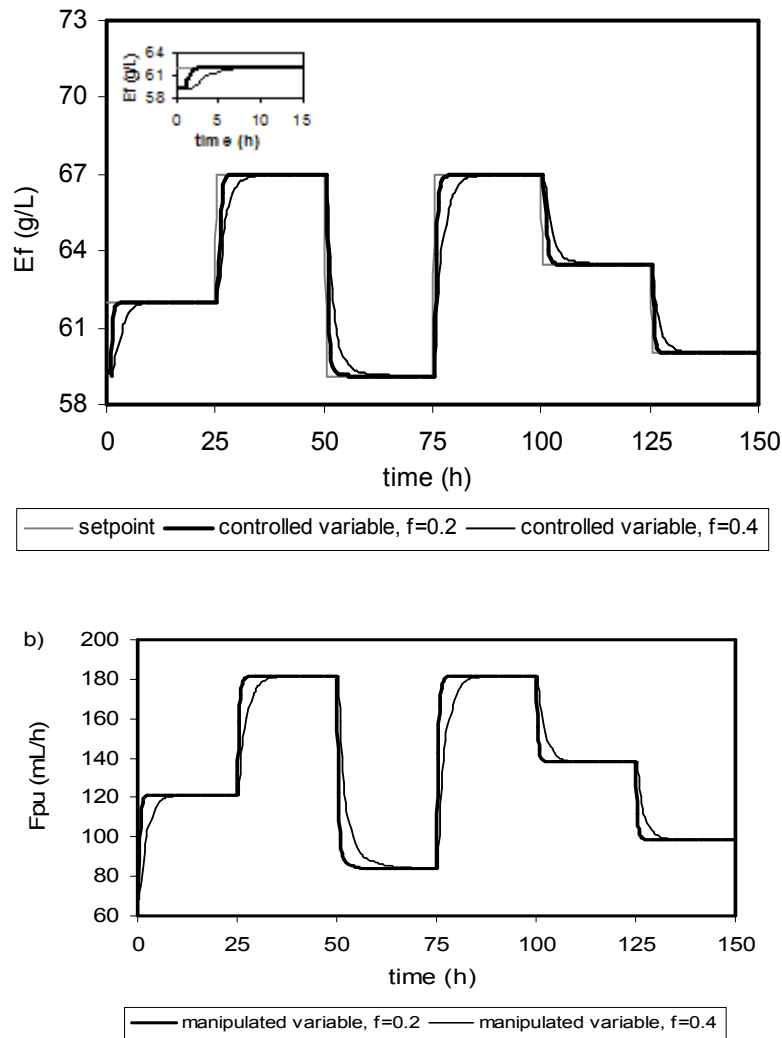


Figure 9.7. Alternative A, controller behavior for *set-point* changes:

a) Controlled variable responses b) Manipulated variable changes.

- Effect of the weighting factor f , on the damping ratio of the closed loop response of the DMC controller with ANN based internal model.

Figure 9.8 shows that the smaller the weighting factor f , the bigger the changes in the controlled variable, so the controller response is faster but with increased oscillation. The bigger the weighting factor f , the bigger the damping effect on the response, so the controller response is slower against perturbations. In these tests the prediction horizon (NP) and the control horizon (NC) remained constant.

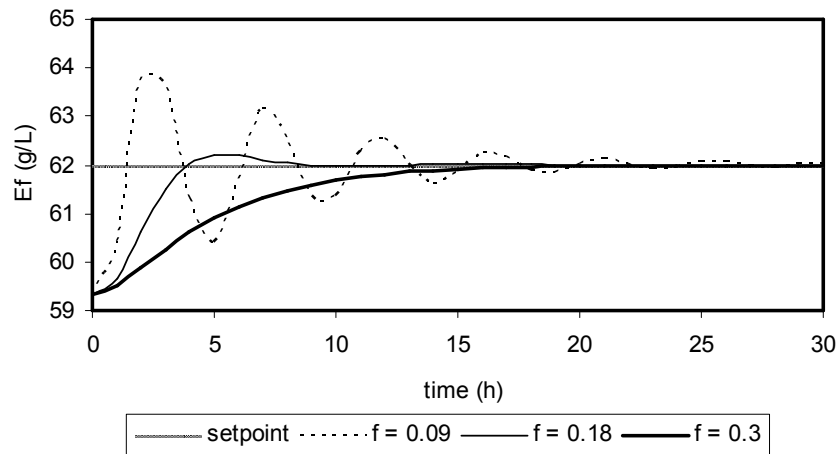


Figure 9.8. Damping effect for different values of f the weighting factor for Δu 's in the controller objective function.

Figure 9.9 shows that the smaller the weighting factor f , the bigger the changes in the Δu 's, and the smaller the damping effect in the response of the controller.

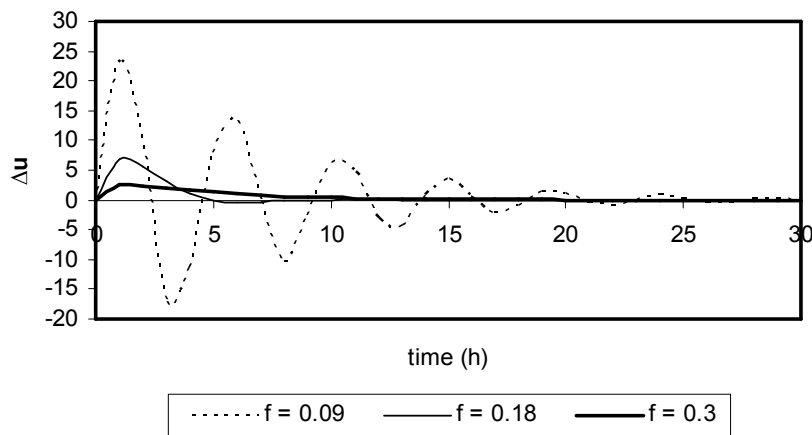


Figure 9.9. Changes in the manipulated variable for different values of f the weighting factor for Δu 's in the controller objective function.

9.5.2. ALTERNATIVE B

Plant model: M 8 with topology 8 :7: 1, described in section Chapter 8, Part 3.

9.5.2.1. ANN internal plant model of the controller

The general ANN internal model of the controller in alternative B is showed in Figure 9.10.

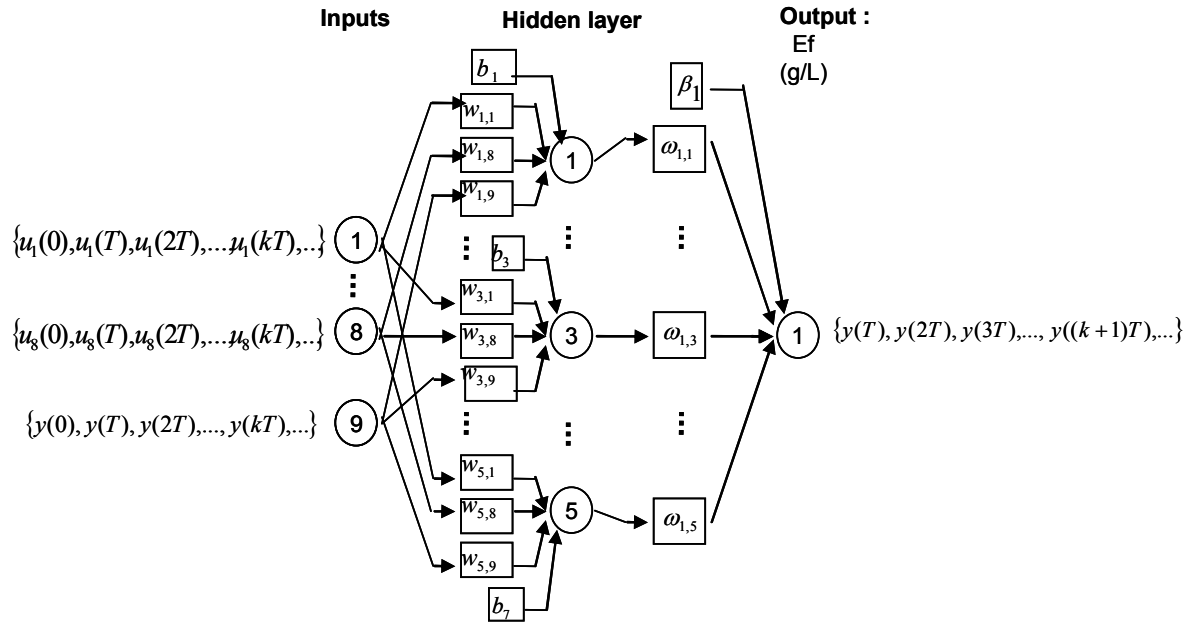


Figure 9.10. General topology of the ANN plant model of the controller.

The principal difficulty to implement this structure is that it is not possible to have available on line measures of all the inputs, sugar concentrations in substrate, biomass, product and other related variables need off line determination at laboratory. In this Chapter are used internal models with a simplified structure in which the inputs are the manipulated variable in the current time, on line available in the plant and the controlled variable also in the current time. The output is the controlled variable a step ahead. This structure is proposed in several works of nonlinear predictive control in the literature (COSTA *et al.*, 2002, 2003; SANTOS *et al.*, 2000).

To train the ANN internal model was used the input – output data which were obtained by sampling the system during the operation time (Werbos, 1992).

The data was split into a training set (75%) and a testing set (25%). In this case after training various models with similar performance were obtained. The criteria used to select the best one and the results are shown in Table 9.1.

Table 9.1. ANN internal model of the controller: values of the criteria used for selecting the model when is used the purge flow rate (Fpu) as manipulated variable.

ANN internal Model	Topology	R ²	Mse	Parameter Number	mswb	COOR	epoch
IMB1	2:3:1	0.99975	0.0030	13	1.49	99.97	255
IMB2	2:2:1	0.99975	0.0030	9	2.63	99.97	137
IMB3	2:2:1	0.99973	0.0032	9	15.13	99.97	221

The ANN plant model IMB1 with three neurons in the hidden layer was selected to be used as internal plant model of the controller because among the alternatives it has the lowest value of the average sum of the squares of the weights and bias (mswb), it is a good characteristic of neural nets and is used in neural networks with Bayesian regulation to prevent overfitting. The magnitude of the training epoch is usually greater when the neural net is more complex, and this is the case of IMB1. The topology of the selected ANN is shown in Figure 9.11.

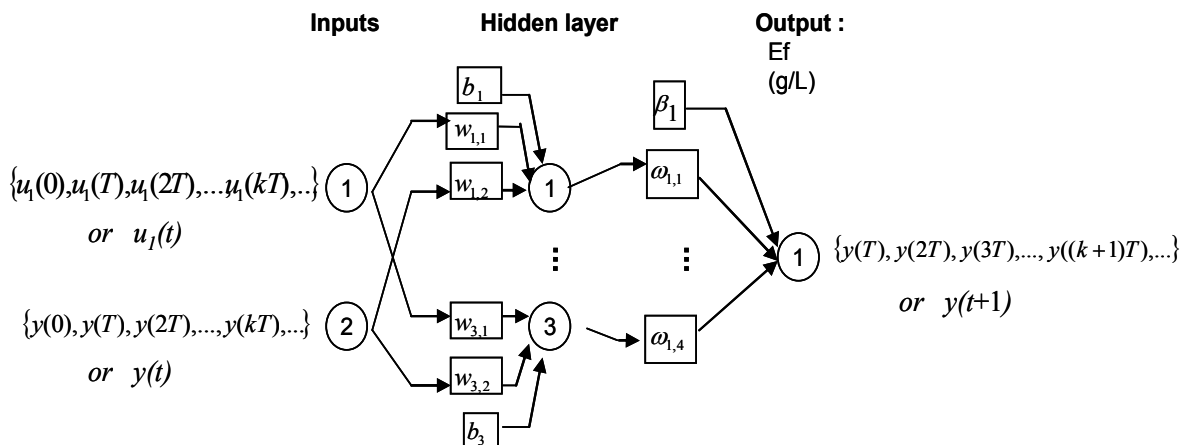


Figure 9.11 Alternative B: Topology of the internal ANN plant model of the controller

In Figure 9.11 y is the controlled variable which is the fermentor ethanol concentration (E_f), and u is the manipulated variable, the purge flow rate (F_{pu}). The arguments $0, T, 2T, \dots, kT, \dots$ denote the given time as a multiple of the sampling interval T .

9.5.2.2. Controller performance

The predictive controllers with nonlinear internal model based on neural networks were subjected to various tests, considering a sampling time of 30 minutes. The parameter values were determined by trial and error; and comparing the controller performance in different cases:

Alternative B

Plant model: M8 (see Chapter 8, Part 3)

Internal plant model of the controller: IMB1

Control structure:

Controlled variable: Fermentor ethanol concentration (E_f)

Manipulated variable: Fermentor purge flow rate (F_{pu})

Prediction horizon, $NP=10$, control horizon, $NC=1$, $f=0.2$; $f=0.4$.

9.5.2.2.1. Regulator performance

In this test, changes were made in the permeate flow rate (F_p), and in the feed flow rate to the fermentor (F_o) to simulate load disturbances and to observe the regulatory behavior of the controllers. In this case the manipulated variable was the fermentor purge flow rate (F_{pu}).

Four cases are mentioned to describe different characteristics on the response. Figures 9.12, 9.13, 9.14 and 9.15 show that the NNMPC structure reject the disturbance and lead the system back to the steady state; therefore it has good regulatory performance.

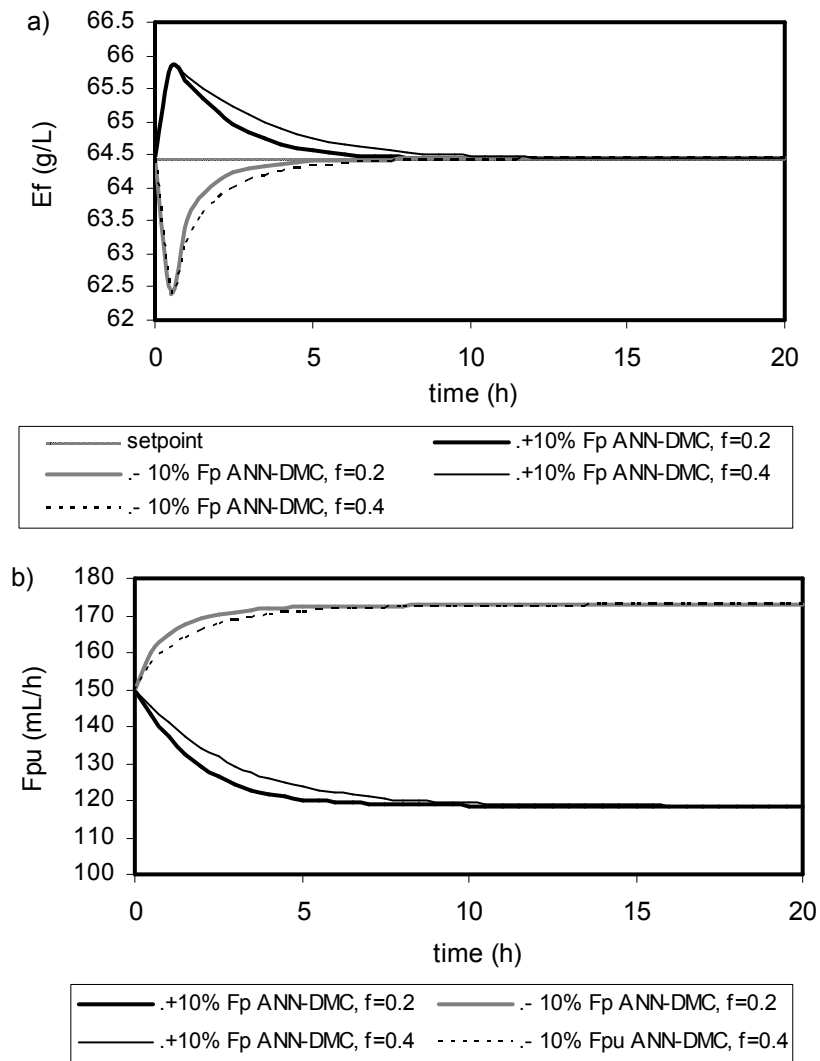


Figure 9.12. Case 1: Alternative B, a) response of the controlled variable, E_f , with time to perturbations of $\pm 10\%$ in the permeate flow rate (F_p). b) manipulated variable: purge flow rate (F_{pu}).

Figures 9.12 and 9.13 show the responses of the control structure for disturbances of $\pm 10\%$ in the value of the permeate flow rate, F_p , in these figures the magnitude of the disturbance is the same but the effect of the disturbance is different. Figures 9.14 and 9.15 show the responses of the control structure for disturbances of $\pm 5\%$ in the value of the feed flow rate, F_o ; in these figures also the magnitude of the disturbance is the same but the effect of the disturbance is

different. In all the cases the control structure is able to return the system to steady state.

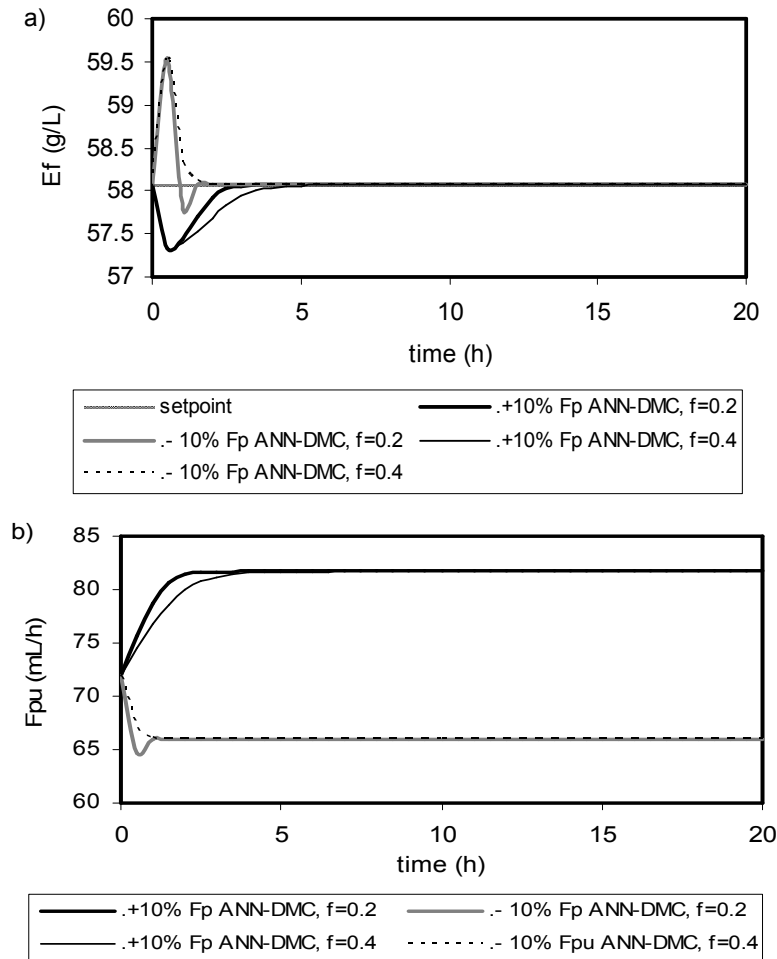


Figure 9.13. Case 2: Alternative B, a) response of the controlled variable, E_f , with time to perturbations of $\pm 10\%$ in the permeate flow rate (F_p), b) manipulated variable: purge flow rate (F_{pu}).

Figure 9.12 and 9.13 show that the regulatory performance of the control structure is better when it is used $f = 0.2$ than $f = 0.4$, in the last case the response of the system is slower.

Figures 9.14 and 9.15 show the regulatory behaviour when changes of $\pm 5\%$ were made in the feed flow rate to the fermentor (F_o) to simulate load disturbances. In Figure 9.14 the ethanol concentration is high with low conversion of

sugars and in Figure 9.15 the ethanol concentration is low with high conversion of sugars.

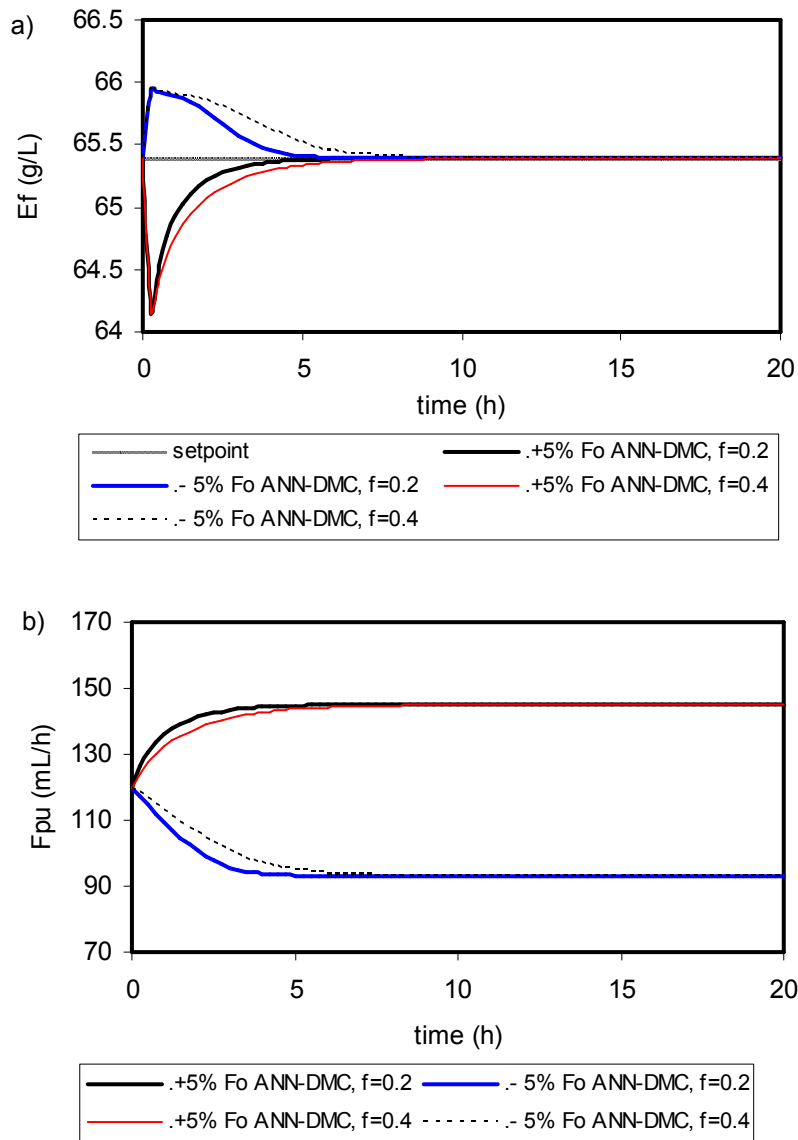


Figure 9.14. Case 3: Alternative B, a) response of the controlled variable, E_f , with time to perturbations of $\pm 5\%$ in the feed flow rate to the fermentor (F_o) b) manipulated variable: purge flow rate (F_{pu}).

Figures 9.14 and 9.15 show different effects of the disturbances that can be explained as follow: In the plant dynamics the effect of the perturbations depend on the concentration of ethanol among other factors i.e. when the ethanol concentration is high the cell viability decreases, it affects the conversion of sugars into ethanol, in

conditions of low conversion of sugars an increase in the feed flow rate produces a decrease in the ethanol concentration, the effect is different when there are conditions of high conversion of sugars, in this case an increase in the feed flow rate produces an increase in the ethanol concentration of the fermentation broth.

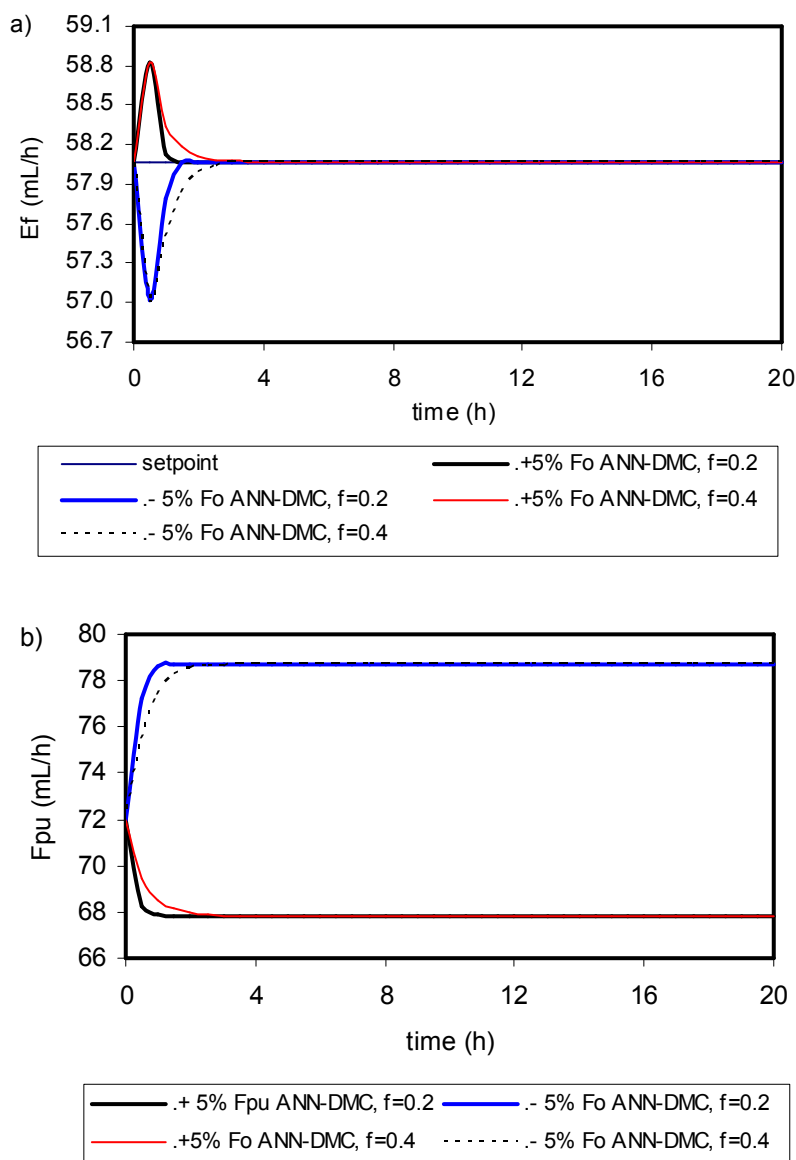


Figure 9.15. Case 4: Alternative B, a) response of the controlled variable, E_f , with time to perturbations of $\pm 5\%$ in the feed flow rate to the fermentor (F_o). b) Manipulated variable: purge flow rate (F_{pu}).

9.5.2.2.2. Servo performance

In this case successive changes were made in the value of the reference trajectory (setpoint). Figure 9.16 show the behavior of the predictive controller with ANN internal model to various changes in the reference trajectory. It is noted that the control structure performed well, bringing the system back to the steady state after each change in the setpoint.

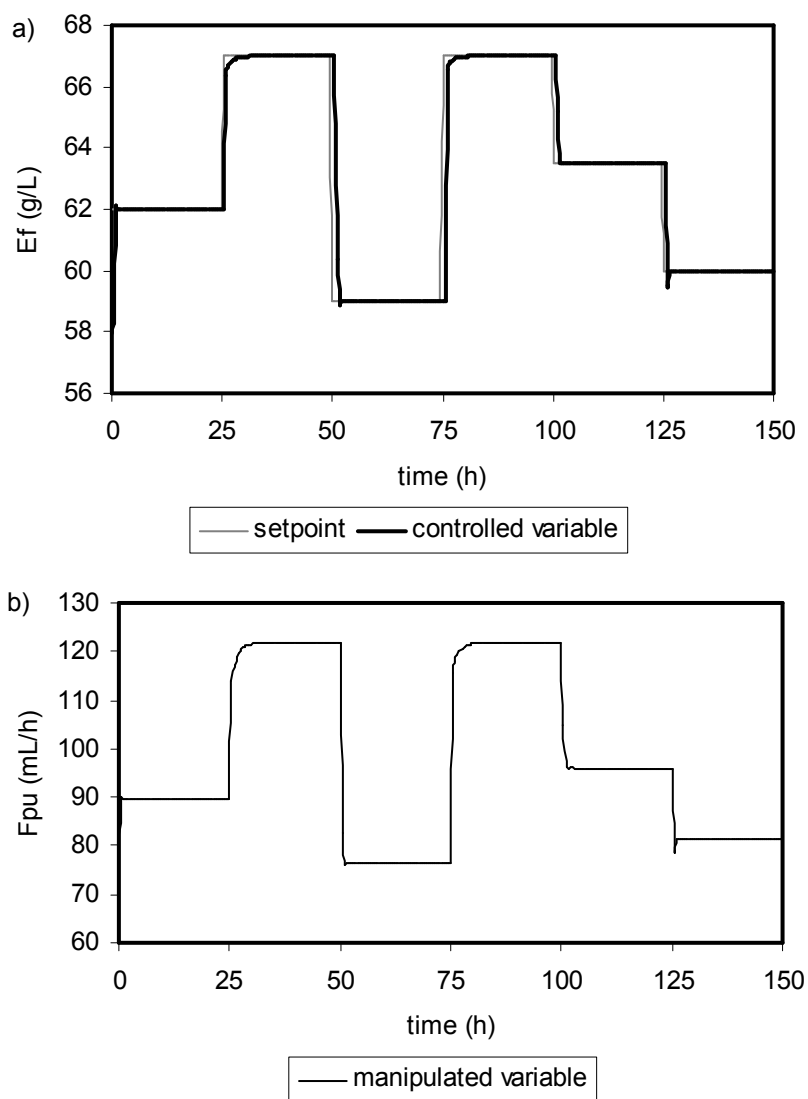


Figure 9.16. Alternative B, servo performance: a) controlled variable E_f and setpoint changes
b) manipulated variable purge flow rate (F_{pu}).

9.5.3. SUBSTITUTION OF THE ANN INTERNAL MODEL OF THE CONTROLLER FROM ALTERNATIVE B IN THE CONTROLLER OF ALTERNATIVE A

The results for regulatory and servo performances are shown in Figures 9.17 and 9.18 to obtain a good performance with the control structure it was necessary to adjust the f parameter by using trial and error tests, in this case best performance were obtained with $f = 0.2$, $NP=10$, $NC=1$. The obtained control structure performed well, bringing the system back to the steady state after the perturbation.

In this case the internal model of the controller developed in alternative B was used as the internal model of the controller of the alternative A. The internal model for alternative A was obtained by performing disturbances in the manipulated variable using the ANN plant model to simulate the plant behavior. In alternative B the ANN internal model was obtained by using the experimental data set to train the model. Therefore both have different topologies. In case of alternative A the topology is 2:4:1 and in case of alternative B it is 2: 3:1

- **Regulator performance**

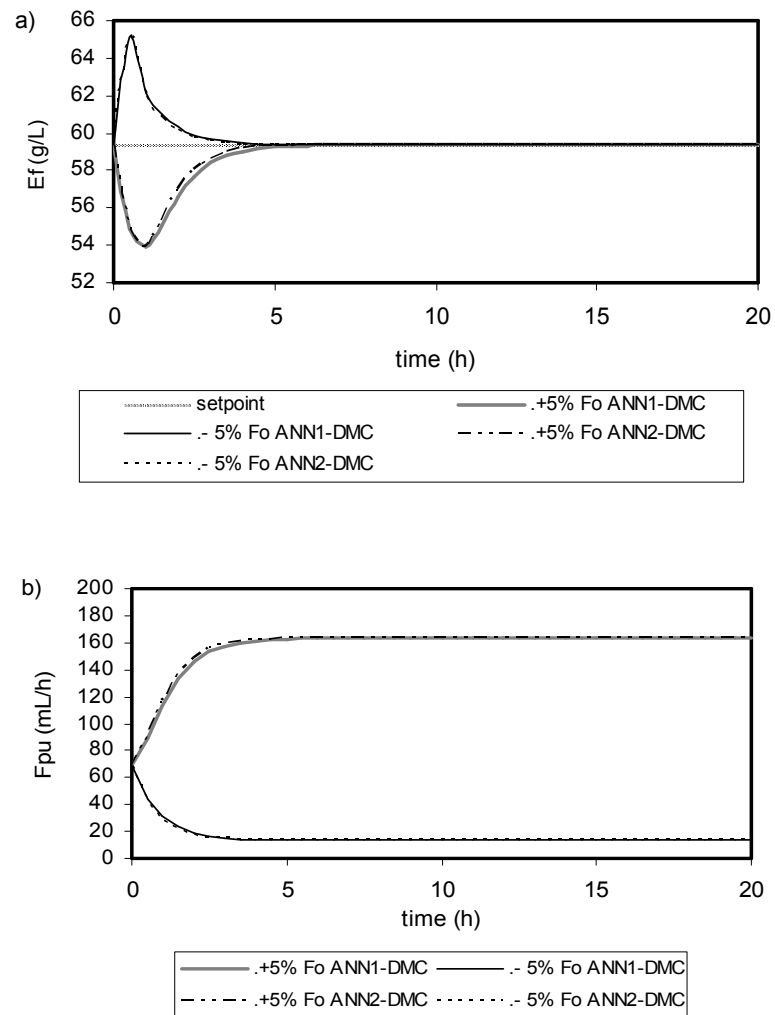


Figure 9.17. Use of the internal model from alternative B in alternative A: controller regulator performance, response of the controlled variable, E_f , with time to perturbations of $\pm 5\%$ in the feed flow rate to the fermentor (F_o). Manipulated variable: purge flow rate (F_{pu}).

- **Servo performance**

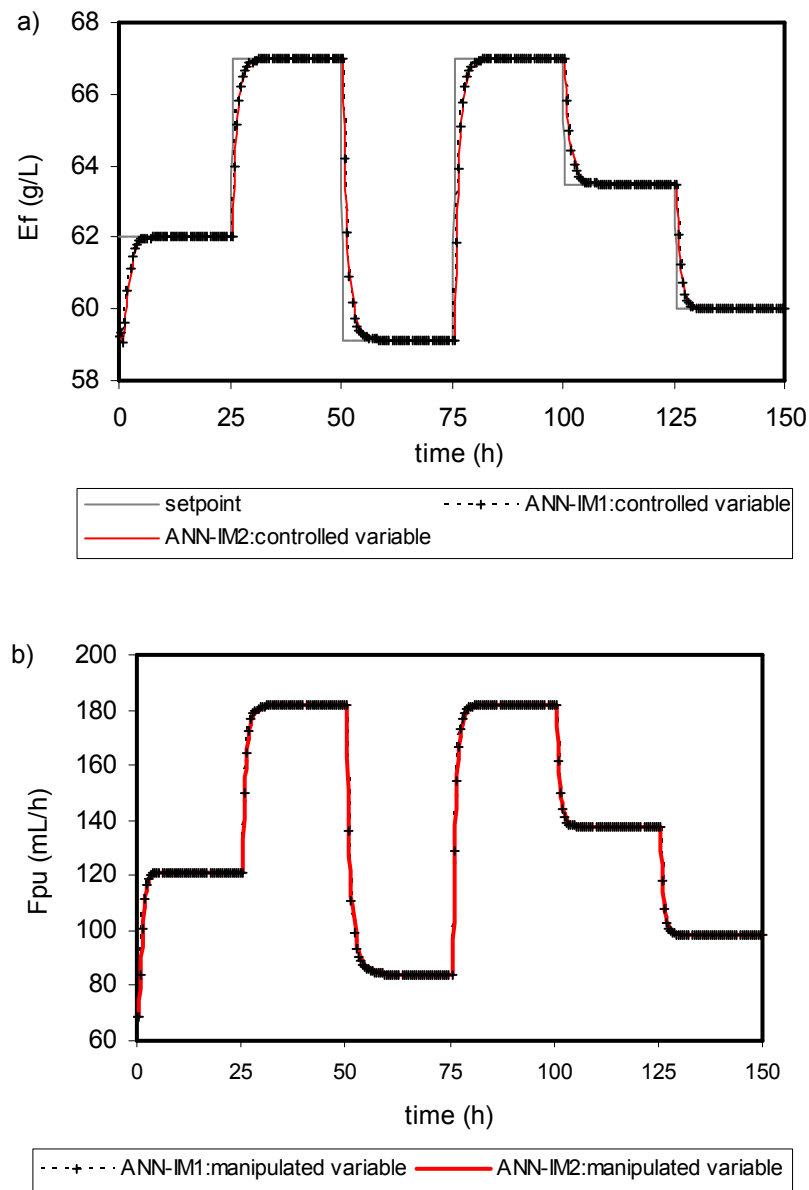


Figure 9.18 Use of the internal model from alternative B in alternative A: controller servo performance: a) Controlled variable E_f and setpoint changes. b) Manipulated variable purge flow rate (F_{pu}).

The results for regulatory and servo performances show that both control structures, the original one for alternative A and the obtained control structure by substitution of the internal model of the controller from alternative B have a quite similar regulatory and servo performances. Therefore the internal model developed by training directly the model with the empirical data set was used efficiently in

different controllers. Therefore, this could be other useful procedure to develop ANN internal models for predictive controllers.

9.6. CONCLUSION

In this chapter a nonlinear control algorithm was developed to control the ethanol concentration in the fermentor of a continuous extractive fermentation process using a feedforward neural network as the internal model in the control strategy based on DMC.

The feedforward network was shown to be suitable for its use as an internal model of the nonlinear controller. One of the most important stages in the development of the nonlinear controller was the development of the internal model. The choice of inputs and outputs of the network, how to generate the set of training data and how to evaluate the predictive ability of the network were very important to the performance of the nonlinear controller. The choice of a network whose inputs are the manipulated and controlled variables in the current time, and the output is the controlled variable one step ahead was adequate in the case study considered. The developed nonlinear controller showed good control performance in all tests that were applied.

9.9. REFERENCES

1. Andrade, R.R.; Ccopa Rivera, E.; Costa, A. C.; Atala, D. I. P.; Maugeri, F.; Maciel Filho, R. *Appl. Biochem. Biotechnol.* 2007, 136 - 140, 753 - 763.
2. Álvarez V.H.; Ccopa Rivera, E.; Costa, A. C.; Maciel Filho, R.; Wolf Maciel M.R.; Aznar, M. *Appl. Biochem. Biotechnol.* 2008, 148, 141-149.
3. Atala, D. I. P.; Costa, A. C.; Maciel Filho, R.; Maugeri, F. *Appl. Biochem. Biotechnol.* 2001, 91-93, 353 - 365.
4. Boza, E.; Costa, A. (2010). The method of Transfer Functions and Block Diagrams to study the contribution of variables in artificial neural networks used for chemical and bio-process modeling. Manuscript submitted for publication.

5. Ccopa Rivera, E.; Atala, D. I. P.; Maugeri, F.; Costa, A. C.; Maciel Filho, R. *Chem. Eng. Process.* 2010, 49, 402-409.
6. Ccopa Rivera, E.; Rabelo, S.C.; Reis Garcia, D.; Maciel Filho, R.; Costa, A. C. *J. Chem. Technol. Biotechnol.* 2010, 85, 983 - 992.
7. Chu, J.; Su, H.; Gao, F.; Wu, J. *Annu. Rev. Contr.* 1998, 22, 59-72.
8. Costa, A. C.; Atala, D. I. P.; Maugeri, F.; Maciel Filho, R. *Process Biochem.* 2001, 37(2), 125-137.
9. Costa, A.C.; Meleiro, L.A.C.; Maciel Filho, R. *Process Biochem.* 2002, 38(5), 743-750.
10. Costa, A.C.; Maciel Filho, R. *Can. J. Chem Eng.* 2003, 81, 1109-1118.
11. Dechechi, E.C. *Predictive adaptative advanced control "DMC Multivariavel Adaptative"*; (PhD thesis) UNICAMP: Campinas - SP, 1998.
12. Fonseca, E.F. (1998). *Predictive control based in neural networks of continuous alcoholic fermentation*; (M.S. Thesis) UFPE, Recife - PE. 1998.
13. Gattu, G.; E. Zafiriou. *Chem. Eng. J.* 1999, 75, 21-29.
14. Gawkowski, P.; Lawrynczuk, M.; Marusak, P.; Sosnowski, J.; Tatjewski, P. In *Proceedings of the International Multiconference on Computer Science and Information Technology*; Ganzha, M; Paprzycki, M; Pelech-Pilichowski; Wisla, Poland, 2007; Vol 2, pp 903–912.
15. Henrique, H.M. *A Contribution to the Study of Neural Networks Applied to Process Control*; (PhD thesis) PEQ/COPPE/UFRJ, Rio de Janeiro, RJ, 1999.
16. Hussain, M.A. *Artif Intell En.* 1999, 13, 55-68.
17. Kambhampati, C.; Mason, J.D.; Warwick, K. *Automatica.* 2000, 36. 485-495.
18. Kwong, W.H. **Introdução ao Controle Preditivo com Matlab**. São Carlos, SP. Brasil: EdUFSCar, 2005.
19. Le Lann, M.V.; Cabassud, M.; Casamatta, G. *Annual Reviews in Control.* 1999, 23, 25-34.
20. Lima, N. M. N. *Modeling and predictive hybrid control by fuzzy logic for polymerization processes*; (M.S. Thesis) UNICAMP: Campinas - SP, 2006.

21. Luyben, W. *Process modeling, simulation and control for chemical engineers*; Second Edition; McGraw-Hill: New York, NY, 1990, pp 287.
22. Meleiro, L. A. C.; Filho, R. M.; Campello, R. J. G. B.; Amaral, W. C. In *Application of Neural Network and Other Learning Technologies in Process Engineering*; Mujtaba, I. M. and Hussain, M.A., Imperial College Press. UK, 2001.
23. Milton, J.S.; Arnold, J.C. *Introduction to Probability and Statistics*; McGraw Hill, New York, NY, 1990.
24. Morari, M.; Lee, J.H. *Comput. Chem. Eng.* 1999, 23, 667- 682.
25. Nagy, Z, K. *Chem. Eng. J.* 2007. 127(1-3), 95-109.
26. Piché, S.; Sayyar-Rodsari, B.; Johnson, D.; Gerules, M. *IEEE Contr. Syst. Mag.* 2000, 20(3), 53-62.
27. Qin, S.J.; Badgwell, T.A. In *Proceedings of the 5th International Conference on Chemical Process Control*; Kantor, J.C. ; Garcia, C.E.; Carnahan, B.; *AIChE Symposium Series No.316*; Tahoe City, CA, 1997; Vol. 93, pp 232-256.
28. Rawlings, J.B. Tutorial overview of model predictive control. *IEEE Contr. Syst. Mag.* 2000, 20(3), 38-52.
29. Rohani, S.; Haeri, M.; Wood, H.C. *Comp. Chem. Eng.* 1999, 23, 279-286.
30. Santos, L.O.; Biegler, L.T. *J. Proc. Cont.* 1999, 9, 233-246.
31. Tatjewski, P.; Lawrynczuk, M. *Int. J. Appl. Math. Comput. Sci.* 2006,16 (1), 7–26.
32. Werbos, P.J. (1992). Neural Networks, System Identification, and Control in the Chemical Process Industries. In D.A. White, D.A. Softge (Eds.), *Handbook of Intelligent Control. Neural, Fuzzy, and Adaptative Approaches*. Van Nostrand Reinhold. New York, NY. 283-356.

CAPÍTULO 10. NON- LINEAR PREDICTIVE CONTROL OF A FERMENTOR IN A CONTINUOUS REACTION-SEPARATION PROCESS

Uma versão anterior deste capítulo foi publicada em :

Boza-Condorena, E.G., Atala, Daniel I.P. and Costa, A.C. (2011). Non-linear predictive control of a fermentor in a continuous reaction-separation process. *Lecture Notes in Engineering and Computer Science. Proceedings of the World Congress on Engineering and Computer Science 2011*. Vol. II, pp. 663-668, San Francisco, CA. USA.

Abstract

In this paper a non-linear predictive controller with an empirical internal model based on Artificial Neural Networks (ANNs) is proposed. The ANN has high non-linear approximation capability and makes possible the use of plant information to generate future control actions. The results show the high potential of the proposed procedure when applied to the continuous extractive fermentation process of bioethanol production, an integrated reaction-separation process with highly complex non-linear dynamics.

Index Terms: Artificial neural networks, continuous extractive fermentation process, non-linear predictive control.

10.1. INTRODUCTION

The increasing demand for energy caused by the global economic growth generates a series of environmental problems and, in this context, there is a great interest in developing technologies for sustainable bioenergy production.

Among several options, the ethanol obtained by fermentation of sugarcane is an attractive biofuel to be used as a substitute for gasoline and can help to reduce gas emissions that produce the greenhouse effect [1]. This characteristic has increased the demand for bioethanol, which makes the development of more efficient production technologies desirable. One alternative is to apply process intensification techniques in continuous production processes. Integrated processes of reaction-separation provide alternative solutions [2], such as the production of ethanol by fermentation with a continuous withdrawal of the ethanol produced using a flash separation unit. This action regulates the concentration of ethanol in the fermentor to ranges where the inhibitory effect of the product on the yeast activity decreases, improving the process productivity [2] [3] [4]. However, these bioprocesses are highly nonlinear, its mathematical modeling is complex and there are additional difficulties to calculate the vapor-liquid equilibrium due to the

composition of the fermented liquid [5]. These characteristics also make the use of classical control techniques to control the process difficult [6].

In this paper neural network model predictive controllers, NNMPC, are designed to control the continuous production process of bioethanol by fermentation and separation via flash evaporation. The controller is based on the dynamic matrix control (DMC) algorithm, which is representative of the MPC technology [7] [8]. The control actions are obtained by using quadratic programming at each sampling time, by this reason the control algorithm is named as QDMC (Quadratic Dynamic Matrix Control). The performance in control of different internal models of the controller and process models are tested using QDMC controllers.

10.2. PROCESS DESCRIPCION

10.2.1. EXPERIMENTAL STAGE

The research experimental stage was conducted at the Bioprocess Engineering Laboratory of the State University of Campinas, Brazil.

The extractive fermentation process is integrated by: one reaction unit (fermentor), one cell recycle system (cross-flow microfiltration), one vacuum separation system (flash tank for ethanol-fermented broth separation and vacuum pump), two helical pumps, three peristaltic pumps, and one condensing unit. The total working volume is approximately 5 L. Figure 10.1 shows a diagram of the extractive fermentation process with vacuum flashing [4].

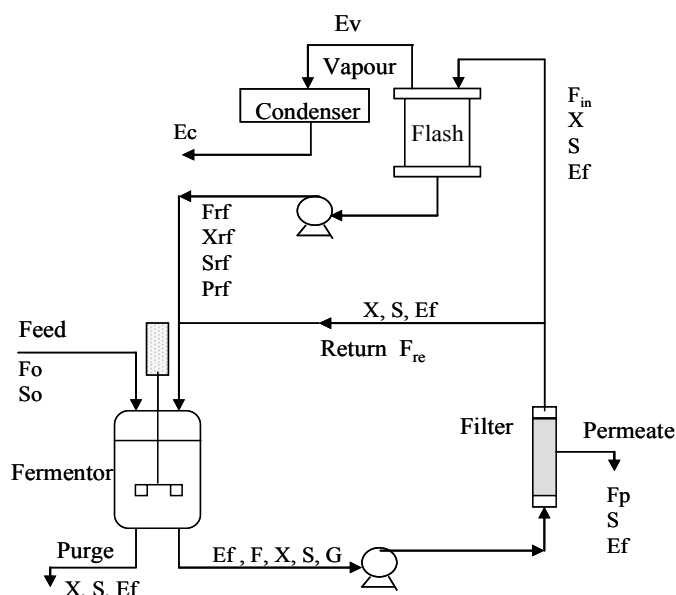


Figure10.1. Diagram of the extractive fermentation process with vacuum flashing

A 3 liters "Bioflo III System" (New Brunswick Scientific Co., Inc., NJ, USA) bioreactor with PID (Proportional, Integral and Differential) control of temperature and agitation was used as the reaction unit. A cross-flow microfiltration unit (Ceraflow model, Millipore Co.) with a filter element made of high purity alumina, 0.22 mm pore and 0.0372 m² filtration area was used in the cell recycle system. The flash tank was a 2.5 liters (working volume) adapted Chemap reactor. The device to measure the input flow was an electromagnetic flowmeter (IFS 400 KC model, signal converter IFC 090 model; Conaut, Brazil) with an operating range from 0 to 1200 L/h. Temperature in the fermentor and flash tank was measured using K type thermocouples (N. Brunswick Scientific Co.). A Cold trap, -25 °C working temperature (MA-055 model, Marconi laboratory equipment, Brazil), was used in the condensing system.

The yeast used was *Saccharomyces cerevisiae* obtained from an industrial fermentation plant. The medium used in the fermentation was sugarcane molasses containing about 77% of purity in sugar, diluted in water with to a concentration of 180 g / L of reducing sugars.

A stage of batch fermentation was initiated shortly after the addition of the inoculum. The objectives in this stage were to promote the total consumption of substrate and to reach a high biomass concentration before the second stage begins. The end of the batch fermentation stage was observed by the stabilization of turbidity and condensate volume readings. The continuous extractive fermentation was initiated by turning on the permeate pump of the filtration system. The removal of fluid with the permeate and purge pumps decreased the fermentation broth level into the fermentor activating the feeding pump (connected to an on-off level controller) that supplied fresh medium, by this action the level was kept constant throughout the fermentation.

The temperature in the fermentor was maintained constant at 33.25 ± 0.25 °C, the feed of substrate concentration was constant at 180 g/L throughout the fermentation process, while the dilution rate of the fermentor was maintained constant in the levels of 0.03 h^{-1} (33.33 h residence time), 0.10 h^{-1} (10 h residence time), 0.15 h^{-1} (6.67 h residence time), 0.20 h^{-1} (5 h residence time) and 0.35 h^{-1} (2.85 h residence time). Working with this last dilution rate it was possible to obtain a productivity of 25 g/L h, which is three times higher than the value obtained in the traditional fermentation process. The flash tank was operated with a feed flow rate of 200 L/h, vacuum pressure at 150 ± 40 mmHg, temperature at 33.8 ± 0.4 °C. The liquid remaining in the flash tank containing the fermentation broth with lower concentration of ethanol returned to the fermentor using a helical pump. The plant was simulated using the ANN models developed in Chapter 8, Part 3.

10.3. NEURAL NETWORK MODEL PREDICTIVE CONTROL (NNMPC) ALGORITHM

The rationale underlying MPC is to transform the control problem into an optimization one, so that at any sampling time instant a sequence of future control values is computed by solving a finite horizon optimal control problem. Then, only

the first element of the computed sequence is effectively used and the overall procedure is repeated at the next sampling time [12] [13].

In the NNMPC algorithm proposed in this work the convolution model used in the linear DMC algorithm is substituted by a non-linear internal model based on ANN, which is trained using data (input/output patterns) generated by the ANN model of the process. The predictive controller estimates a control action sequence (future inputs) that leads the controlled variables (outputs) to follow an optimal path to achieve a reference trajectory. This optimal path is determined by optimizing the following quadratic objective function [14] [15].

$$J = \sum_{i=1}^{NP} [y_{k+i}^c - ysp_k]^2 + \lambda^2 \sum_{i=1}^{NC} [\Delta u_{k+i-1}]^2 \quad (10.2)$$

Where:

ysp = set point, λ = weighting factor (suppression factor), which prevent large swings in the manipulated inputs, Δu = value of the future change in the manipulated variable that minimize the performance index J on the control horizon NC.

y^c = prediction made by the neural network model on a prediction horizon NP, corrected according to (10.7).

Certain restrictions on the changes of the manipulated variable are also considered, so the optimal control action sequence is obtained by solving a constrained quadratic programming (QP) problem at each sampling time.

The optimization problem is written as a quadratic program:

$$\min \phi(\Delta U) = \frac{1}{2} (\Delta U)^T H(\Delta U) + f^T \Delta U \quad (10.3)$$

Subject to:

$$\Lambda^T (\Delta U) = \eta \quad (10.4)$$

$$U_{\min} \leq \Delta U \leq U_{\max} \quad (10.5)$$

Where the matrices H and Λ , are related to tuning parameters [16].

$$\Delta U = [\Delta u_{k+i-1}]^T \quad (10.6)$$

(k = sampling instant; $i = 1, \dots, NC$)

The vectors η and ΔU are linear functions of the output prediction vector, $\hat{y}$, as well as of the tuning parameters.

Although the future control actions are calculated over a control horizon NC ($NC \leq NP$), only the first control action is utilized. The optimization algorithm was implemented in Matlab using the *quadprog* routine of the optimization toolbox.

The algorithm uses a corrected value for the prediction y^c in (10.2), incorporating a feedback strategy. At instant k , the predicted value of the output is compared to the measured value; the difference is used to correct the predicted value $\hat{y}$ in the NP moments ahead. For example, for the moment $k + i$:

$$y_{k+i}^c = \hat{y}_{k+i} + (y_k - \hat{y}_k) \quad (10.7)$$

10.4. RESULTS AND DISCUSSION

In this work two control alternatives using the model-based approach to control system design [9] were compared. The control objective is to regulate the ethanol concentration in the fermentor (E_f). In each alternative an empirical dynamic model of the process based on ANN was previously developed. This approach was used instead of developing a phenomenological model of the process, as the modeling of the flash tank has been shown to be complex [5] and lead to inaccurate results.

10.4.1. ALTERNATIVE I

10.4.1.1. Process modeling

After a preliminary study of the process, the following variables were considered as input variables to develop the ANN model : 1) Permeate flow rate (F_P) ; 2) Fermentor purge flow rate (F_{PU}); 3) Feed flow rate to the fermentor (F_0); 4) Fermentor biomass concentration (X) ; 5) Fermentor substrate concentration (S);

6) Residence time (t_r) or Fermentor dilution ($D=1/t_r$); 7) Fermentor Glycerol concentration (G); 8) Cell viability (C_v); 9) Fermentor outlet flow rate (F); 10) Fermentor temperature (T_{ferm}); 11) Flash tank inlet flow rate (F_{in}); 12) Flash tank temperature (T_{flash}); and 13) Flash tank pressure (P_{flash}). This model was identified as M13 and was described in Chapter 8, Part 3.

10.4.1.2. Dynamic behavior study

A procedure to determine the effect of input variables, on the output variables is by making changes or disturbances in the values of the inputs one by one and observing the changes in the outputs. In this work random disturbances were made on the selected inputs one by one by simulating the dynamic behavior of the process.

As inputs were utilized: a) purge flow rate, b) permeate flow rate, c) feed flow rate to the fermentor, and d) the fermentor dilution; and as output it was used the fermentor ethanol concentration, E_f .

The highest associations between variables were obtained in two cases: 1) between the purge flow rate, F_{pu} , and the ethanol concentration in the fermentor, E_f , with a correlation coefficient, $R = 0.99$ and a range of variation of the output as a result of the variations in the input, $r = 13.48$ g/L ($r = \max(\text{output}) - \min(\text{output})$) and 2) between the feed flow rate to the fermentor, F_o , and the ethanol concentration in the fermentor, E_f , with a correlation coefficient, $R = -0.91$ and a range of variation of the output as a result of the variations in the input, $r = 28.22$ g/L. Therefore F_{pu} and F_o could be appropriate variables to be considered as manipulated variables in order to control E_f .

Figure 10.2 show graphically the behavior of the inputs: purge flow rate, F_{pu} , and feed flow rate to the fermentor, F_o ; and the output ethanol concentration in the fermentor, E_f , with normalized values in the range $[0.1, 0.9]$.

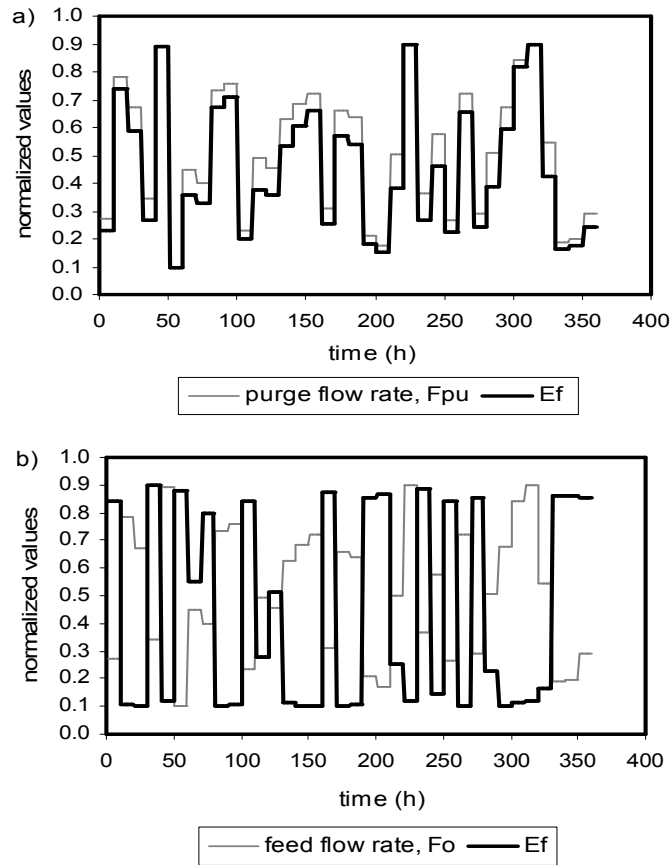


Figure 10.2. Normalized values of inputs and outputs:
a) Purge flow rate, F_{pu} , and ethanol concentration, E_f .
b) Feed flow rate to the fermentor, F_o , and Fermentor ethanol concentration, E_f .

From the results two controller structures were proposed: control structure 1 with the purge flow rate, F_{pu} , as manipulated variable, and control structure 2 with the feed flow rate to the fermentor, F_o , as manipulated variable. The controlled variable is the ethanol concentration in the fermentor, E_f . Each controller structure has an internal model.

10.4.1.3. Internal model training

The DMC algorithm uses an internal model (convolution model) to generate predictions of future control actions. In this work the internal model is a neural network with two inputs: the controlled and manipulated variables at the present

sampling instant, t , and one output: the controlled variable one step ahead, $(t + 1)$, as is shown in Figure 10.3.

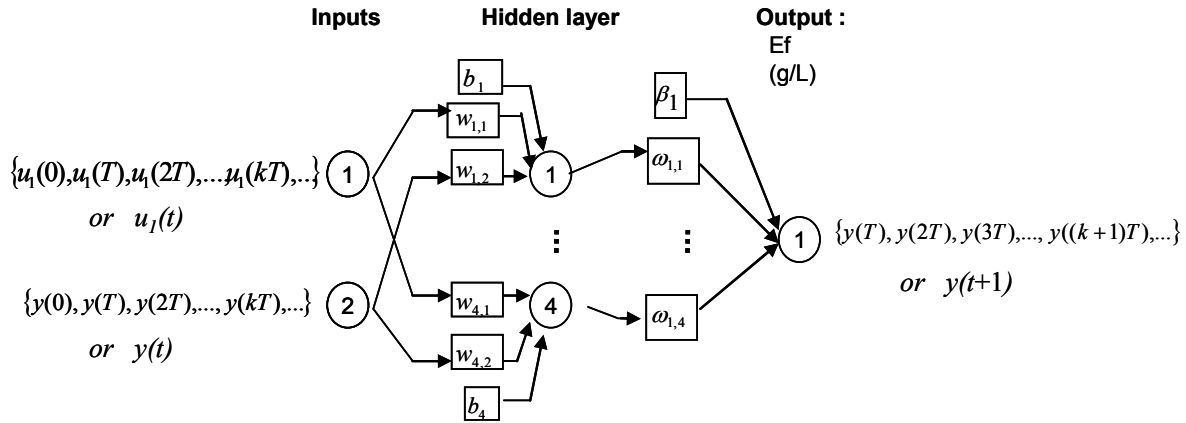


Figure 10.3 Controller ANN plant model topology

The training data (input/output patterns) were obtained by making 24 successive random disturbances on the inputs of the process model M13 (Chapter 8, Part 3) and determining the respective output. The interval between disturbances was chosen to ensure that the system reaches new steady states, as suggested by Santos and Biegler [17]. Both internal models based on ANN has a 2:4:1 topology with 4 neurons in the hidden layer. The internal model performance in describing the training data was determined using the following equation, which was suggested by Milton and Arnold [18]:

$$CORR = (1 - \frac{SEE}{S\tau\tau})100\% \quad (10.8)$$

$$\text{Where } SEE = \sum_{k=1}^N (y_e(k) - y(k))^2,$$

$$S\tau\tau = \sum_{k=1}^N (y_e(k) - \overline{y_e})^2,$$

$y_e(k)$ is an experimental output, $y(k)$ is the corresponding output of the neural model, $\overline{y_e}$ is the average of experimental outputs and N is the number of experimental data. The two internal models describe the training data with correlations ($CORR$) of: 99.05%, and 99.12%; respectively.

Another test suggested by various authors [15] [17] [19] and that was done with satisfactory results was to determine each model's ability to predict steady states of the process.

The practical experience from the present work has shown that additionally two more characteristics are important when testing the performance of the ANN internal model: 1) an accurate open loop response, and 2) a wide operation range.

10.4.1.4. Controller performance

The predictive controllers with nonlinear internal model based on neural networks were subjected to various tests, considering a sampling time of 30 minutes. The parameter values were determined by trial and error; and comparing the controller performance in different cases:

1) Control structure 1A (manipulated variable: purge flow rate): prediction horizon, NP =10, control horizon, NC=3, $f = 0.22$.

2) Control structure 2A (manipulated variable: feed flow rate to the fermentor): prediction horizon, NP =10, control horizon, NC=1, $f = 0.05$.

10.4.1.4.1. Regulator performance

In this test, changes were made in different inputs variables: F_o , F_{pu} , F_p , and D ; to simulate load disturbances and to observe the regulatory behavior of the controllers [20]. In each case disturbance variables were different from the manipulated variables. Two cases are mentioned to describe different characteristics on the response. Figure 10.5 shows the responses of the control structure 2A for disturbances of $\pm 20\%$ in the value of the purge flow rate, F_{pu} . Figure 10.4 shows the responses of the control structure 1A for disturbances of $\pm 5\%$ in the value of the feed flow rate, F_o .

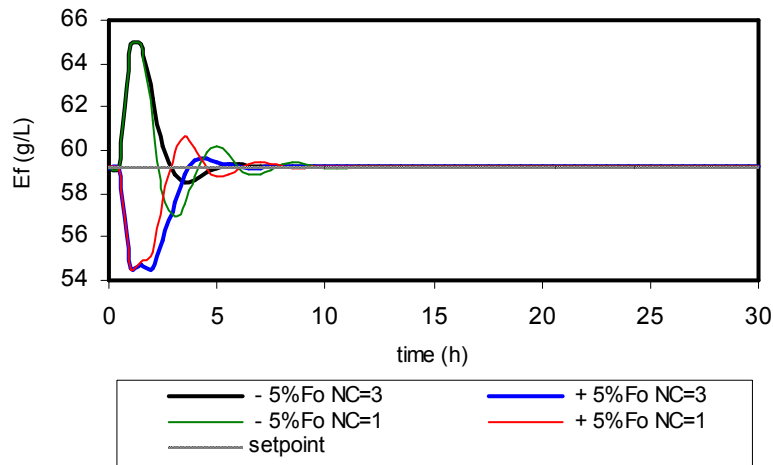


Figure 10.4. Control structure 1A: Response of the controlled variable, E_f , with time to perturbations of $\pm 5\%$ in the feed flow rate to the fermentor, F_o . Manipulated variable: purge flow rate, F_{pu} .

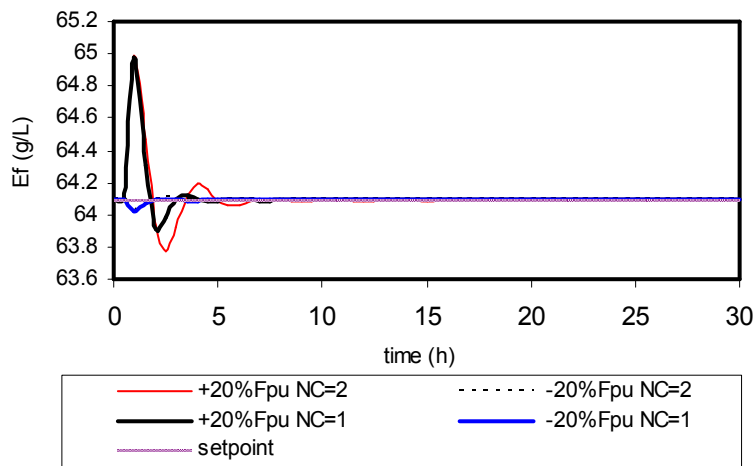


Figure 10.5. Control structure 2A: Response of the controlled variable, E_f , with time to perturbations of $\pm 20\%$ in the purge flow rate, F_{pu} . Manipulated variable: feed flow rate to the fermentor, F_o .

Figures 10.4 and 10.5 show that both NN MPC structures reject the disturbance and bring the system back to the steady state; therefore they have good regulatory performance. Three criteria were useful to describe the regulatory performance of the controllers: 1) the maximum deviation of the response from the reference trajectory, 2) the settling time, and 3) the IAE criterion (integral of the absolute value of the error), which provides controller settings that are between the most conservative settings, given by the ITAE criterion (integral of the time-weighted absolute error) and the most aggressive settings, given by the ISE

criterion (integral of the squared error) [9]. Figure 10.4 show that $NC=3$, is the best selection for control structure 1A, and Figure 10.5 show that $NC=1$, is the best selection for control structure 2A. The deviation (d) from the reference trajectory (setpoint) is higher for control structure 1A ($d=5.66$) than for control structure 2A ($d = 0.90$). The settling time is lower with control structure 2A than with 1A. Control structure 1A leads to a higher IAE value ($IAE=104.98$) than control structure 2A ($IAE=13.55$).

From the tests that were performed we can state that a carefully determination of the: “dynamic matrix” elements, suppression factor f , control horizon NC , and prediction horizon NP , is important to increase the performance of the NNMPC controllers. Although control structure 2A appears to have the best regulatory performance, it is necessary to do more tests using different disturbances of similar magnitude in order to have a better conclusion. In the present case there are strong interactions between input variables, e.g. feed flow rate to the fermentor, F_o , and permeate flow rate, F_p ; Fermentor dilution, D , and F_o ; that give difficulty in stating a definitive conclusion.

10.4.1.4.2. Servo performance

In this case successive changes were made in the value of the setpoint.

Figures 10.6 and 10.7 show the behavior of the predictive controllers with neural internal model to various changes in the reference trajectory (setpoint). It is noted that both control structures performed well, bringing the system back to the steady state after each change in the setpoint (represented in this case by ethanol concentration in the fermentor, E_f).

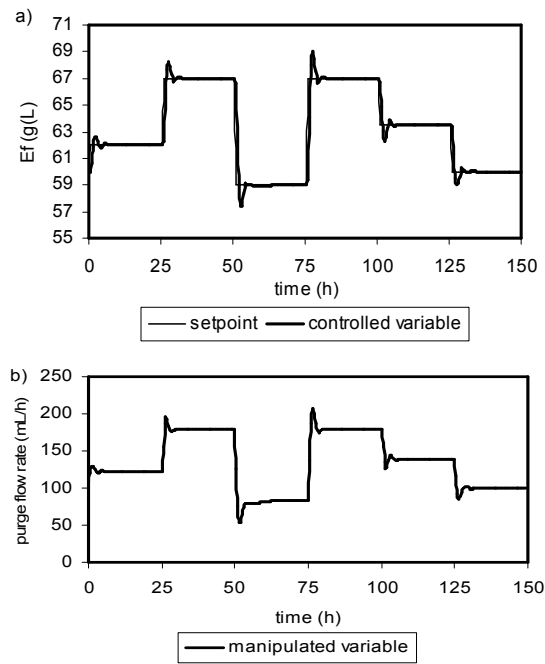


Figure 10.6. Servo performance of the control structure 1A:
a) Controlled variable F_{pu} and setpoint changes
b) Manipulated variable behavior (control actions).

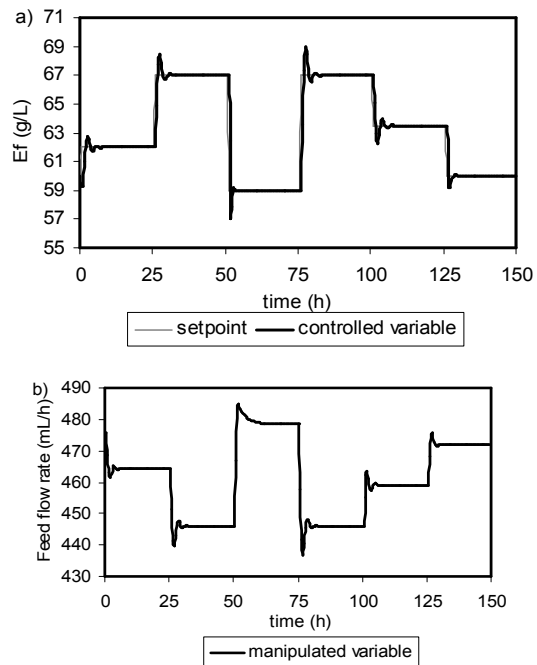


Figure 10.7. Servo behavior of control structure 2A:
a) Controlled variable F_o and changes in the setpoint.
b) Manipulated variable (control actions).

10.4.2. ALTERNATIVE II

In Chapter 8, Part 3, it was developed an ANN process model with eight input variables to explain the behavior of the fermentor ethanol concentration (output), the input variables are:

1) Permeate flow rate (F_P) ; 2) Fermentor purge flow rate (F_{PU}); 3) Feed flow rate to the fermentor (F_0); 4) Fermentor biomass concentration (X) ; 5) Fermentor substrate concentration (S); 6) Residence time (tr) or Fermentor dilution ($D=1/tr$); 7) Fermentor outlet flow rate (F); and 8) Flash tank pressure (P_{flash}).

In this chapter is used this model as a virtual plant to be controlled in alternative II. The ANN internal models of the controller are obtained by training directly with the experimental data set in order to give to the models the dynamics of the process. Each internal model of the controller has as inputs the controlled and the manipulated variables at the current time and as output the controlled variable a step ahead.

In alternative II, two control structures were tested, each one has an ANN internal model of the controller, in the first one (1B) the manipulated variable is the purge flow rate (F_{pu}) and in the second one (2B) the manipulated variable is the feed flow rate to the fermentor (F_0).

Each ANN internal model of the controller was selected after training by using the performance results obtained in prediction test with input-output patterns of the test data set, which were not used for training.

The performance criteria were the values of R^2 , mean squared error (mse), net parameter number, the average sum of the squares of the weights and biases, the correlation coefficient (COOR) determined by using the above mentioned equation suggested by Milton and Arnold [18] and the number of training epochs, the results are presented in Table 10.1. The selected model is identified as IM1-II.

Table 10.1 ANN internal model of the controller: values of the criteria used for selecting the model when is used the feed flow rate to the fermentor (F_o) as manipulated variable.

ANN internal Model	Topology	R^2	Mse	Parameter Number	Mswb	COOR	Epoch
IM1-II	2:2:1	0.99985	0.0034	9	11.873	99.971	150
IM2-II	2:2:1	0.99985	0.0034	9	11.888	99.971	237
IM3-II	2:3:1	0.99986	0.0033	13	0.5356	99.972	229
IM4-II	2:3:1	0.99986	0.0033	13	0.9383	99.972	455

The ANN internal model of the controller when the purge flow rate is used as manipulated variable is the same that was used in Chapter 9, the values of the criteria used to select this model are summarized as follows:

ANN internal Model	Topology	R^2	Mse	Parameter Number	Mswb	COOR	Epoch
IM1-I	2:3:1	0.99975	0.0030	13	1.49	99.97	255

The values of R^2 and mean squared error (mse) inform that bouth internal models fit very well the experimental data.

The parameter values were determined by trial and error; and comparing the controller performance in different cases:

Therefore the following control structures were tested in alternative B, the parameter values were determined by trial and error:

1) Control structure 1B (manipulated variable: purge flow rate): prediction horizon, $NP = 10$, control horizon, $NC = 1$, $f = 0.25$.

2) Control structure 2B (manipulated variable: feed flow rate to the fermentor): prediction horizon, $NP = 10$, control horizon, $NC = 1$, $f = 0.4$.

The test performed had as objective to assess the regulator and servo performance of the control structures these tests are widely used in the control literature to evaluate the performance of the controllers.

10.5.2.1 Regulator performance

In this test, changes were made in F_o and F_{pu} ; to simulate load disturbances and to observe the regulatory behavior of the controllers. In each case disturbance variables were different from the manipulated variables. Different cases are mentioned to describe characteristics on the response.

Figure 10.8 shows the responses of the control structure 1B for disturbances of $\pm 5\%$ in the value of the feed flow rate, F_o .

Figure 10.9 shows the responses of the control structure 2B for disturbances of $\pm 20\%$ in the value of the purge flow rate, F_{pu} .

Figures 10.8 and 10.9 show that the NNMPc control structures 1B and 2B reject the disturbance and lead the system back to the steady state, therefore they show a good regulatory performance.

Figures 10.8 a) and b), shows that the effect of the change in the value of the control horizon N_C in controller structure 1B from $N_C=1$ to $N_C=3$ is not significative because the responses are quite similar. The same behavior was observed in controller structure 2B. In the IAE values of controller structure 2B were $IAE=20.58$ when $N_C=1$ and $IAE=20.71$ when $N_C=3$ while parameter f is maintained in 0.25. Then in both alternatives it is better to use $N_C=1$.

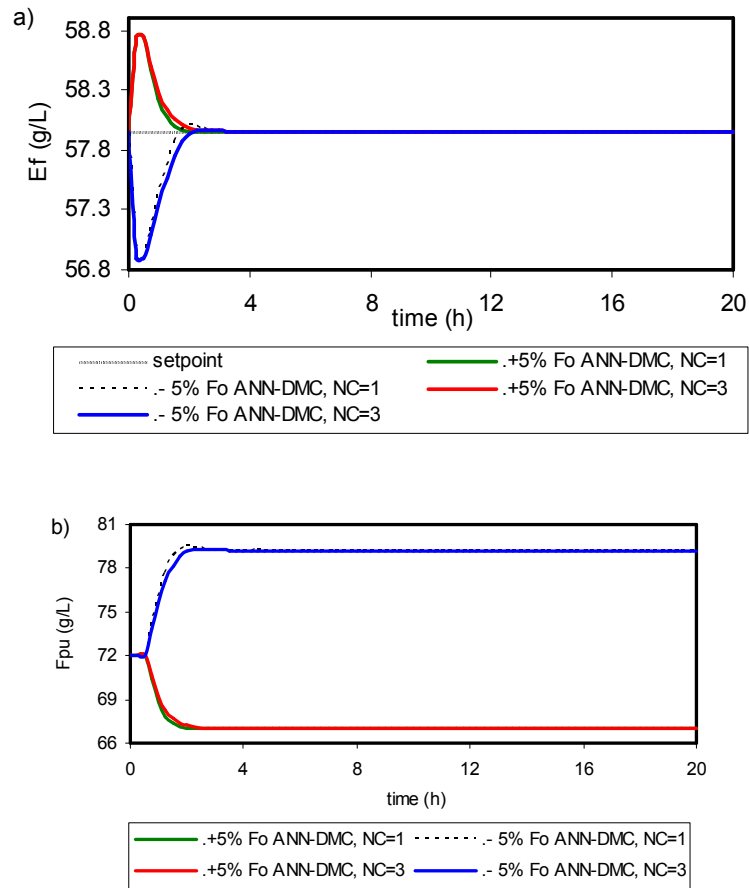


Figure 10.8. Control structure 1B: Response of the controlled variable, E_f , with time to perturbations of $\pm 5\%$ in the feed flow rate to the fermentor, F_o . Manipulated variable: purge flow rate, F_{pu} .

A comparison of the results between alternative I and alternative II show that the response in alternative II is slower than the one of alternative I however this difference maybe attributed to different factors one of them is the use of different values of f , in alternative II is used $f = 0.25$ while in alternative I is used $f = 0.05$

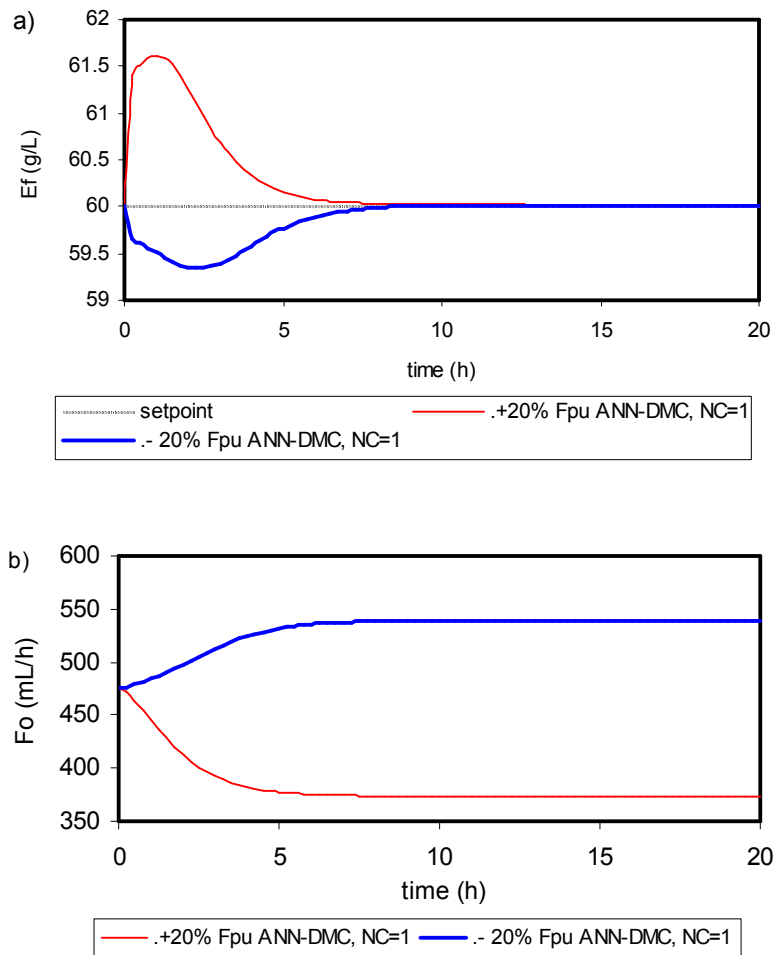


Figure 10.9. Control structure 2B: Response of the controlled variable, E_f , with time to perturbations of $\pm 20\%$ in the purge flow rate, F_{pu} . Manipulated variable: feed flow rate to the fermentor, F_o .

10.5.2.2. Servo performance

In this case successive changes were made in the value of the setpoint.

Figures 10.10 and 10.11 show the behavior of the predictive controllers with neural internal model in the controller to various changes in the reference trajectory (setpoint). It is noted that both control structures 1B and 2B performed well, bringing the system back to the steady state after each change in the setpoint (represented in this case by ethanol concentration in the fermentor, E_f).

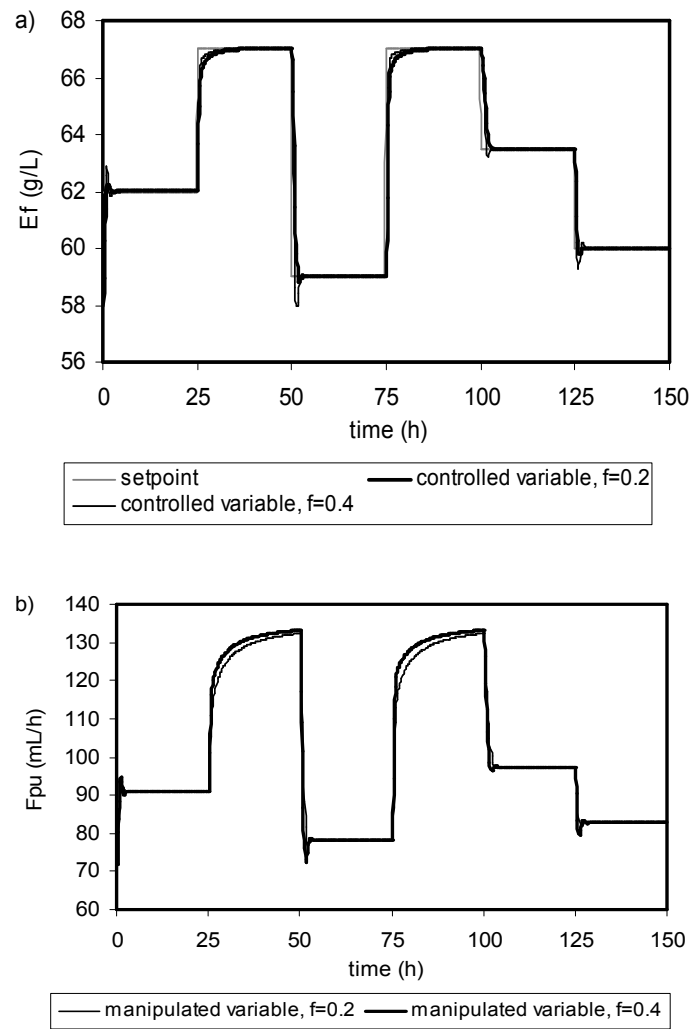


Figure 10.10. Servo performance of the control structure 1B:

- a) Controlled variable E_f and setpoint changes
- b) Manipulated variable behavior F_{pu} (control actions).

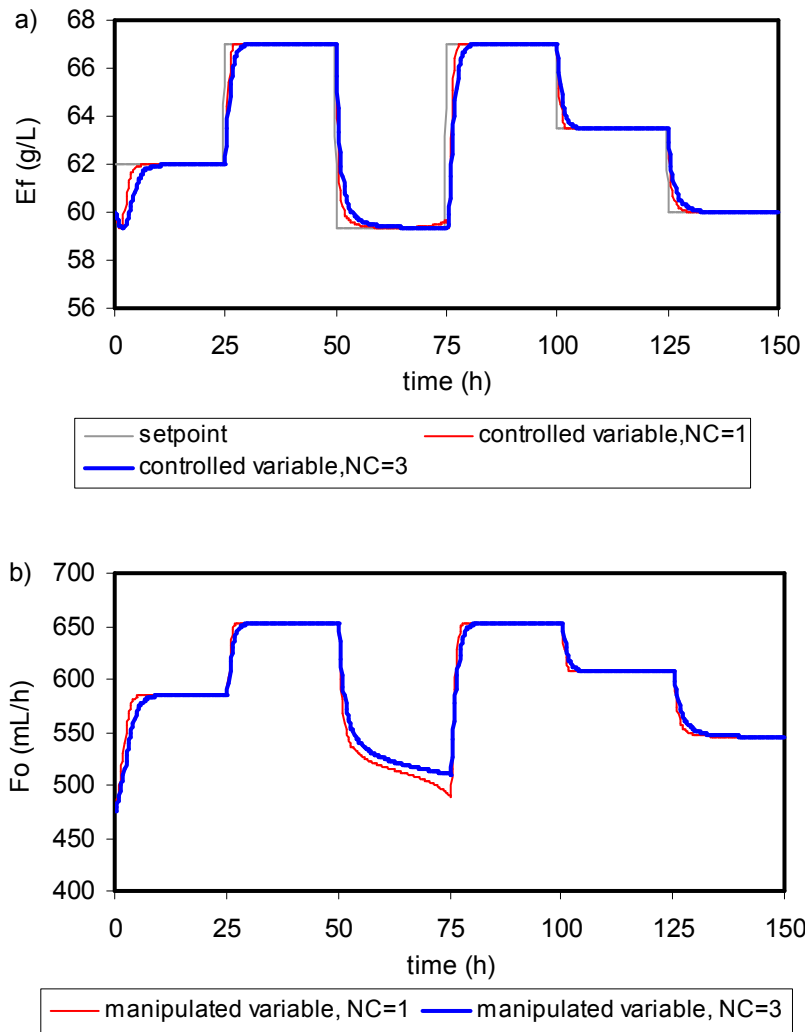


Figure 10.11. Servo performance of the control structure 2B:

- a) Controlled variable E_f and setpoint changes
- b) Manipulated variable behavior F_o (control actions).

A comparison of the results for servo performance between alternative I and alternative II show that the response of the plant model in alternative I has more overshoot than in case of alternative II, this indicates that the model responds quickly to reach the new reference trajectory, and maybe associated with the different training characteristics for both models, in alternative I the training sequence is: $\{u_i(0), u_i(T), u_i(2T), \dots, u_i(kT), \dots\}$ for inputs and $\{y(0), y(T), y(2T), \dots, y(kT), \dots\}$ for outputs, while in alternative II the training sequence

is $\{u_i(0), u_i(T), u_i(2T), \dots, u_i(kT), \dots\}$ for inputs and $\{y(T), y(2T), y(3T), \dots, y((k+1)T), \dots\}$ for outputs. In the literature Zakrzewska (2009) states that the quick response of the model can be useful for controlling a reactor in real time, because the model should be able to determine the system response to changes in the inlet variables almost instantly.

10.5.3. SUBSTITUTION OF THE ANN INTERNAL MODEL OF THE CONTROLLER FROM ALTERNATIVE II IN THE CONTROLLER OF ALTERNATIVE I

In this case the internal model developed for alternative II was used instead of the internal model trained for the plant model in alternative I. To train internal models using data provided by the plant model is a procedure suggested in the literature, mainly when there is a theoretical model of the process, however in the study case the dynamics of the process are implicit in the empirical data set, for this reason the empirical data set was used to train directly the internal model of the controller in alternative II.

The result of use the internal model of alternative II as internal model of the controller in alternative I is shown in Figures 10.12 and 10.13.

Figure 10.12 shows the regulator performance, before to obtain a good regulator performance with the control structure it was necessary to adjust the f parameter in this case best performance were obtained with $f = 0.12$, $NP=10$, $NC=1$.

Figure 10.13 shows the result for the servo performance when using values of $f = 0.18$, $NC=1$ and $NC = 3$, it is observed that there is a greater overshoot with $NC=3$ than with $NC=1$. The IAE values are $IAE= 38.2$ and $IAE= 42.40$ respectively.

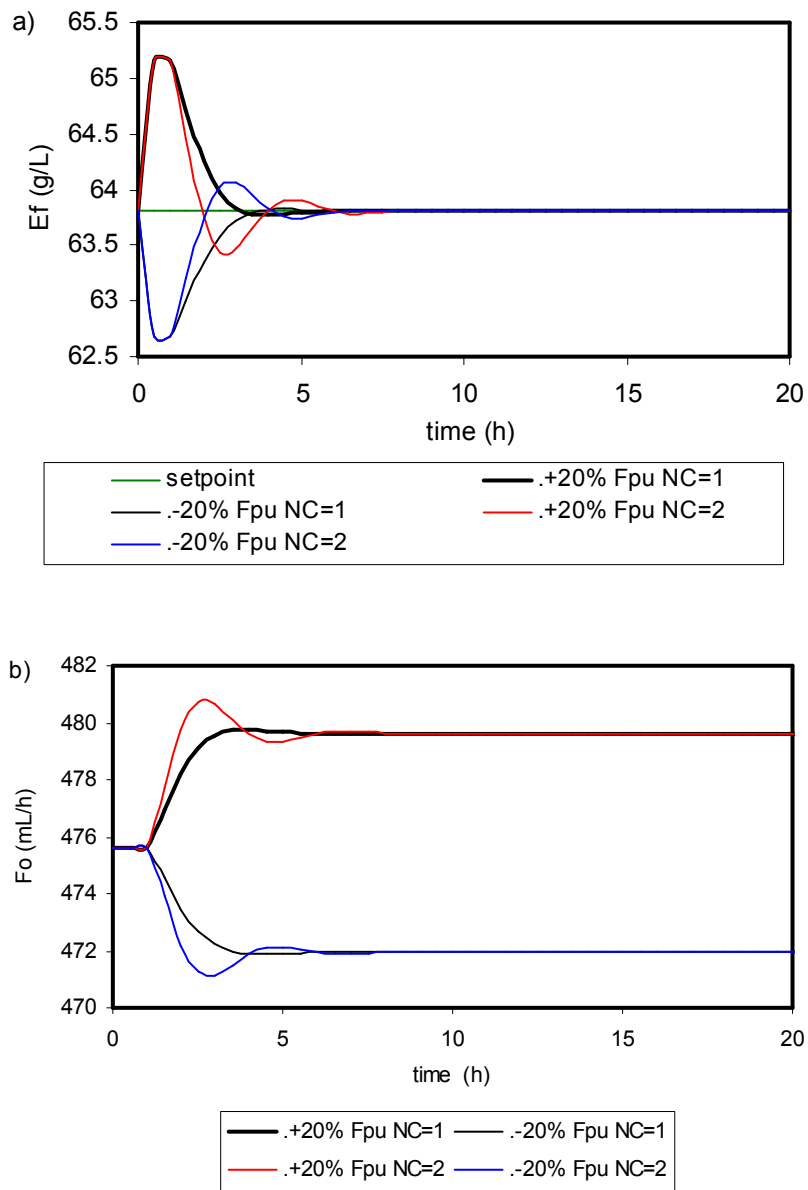


Figure10.12. Regulator performance of the control structure 2A from alternative I with controller internal model **IM1-II**: a) Response of the controlled variable, E_f , with time to perturbations of $\pm 20\%$ in the purge flow rate, F_{pu} . b) Manipulated variable: feed flow rate to the fermentor, F_o .

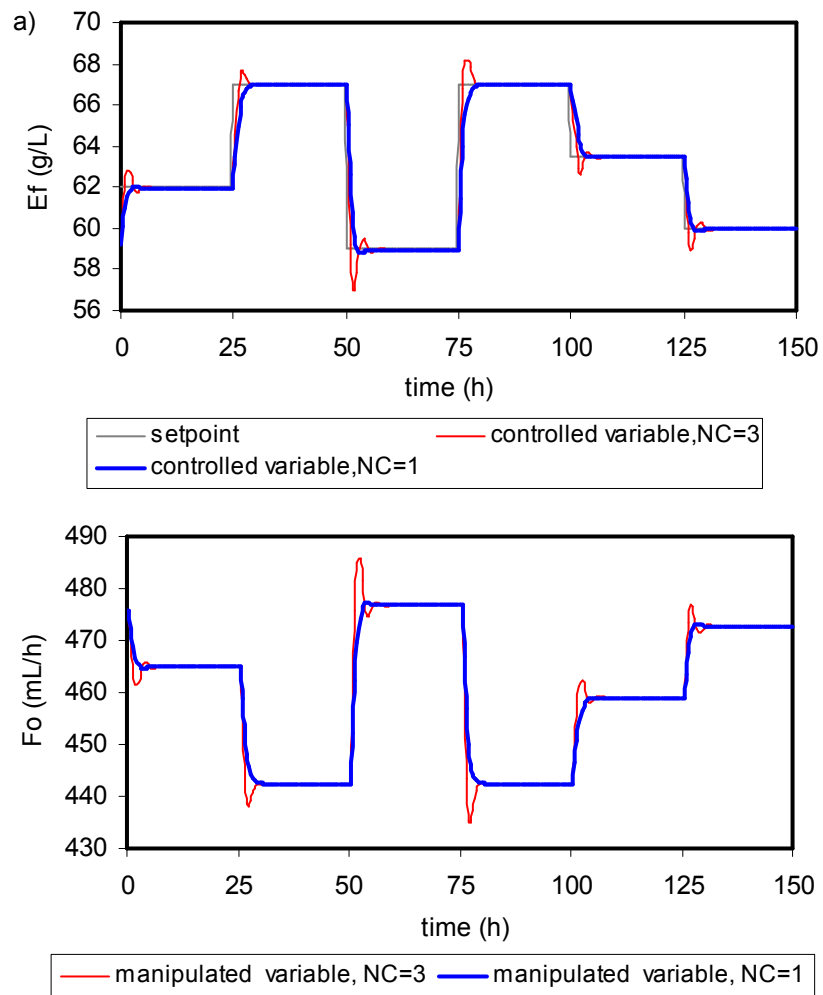


Figure 10.13. Servo performance of the control structure from alternative I with controller internal model IM1-II: a) Controlled variable E_f and setpoint changes, b) Manipulated variable behavior F_o (control actions).

Figures 10.12 and 10.13 show that the internal model developed by training directly the model with the empirical data set was used efficiently in the controller of alternative I, therefore this could be other useful procedure to develop ANN internal models for predictive controllers.

10.6. CONCLUSION

In this paper two ANN models with different number of variables for the plant to be controlled and four ANN internal models for predictive controller were developed. The non-linear neural networks model predictive controllers (NNMPC) were designed using the QDMC algorithm. It was shown that the non-linear models based on ANN can accurately identify the experimental behavior of the studied process and have a good performance as the internal model in predictive controller.

The use of an ANN model to simulate the dynamic behavior of the process allowed the design of the proposed control system even when an accurate phenomenological model was not available. The predictive performances of the controllers were tested with load disturbances and setpoint disturbances with good results, which demonstrate the high potential of the procedures used in this study for control system design in integrated reaction-separation processes.

10.7. REFERENCES

- [1] E. Smeets, M. Junginger, A. Faaij, A. Walter, and P. Dolzan. (2006, August). **Sustainability of Brazilian Bio-ethanol**. Utrecht University. Available: < <http://igitur-archive.library.uu.nl/chem/2007-0628-202408/NWS-E-2006-110.pdf> >
- [2] C.A. Cardona and O.J. Sánchez, "Fuel ethanol production: Process design trends and integration opportunities," *Bioresour. Technol.*, vol. 98, no. 12, pp. 2415-2457, Mar. 2007.
- [3] A. C. Costa, D. I. P. Atala, F. Maugeri, R. Maciel Filho, "Factorial design and simulation for the optimization and determination of control structures for an extractive alcoholic fermentation," *Process. Biochem.*, vol. 37, no. 2, pp.125-137, Mar. 2001.
- [4] D.I.P. Atala, "**Assembly, instrumentation, control and experimental development of an extractive fermentative process of ethanol production**," PhD dissertation, UNICAMP, SP, Brazil. 2004.

- [5] V.H. Álvarez, E. Ccopa Rivera, A. C. Costa, R. Maciel Filho, M.R. Wolf Maciel, M. Aznar, "Bioethanol Production Optimization: A Thermodynamic Analysis," *Appl. Biochem. Biotechnol.*, vol. 148, no.1-3, pp.141-149, Jan. 2008.
- [6] A. Ashoori, B. Moshiri, A. Khaki-Sedigh, M. Reza, "Optimal control of a nonlinear fed-batch fermentation process using model predictive approach," *J. Proc. Cont.*, vol. 19, pp. 1162-1173, Mar. 2009.
- [7] D. Dougherty, D. A. Cooper, "A practical multiple model adaptive strategy for single-loop MPC," *Control Eng. Practice*, vol. 11, pp. 141–159, 2003.
- [8] S. J. Qin and T.A. Badgwell, "A survey of industrial model predictive control technology," *Control Eng. Practice*, vol. 11, no. 7, pp. 733-764, 2003.
- [9] D.E. Seborg, T.F. Edgar and D.A. Mellichamp, **Process Dynamics and Control**, Second Edition, Hoboken, NJ: John Wiley & Sons, Inc., 2003.
- [10] C.L. Nascimento and T. Yoneyama, **Artificial Intelligence in Control and Automation**, Sao Paulo, Brazil: Blucher- FAPESP. 2008.
- [11] E.C. Rivera, D. I. P. Atala, F. Maugeri, A. C. Costa, R. Maciel Filho, "Development of real-time state estimators for reaction-separation processes: A continuous flash fermentation as a study case," *Chem. Eng. Proc.: Process intensification*, vol. 49, pp. 402-409, Mar. 2010.
- [12] R. Scattolini. "Architectures for distributed and hierarchical Model Predictive Control – A review," *J. Proc. Cont.*, vol. 19, pp. 723–731, Feb. 2009.
- [13] Y. Zheng, S. Li, X. Wang. "Distributed model predictive control for plant-wide hot-rolled strip laminar cooling process," *J. Proc. Cont.*, vol. 19, pp. 1427-1437, Apr. 2009.
- [14] W. Luyben. **Process modeling, simulation and control for chemical engineers**, Second Edition, New York, NY: McGraw-Hill, 1990.
- [15] A.C. Costa, R. Maciel Filho. "Non linear Predictive Control of a Three-Phase Catalytic Reactor ," *Can. J. Chem. Eng.*, vol. 81, pp. 1109-1118, Oct. 2003.
- [16] B.A. Ogunnaike and W.H. Ray. **Process dynamics, modeling, and control**, New York, NY: Oxford University Press, 1994.
- [17] L.O. Santos, L.T. Biegler. "A tool to analyze robust stability for model predictive controllers," *J. Proc. Cont.*, vol. 9, pp. 233-246, 1999.

- [18] J.S. Milton, J.C. Arnold. **Introduction to Probability and Statistics**, New York, NY: McGraw Hill, 1990.
- [19] A.C. Costa, L.A.C. Meleiro, R. Maciel Filho. "Non-linear predictive control of an extractive alcoholic fermentation process," *Process Biochem.*, vol. 38, no. 5, pp. 743-750, Dec. 2002.
- [20] A.R. Soeterboek, "**Predictive Control a unified approach**," PhD dissertation, Technische Universiteit Delft, Rotterdam, Holand, 1990.
- [21] Zakrzewska and Jaworski. An Empirical Model of a Multiphase Reactor Based on Artificial Neural Network. ICheaP-9 The ninth International Conference on Chemical & Process Engineering 10-13 May 2009 - Rome, Italy. <http://www.aidic.it/icheap9/webpapers/202Zakrzewska.pdf>

**CAPÍTULO 11. DECENTRALIZED NON-LINEAR PREDICTIVE
CONTROL FOR MIMO REACTION-SEPARATION PROCESSES: A
CONTINUOUS EXTRACTIVE FERMENTATION AS STUDY CASE**

Abstract

In this paper, a decentralized control system based on model predictive control (MPC) to be applied in the continuous extractive fermentation with vacuum flashing is proposed. The process has a highly complex non-linear dynamic. In the design approach, first, based on experimental data from the plant operation, an empirical multiple input multiple output (MIMO) model based on Artificial Neural Networks (ANN) is developed to be used as a virtual plant. Systematic step perturbations of inputs, Bristol's relative gain array method (RGA) and Morari resiliency index (MRI) are then used to obtain the best loop pairing.

The process industry's standard for MPC, the dynamic matrix control (DMC), with internal model based on ANN and quadratic programming optimization, is then used to develop the control structure.

Finally, the loop controllers are selected and independently designed based on data provided by the virtual plant. The performance of the proposed control system is verified by several tests of the servo and regulatory cases demonstrating the high potential of the design approach to develop decentralized control systems for MIMO processes when applied to the study case.

Key words: Artificial neural networks, continuous extractive fermentation process, non-linear predictive control.

11.1. INTRODUCTION

Environmental problems associated with the use of fossil fuels and the increase in their price have led to extensive research in order to replace them with biofuels. In this context, the use of bioethanol as a substitute of gasoline has been proved to be a cleaner alternative; therefore there is an increased demand and the need to increase the production of this biofuel. However, increasing the production of bioethanol involves several challenges; one of them is to improve the process efficiency. Cardona & Sanchez (2007) proposed as an important alternative the integration of reaction and separation processes, integrated schemes of this type

couple continuous fermentation with: vacuum flashing, gas stripping, filtration, perstraction, pervaporation, membrane distillation, liquid-liquid extraction.

The extractive fermentation process with vacuum flashing is a particularly interesting choice for the intensification of ethanol production because it is designed to remove ethanol continuously from the fermentation broth, reducing or neutralizing the inhibition of the cells by high ethanol concentrations and leading to higher productivities (COSTA *et al.*, 2001; ATALA, 2004; CARDONA & SANCHEZ, 2007). Other advantages are: the possibility of eliminating the need for heat exchangers in the fermentation stage (ATALA, 2004) and the potentiality of eliminating one distillation column from the process (RIVERA *et al.*, 2010a), which reduces costs in equipment and maintenance. However, the process presents highly complex dynamics and it is very complicated to develop a phenomenological mathematical model of the vapor-liquid equilibrium for the fermented broth system because it has a complex phase behavior. Besides, most of the substances being polar components, the very low concentration of many other components (apart from ethanol and water), called congeners, lead to difficulties in correlating and predicting the concentration of the distilled product in the continuous flash extractive fermentation (ÁLVAREZ *et al.*, 2008).

In this work a decentralized non-linear control system was proposed to control the extractive fermentation process with vacuum flashing. In order to develop the control algorithm an empirical process model was built with data stored in the computer based data acquisition system to be used as a virtual plant, avoiding the costs of performing experiments in the pilot plant. A selected MIMO process model based on Artificial Neural Network (ANN) was used to describe the two variables to be controlled: the ethanol concentration in the fermentor and the ethanol concentration in the condensed stream from the flash tank separator system. ANNs are described in the specialized literature as a rapid and accurate modeling technique for biotechnological processes (RIVERA *et al.*, 2010 a and b).

Analyses based on multiple criteria were performed to determine how the controlled and manipulated variables should be paired in the decentralized control

system. The loop controllers were designed using the Model Predictive Control (MPC) approach. MPC is an important advanced control technology for difficult multivariable control problems and has had remarkable success in multiple industrial applications, such as refining, petrochemical, chemical, paper and food industries (PICHÉ *et al.* 2000; QIN and Badgwell, 2003; SEBORG, *et al.*, 2004; Tatjewski, 2006). Furthermore, predictive controllers are recognized as robust and easy to tune (Soeterboek, 1990). In the present work the Dynamic Matrix Control (DMC) algorithm, a representative algorithm of the MPC technology with recognized impact on industrial process control (QIN and BADGWELL, 2003) was selected to develop the loop predictive controllers, in which were used MISO (multi-input single-output) internal models based on ANN.

The effectiveness of the proposed control system was verified by several tests performed to assess the servo and regulatory performances.

11.2. PROCESS DESCRIPCION

11.2.1. EXPERIMENTAL STAGE

The research experimental stage was conducted at the Bioprocess Engineering Laboratory of the State University of Campinas, Brazil (ATALA, 2006). The extractive fermentation process is integrated by: one reaction unit (fermentor), one cell recycle system (cross-flow microfiltration), one vacuum separation system (flash tank for ethanol-fermented broth separation and vacuum pump), two helical pumps, three peristaltic pumps, and one condensing unit. The total working volume is approximately 5 L.

A 3 liters "Bioflo III System" (New Brunswick Scientific Co., Inc., NJ, USA) bioreactor with PID (Proportional, Integral and Differential) control of temperature and stirring system was used as reaction unit.

A cross-flow microfiltration unit (Ceraflow model, Millipore Co.) with filter element made of high purity alumina, 0.22 mm pore and 0.0372 m² filtration area was used in the cell recycle system. The flask tank was a 2.5 liters (working volume) adapted Chemap reactor. The device to measure the input flow was an

electromagnetic flowmeter (IFS 400 KC model, signal converter IFC 090 model; Conaut, Brazil) with an operating range from 0 to 1200 L/h.

Figure 11.1 shows a diagram of the extractive fermentation process with vacuum flashing.

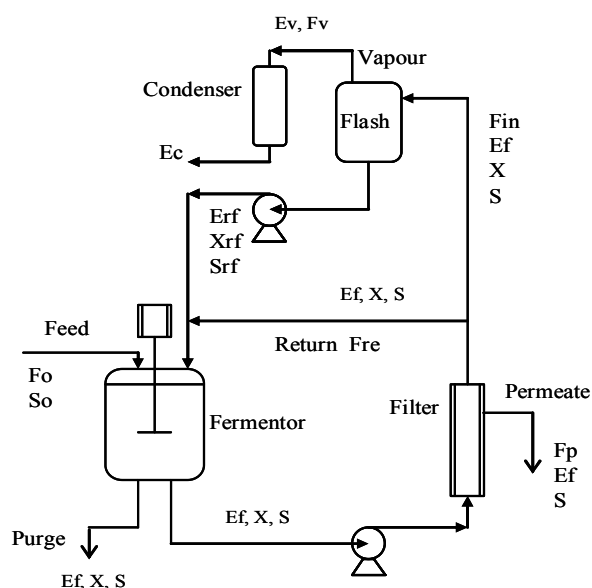


Figure 11.1. Diagram of the extractive fermentation process with vacuum flashing

Temperature in the fermentor and flash tank was measured using K type thermocouples (N. Brunswick Scientific Co.). A Cold trap, -25 °C working temperature (MA-055 model, Marconi laboratory equipment, Brazil) was used in the condensing system.

The yeast used was *Saccharomyces cerevisiae* obtained from an industrial fermentation plant. The medium used in the fermentation was sugarcane molasses containing about 77% of purity in sugar, diluted in water to a concentration of 180 g/L of reducing sugars.

The fermentation process was divided in two stages: extractive batch fermentation and continuous extractive fermentation; to facilitate the control and operation of the system. The stage of batch fermentation was initiated shortly after the addition of the inoculum. The objectives in this stage were to promote the total

consumption of substrate and to reach a high biomass concentration before the second stage begins. The end of the batch fermentation stage was observed by the stabilization of turbidity and condensate volume readings. The continuous extractive fermentation was initiated by turning on the permeate pump of the filtration system. The removals of fluid with permeate and purge pumps decreased the fermentation broth level in the fermentor activating the feeding pump (connected to an on-off level controller) that supply fresh medium, by this action the level was kept constant throughout the fermentation.

The temperature in the fermentor was maintained constant at 33.25 ± 0.25 °C, the feed substrate concentration was of 180 g/L throughout the fermentation process, while the dilution rate of the fermentor was maintained constant in the levels of 0.03 h^{-1} (33.33 h residence time), 0.10 h^{-1} (10 h residence time), 0.15 h^{-1} (6.67 h residence time), 0.20 h^{-1} (5 h residence time) and 0.35 h^{-1} (2.85 h residence time). Using the highest dilution rate productivity of 25 g/L h was obtained, which is three times higher than the value obtained in the traditional fermentation process (ATALA, 2006). The flash tank was operated with a feed flow rate of 200 L/h, vacuum pressure at 150 ± 40 mmHg and temperature of 33.8 ± 0.4 °C. The liquid remaining in the flash tank containing the fermentation broth with lower concentration of ethanol was returned to the fermentor using a helical pump.

The analytical monitoring was performed taken samples periodically from the fermentor for off line analysis of total sugar (sucrose, glucose and fructose), glycerol, ethanol, biomass, cell viability and cell volume. In the condensed stream the ethanol concentration and volume collected were analyzed. The concentrations of sugars, ethanol and glycerol were determined by HPLC (Varian 9010 model, Varian, Inc. Scientific Instruments, Hansey Way, USA) using an ion partition column (Aminex HPX-87H model, Bio-Rad Laboratories, Inc, CA, USA) at 30°C. Ultrapure water, obtained from Milli-Q purification system (Millipore Corporation, Concord Road, USA), containing Millipak membrane of 0.22 μm pore, was mixed with H_2SO_4 (pH 1.4) and used as the eluent at 0.7 mL/min. The standards were

mixed solutions of sucrose, glucose, fructose and ethanol at concentrations from 0.1 to 40 kg/m³. Cells concentration in dry weight was determined by gravimetric analysis. Turbidity was measured with a turbidity transmitter (FSC 402 Mettler Toledo Ingold Inc., USA); pH was measured using a glass pH-electrode (Cole-Parmer Instrument, London, England). On line measurements made in different process units were stored in computer using a data acquisition system with management software implemented in LabVIEW 6.1 (National Instruments, Austin TX, USA).

11.2.2. VARIABLE SELECTION

In this work a model-based approach to control system design (SEBORG, EDGAR & MELLICHAMP, 2004) was applied in a complex process. Therefore, after formulating the control objective: to regulate the ethanol concentration in the fermentor (E_f) and the ethanol concentration in the condensed stream from the flash tank (E_c), an empirical MIMO model of the process based on ANN was developed to be used as a virtual plant, so that the control system can be properly designed. This modeling alternative was used instead of a phenomenological model of the process, because the modeling of the flash tank has been shown to be complex (ÁLVAREZ, *et al.*, 2008) and lead to inaccurate results.

From many different models that have been proposed to describe the dynamics of ethanol fermentation (JARZEBSKI, MALINOWSKI, & GOMA, 1989; MONBOUQUETTE, 1992; ATALA, *et al.*, 2001; RIVERA *et al.*, 2010a) and because in the control problem it is necessary to consider flows as manipulated variables to reach an adequate control of concentrations, the following variables were initially considered as input variables to develop a plant model based on ANN: 1) Permeate flow rate (F_P); 2) Fermentor purge flow rate (F_{PU}); 3) Feed flow rate to the fermentor (F_0); 4) Fermentor outlet flow rate (F); 5) Fermentor biomass concentration (X); 6) Fermentor substrate concentration (S); 7) Residence time (tr) or Fermentor dilution ($D=1/tr$); 8) Fermentor Glycerol concentration (G); 9) Cell viability (C_v); 10) Fermentor temperature (T_{ferm}); 11) Flash tank inlet flow rate (F_{in});

12) Flash tank temperature (T_{flash}); and 13) Flash tank pressure (P_{flash}). While the fermentor ethanol concentration (E_f) and ethanol concentration in the condensed stream from the flash tank (E_c) were considered as the output variables.

A preliminary study was performed to determine the importance of the effect of each input variable on the fermentor ethanol concentration (BOZA-CONDORENA *et al.*, 2011). In this study the method of transfer functions was applied (BOZA-CONDORENA & COSTA, 2011). This method uses transfer functions expressed in terms of ANN interconnection weights and block diagrams to estimate the importance of the influence of the inputs on the output.

The results have shown that the permeate flow rate (F_p) and the fermentor purge flow rate (F_{pu}) had the highest effects on the fermentor ethanol concentration, and that the fermentor temperature (T_{ferm}), flash tank inlet flow rate (F_{in}), flash tank temperature (T_{flash}) had the lowest effects. The obtained results fit well with the plant dynamics, the values of the last three variables are approximately constant during all the experiment, because they are controlled variables in classical control loops. Other variable that is regulated by control loop is the flash tank pressure (P_{flash}). Because of this these variables were not considered to test by simulation the performance of additional predictive controller for the process plant. The experimental behavior of the selected variables is shown in Figure 11.2.

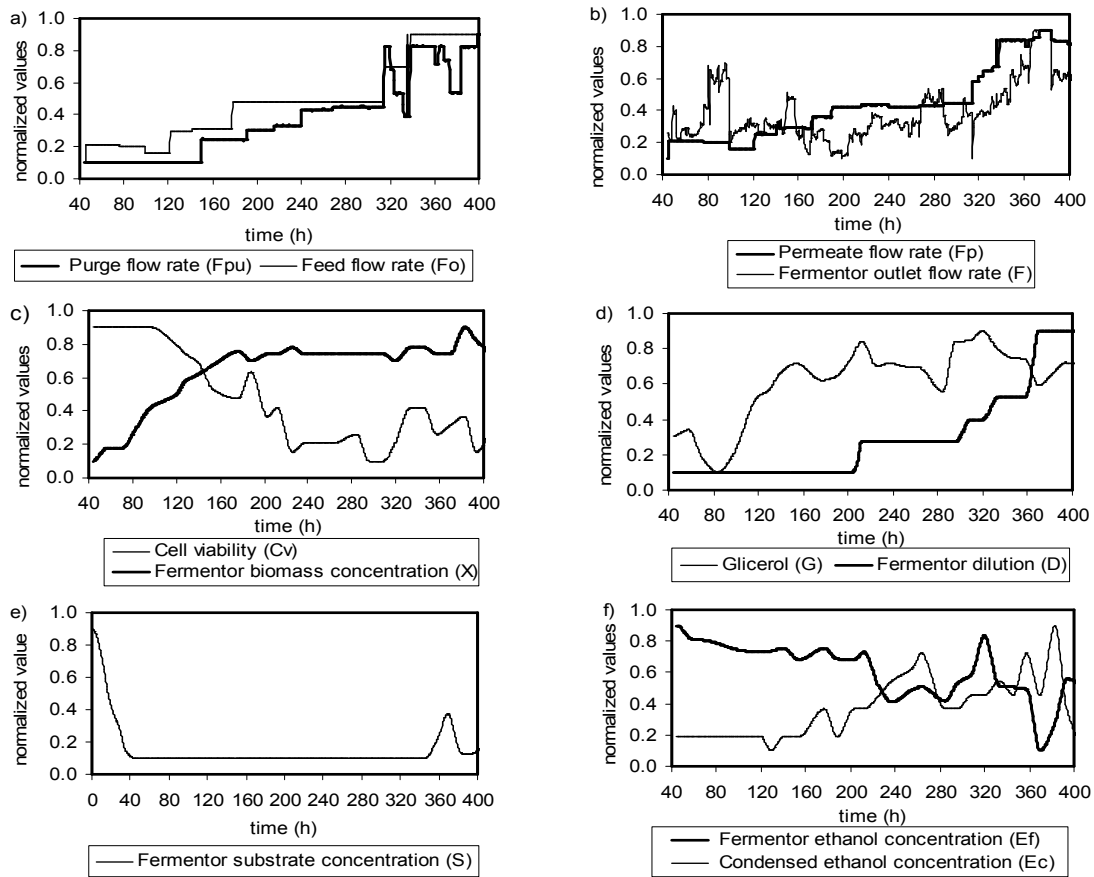


Figure 11.2. Experimental behavior of the selected variables in the continuous fermentation process with vacuum flashing.

In Figure 11.2 the data set was normalized in the [0.1, 0.9] range. Figure 11.2f shows the increasing non linear behavior of the ethanol concentration in the fermentor, E_f , and the condensed ethanol concentration, E_c , by effect of the changes introduced in the fermentor dilution rate, D in the operation period between 200 h and 400 h (Figure 11.2d), that finally led to an unstable process, which denotes a non satisfactory process control; therefore the necessity of applying other control strategies to improve the process performance and the operability of the plant.

11.3. MODELING AND CONTROL

11.3.1. ARTIFICIAL NEURAL NETWORK PROCESS MODELING

A two layer feedforward neural network was employed to model the process with a tan-sigmoid transfer function on the hidden layer and a linear transfer function on the output layer, because according to Cybenko's theorem, with this structure, the ANN models are able to approximate continuous functions at any desired level (NASCIMENTO & YONEYAMA, 2008; RIVERA *et al.*, 2010a). The Levenberg-Marquardt method for backpropagation training was used to train the ANN.

In the ANN structure selected the inputs, x_i , the weights that connect inputs to neurons in the hidden layer, W_{ji} , the weights that connect neurons in the hidden layer to neurons in the output layers, W_{kj} , the bias, b , the activation functions, f , and the value of the output variable E_k^O are related by the following equation:

$$E_k^O = f^2 \left(\sum_{j=1}^N W_{kj} f^1 \left(\sum_{i=1}^M W_{ji} x_i + b_j^1 \right) + b_k^2 \right) \quad (11.1)$$

Where: $j = 1, \dots, N$ (hidden layer neurons) ; $i = 1, \dots, M$ (inputs); $k = 1, \dots, K$ (output layer neurons). The weight W_{ji} , connects the i th input (x_i), and the j th neurons on the hidden layer (layer 1). The weight W_{kj} , connects the j th neuron of the hidden layer, and the k th neurons of the output layer (layer 2). Superscript 1 indicates layer 1. Superscript 2 indicates layer 2, Superscript 0 indicates output layer. The ethanol concentration estimated in the k th neuron of the output layer is represented by E_k^O .

The number of neurons in the hidden layers was determined by selecting the lowest mean square error (mse) when using the cross-validation technique, which is used to avoid model overfitting and to evaluate the performance of neural network by its ability to predict the elements of a validation data set which was not used when training the neural network.

In the present work the empirical MIMO model of the Plant based on ANN was developed using a representative data base containing 800 input/output patterns corresponding to 400 h of continuous flash fermentation at different dilution rates (0.1, 0.15, 0.20 and 0.35 h⁻¹); 600 input/output patterns (75%) were used for ANN training and the remaining 200 input/output patterns (25%) were used to validate the trained ANN.

11.3.2. NON-LINEAR MODEL PREDICTIVE CONTROL ALGORITHM

The rationale underlying MPC is to transform the control problem into an optimization one, so that at any sampling time instant a sequence of future control values is computed by solving a finite horizon optimal control problem. Then, only the first element of the computed sequence is effectively used and the overall procedure is repeated at the next sampling time (SCATTOLINI, 2009; ZHENG, Li, & WANG, 2009). A general scheme of the non linear predictive controller is shown in Figure 11.3.

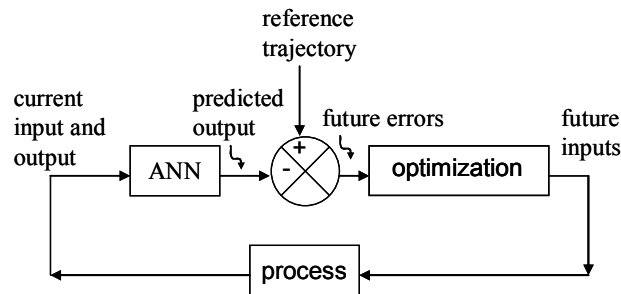


Figure 11.3 General scheme of the ANN-MPC predictive controller

In the ANN-MPC algorithm, the internal model used in the conventional formulation is substituted by an internal model based on ANN, which is trained using data (input/output patterns) generated by the ANN model of the process. The predictive controller estimates a control action sequence (future inputs) that leads the controlled variables (outputs) to follow an optimal path to achieve a reference trajectory. This optimal path is determined by optimizing an objective function.

In the present work the Quadratic Dynamic Matrix (QDMC) formulation is used (OGUNNAIKE & RAY, 1994, COSTA & MACIEL, 2003). In this alternative, the optimal control action sequence is obtained by solving a constrained quadratic programming (QP) problem at each sampling time and the objective function is represented by the following equation:

$$J = \sum_{i=1}^{H_p} [y_{k+i}^c - ysp_k]^2 + F^2 \sum_{i=1}^{H_c} [\Delta u_{k+i-1}]^2 \quad (11.2)$$

Where:

ysp = set point, F = weighting factor (suppression factor), which prevent large swings in the manipulated inputs,

Δu = value of the future change in the manipulated variable that minimize the performance index J on the control horizon H_c .

y^c = prediction made by the neural network model on a prediction horizon H_p , corrected according to Eq. (11.7).

Certain restrictions on the changes of the manipulated variable are also considered, so the optimization problem is written as a quadratic program:

$$\min \phi(\Delta U) = \frac{1}{2} (\Delta U)^T H (\Delta U) + f^T \Delta U \quad (11.3)$$

Subject to:

$$\Lambda^T (\Delta U) = \eta \quad (11.4)$$

$$U_{\min} \leq \Delta U \leq U_{\max} \quad (11.5)$$

Where the matrices H and Λ , are related to tuning parameters and f is a vector of coefficients of ΔU .

$$\Delta U = [\Delta u_{k+i-1}]^T \quad (11.6)$$

(k = sampling instant; $i = 1, \dots, H_c$)

The vectors η and ΔU are linear functions of the output prediction vector, $\hat{y}$, as well as of the tuning parameters. Although the future control actions are calculated over a control horizon H_c ($H_c \leq H_p$), only the first control action is utilized. The

optimization algorithm was implemented in Matlab using the *quadprog* routine of the optimization toolbox.

The algorithm uses a corrected value for the prediction y^c in Eq. (11.2), incorporating a feedback strategy. At instant k , the predicted value of the output is compared to the measured value; the difference is used to correct the predicted value $\hat{y}$ in the H_p moments ahead. For example, for the moment $k + i$:

$$y_{k+i}^c = \hat{y}_{k+i} + (y_k - \hat{y}_k) \quad (11.7)$$

Although the optimization is based on the control horizon (H_c), only the first control action is implemented and the optimization problem is solved at each sampling instant. Scheme of the decentralized control system based on non-linear Model Predictive Control is shown in Figure 11.4.

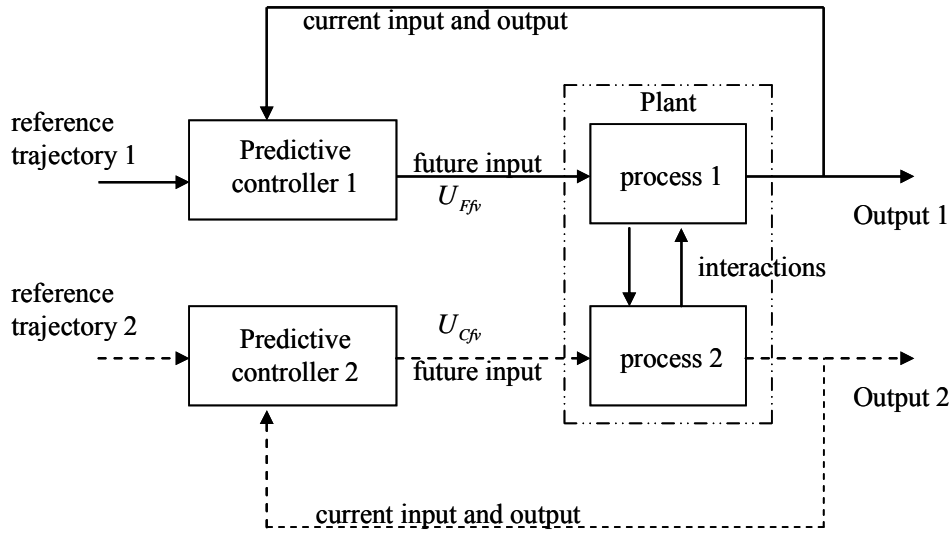


Figure 11.4. General scheme of decentralized control system.

11.4. RESULTS AND DISCUSSION

11.4.1. PROCESS MODELING

The design of controller based on ANN-MPC, started with the modeling of the process. Neural networks of configurations 9: N: 2, where N is the number of neurons in the hidden layer ($N \in [11, 20]$), were tested in order to select the best

ANN based model of the process. The lowest mse and R^2 values obtained when paring the experimental validation set with the ANN predicted values were used as indicators of the best model. A neural network with topology 9:15:2, with 15 neurons in the hidden layer led to the lowest mse and R^2 values, so it was selected as a model of the process. Figure 11.5 shows the ANN selected model topology. Figures 11.6a and 11.6b show the experimental and ANN predicted values for the fermentor ethanol concentration and the condensed ethanol concentration in the validation test.

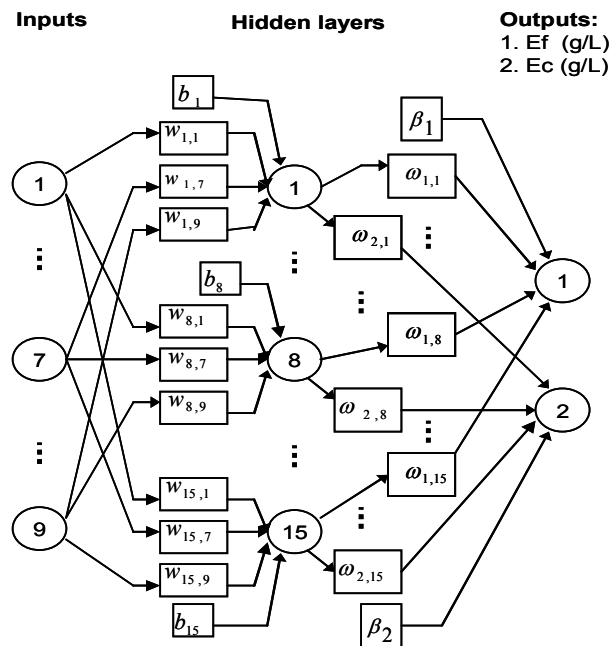


Figure 11.5. ANN model topology

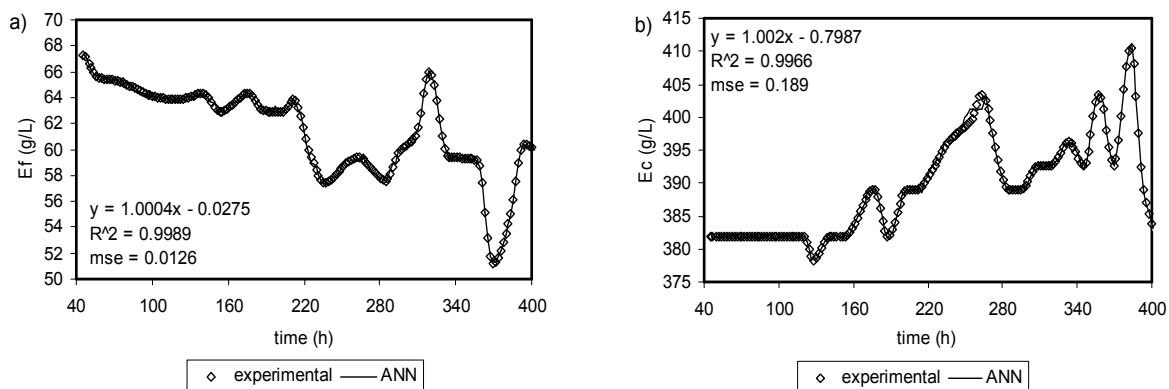


Figure 11. 6. Output of ANN process model and experimental values of the validation data set.
a) Fermentor ethanol concentration, E_f . b) Condensed ethanol concentration, E_c .

Figures 11.6a and 11.6b show that the selected ANN had a good performance in predicting experimental target values which were unknown by this net in the training step. The performance is slightly better when predicting the fermentor concentration. The model of the process should have the best accuracy while avoiding overfitting to obtain the best performance of the controllers based on MPC, since the internal ANN models of the controllers are trained with the information given by the process model.

11.4.2. DYNAMIC BEHAVIOR STUDY

A procedure to determine the effect of input variables on the output variables is by making changes or disturbances in the values of the inputs one by one and observing the changes in the outputs. In this work random disturbances were made on the selected inputs one by one by simulating the dynamic behavior of the process. The inputs were: a) purge flow rate, b) permeate flow rate, c) feed flow rate to the fermentor, and d) the fermentor dilution rate. The outputs were fermentor ethanol concentration, E_f , and the condensed ethanol concentration, E_c .

Two statistical parameters were of great importance to interpret the results, the correlation coefficient, r , and the output variation scope, OVS ($OVS = \max(\text{output}) - \min(\text{output})$). The first one was used to measure the association between input and output, and the second one to measure the impact magnitude of

the input change on the output. With this information it was straightforward to infer a preliminary result of variable pairing.

The highest associations between variables were obtained in two cases: 1) between the permeate flow rate, F_p , and the ethanol concentration in the fermentor, E_f , with a correlation coefficient, $r = 0.98$ and an output variation scope as a result of the disturbances in the input, $OVS = 12.60$ g/L ; and 2) between the purge flow rate, F_{pu} , and the condensed ethanol concentration, E_c , with a correlation coefficient, $r = 0.98$ and a range of variation of the output as a result of the variations in the input, $OVS = 32.58$ g/L. Therefore from this analysis, F_p could be an appropriate manipulated variable to control E_f and F_{pu} to control E_c .

Figure 11.7 show graphically the behavior of the inputs: permeate flow rate, F_p , and purge flow rate, F_{pu} ; and of the outputs: fermentor ethanol concentration, E_f , and condensed ethanol concentration, E_c , with normalized values in the range [0.1, 0.9].

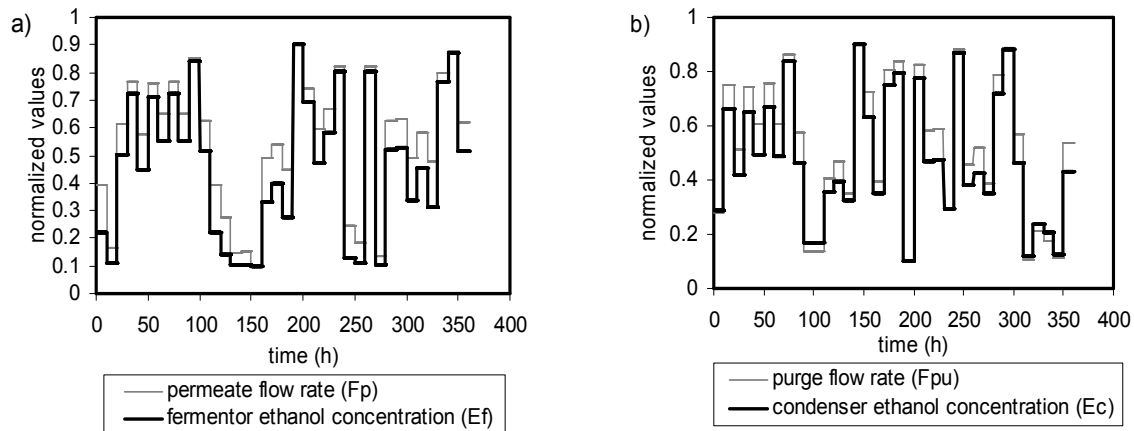


Figure 11.7. Effects of the input random step disturbances. a) Input: permeate flow rate, F_p . Output: fermentor ethanol concentration, E_f . b) Input: purge flow rate, F_{pu} . Output: condenser ethanol concentration, E_c .

11.4.3. PAIRING OF VARIABLES AND RESILIENCY

An analysis of how the controlled and manipulated variables should be paired using multiple criteria is important in a multiloop control scheme because an incorrect pairing can result in poor control system performance and reduced stability margins (SEBORG, EDGAR & MELLICHAMP, 2004). In this work additionally to the preliminary analysis, the Bristol's relative gain array method (RGA) and the Morari resiliency index (MRI) were used. MRI was computed to have an indication of the inherent controllability of the process for different choices of manipulated and controlled variables (MORARI, 1983). The algorithm was implemented in Matlab. The results are shown in Table 11.1.

Bristol's approach provides information about the process interaction and a recommendation concerning the most effective pairing of controlled and manipulated variables (SEBORG, EDGAR & MELLICHAMP, 2004). When using this approach, parameter λ defines all the RGA method information.

Table 11.1. Indicators of effective pairing of variables and resiliency.

Manipulated Variables	RGA	Pairing	MRI
1) F_{PU}, F_P	$\begin{bmatrix} 0.38 & 0.62 \\ 0.62 & 0.38 \end{bmatrix}$	$F_{PU} - E_C / F_P - E_F$	1.09
2) F_{PU}, F_0	$\begin{bmatrix} -1.78 & 2.78 \\ 2.78 & -1.78 \end{bmatrix}$	$F_{PU} - E_C / F_0 - E_F$	0.18
3) F_{PU}, D	$\begin{bmatrix} -0.24 & 1.24 \\ 1.24 & -0.24 \end{bmatrix}$	$F_{PU} - E_C / D - E_F$	0.78
4) F_P, F_0	$\begin{bmatrix} 0.51 & 0.49 \\ 0.49 & 0.51 \end{bmatrix}$	$F_P - E_F / F_0 - E_C$	1.20
5) F_P, D	$\begin{bmatrix} 0.24 & 0.76 \\ 0.76 & 0.24 \end{bmatrix}$	$F_P - E_C / D - E_F$	1.73
6) F_0, D	$\begin{bmatrix} -0.45 & 1.45 \\ 1.45 & -0.45 \end{bmatrix}$	$F_0 - E_C / D - E_F$	0.57

Table 11.1 shows that in cases 2, 3 and 6; $\lambda < 0$, then the control loops interact by trying to “fight each other” and the closed- system may become unstable (MC AVOY, 1983; SHINSKEY; 1996; SEBORG, EDGAR & MELLICHAMP, 2004). Also, the MRI values are the lowest and thus the process is less controllable or resilient. In these cases Niederlinski index is also negative, $NI < 0$, therefore the system will be unstable for any controller settings if integral action is used in all loops (LUYBEN, 1990). In case 4, $\lambda = 0.51$, then the interaction between the two control loops is the largest one (MC AVOY, 1983; SHINSKEY; 1996; SEBORG, EDGAR & MELLICHAMP, 2004).

Case 5, could be a good alternative, because $\lambda=0.75$ shows that there is an interaction of intermediate magnitude between the two control loops; $\lambda=0.75$ is the mean between $\lambda=0.5$ (highest interaction) and $\lambda=1$ (there is no interaction); $MRI=1.73$ means good controllability or resilience, but as was pointed out in the previous analysis the output variation scope for E_f is small in this case, $OVS=7.45$ g/L, which means that when 36 random step perturbations were made in D , the output E_f had a small variation range. Therefore the ANN-DMC controller structures using D as manipulated variable will have a limited control scope.

The remaining case 1 has also good controllability or resilience, as $MRI = 1.09$. The RGA method shows the existence of interactions between the two control loops, $\lambda = 0.38$ and $1-\lambda = 0.62$, so a pairing with 1-2/2-1 configuration should be utilized. Luyben (1990) point out that interaction is not necessarily bad, and in fact in some systems it helps to reject the effect of load disturbances. In case 1, in the first control loop, F_p is the manipulated variable and E_f is the controlled variable. The output variation scope, OVS value of E_f ($OVS=12.6$ g/L), is bigger than the OVS value of E_f in case 5 ($OVS=7.45$ g/L). Therefore, case 1 was selected for pairing the control loops, in this case in the second control loop, F_{pu} is manipulated variable and E_c is controlled variable. This result is the same obtained before with the dynamic behavior study.

11.4.4. INTERNAL MODEL TRAINING

In this study the controllers used MISO internal models based on ANN with two inputs: the controlled and manipulated variables at the present sampling instant, t , and one output: the controlled variable one step ahead ($t+1$). The training data (input/output patterns) were obtained by making 24 successive random disturbances in the inputs of the process model and then calculating the respective output. The interval between disturbances was chosen to ensure that the system reaches new steady states, as suggested by Santos and Biegler (1999). In both cases the internal model based on ANN has a 2:4:1 topology with 4 neurons in the hidden layer. The internal model performance in describing the training data was determined using the equation suggested by Milton y Arnold (1990):

$$CORR = (1 - \frac{SEE}{S\tau\tau})100\% \quad (11.8)$$

$$\text{Where } SEE = \sum_{k=1}^N (y_e(k) - y(k))^2,$$

$$S\tau\tau = \sum_{k=1}^N (y_e(k) - \overline{y_e})^2,$$

$y_e(k)$ is the target output, $y(k)$ is the corresponding output of the neural model, $\overline{y_e}$ is the average of target outputs and N is the number of experimental data.

The internal model of the predictive controller C1 for the first control loop, described the training data with a correlation ($CORR$) of 99.15 % and the internal model of the predictive controller C2 for the second control loop described the training data with a correlation ($CORR$) of 99.11%. Another test performed was the evaluation of the predictive ability of the internal models in describing the steady states of the process, as was suggested by various authors (FONSECA, 1998; SANTOS & BIEGLER, 2000; COSTA *et al.*, 2002; COSTA & MACIEL, 2003). Both internal models based on ANN were able to describe well the steady states of the process.

11.4.5. DESCENTRALIZED CONTROL SYSTEM PERFORMANCE

Rigorous tests were conducted to evaluate the servo and regulatory performances of the loop controllers of the proposed decentralized control system. In all the cases the sampling time was 30 minutes. The loop controllers are described in Table 11.2. The controller parameters were determined by trial and error.

Table 11.2. Control structure tested to implement the Decentralized Control System

Control structure Component	LOOP 1	LOOP 2
Controllers	Predictive controller C1	Predictive controller C2
ANN internal models	$E_{f_{t+1}} = f_{ann}(E_{f_t}, F_{p_t})$	$E_{C_{t+1}} = f_{ann}(E_{C_t}, F_{pu_t})$
Optimization	Successive quadratic programming	Successive quadratic programming
Parameters Setting	$F = 0.12$, NC=3, NP=10.	$F = 0.6$, NC=2, NP=10.

11.4.5.1. Regulator performance

In this test changes were made in the input variables of the MIMO process model that represent the virtual plant to simulate load disturbances and to observe the regulator performance of the predictive control structures in both control loops. Disturbances of $\pm 10\%$ in the feed flow rate to the fermentor, F_o , and of $\pm 10\%$ in the dilution rate, D , were performed simultaneously. The results of the regulator performance in both control loops are presented in Figures 11.8 and 11.9. The fermentor ethanol concentration is controlled in loop 1 and the condensed ethanol concentration is controlled in loop 2.

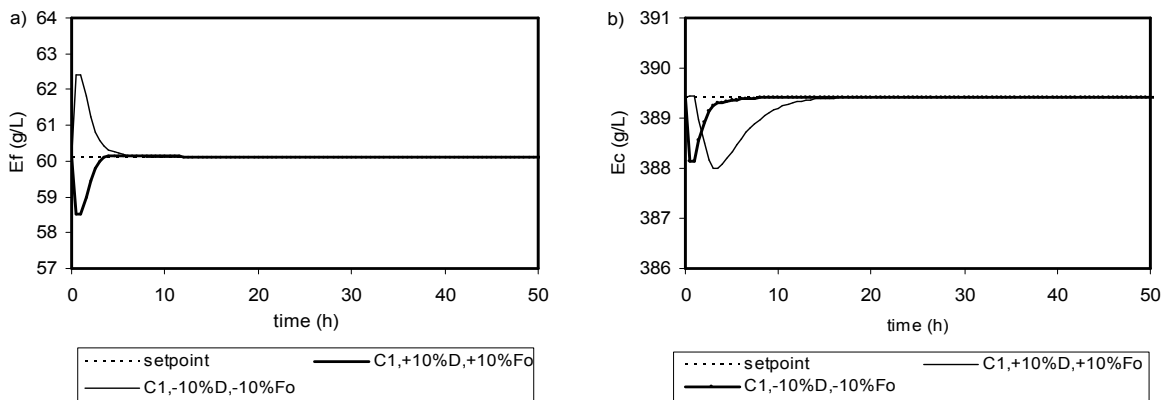


Figure 11.8. Controlled variables in the regulator performance to disturbances of $\pm 10\%$ in the feed flow rate to the fermentor, F_o , and of $\pm 10\%$ in the dilution rate, D . a) Control loop 1, controlled variable: Fermentor ethanol concentration, E_f . b) Control loop 2, controlled variable: Ethanol concentration in the condensed stream, E_c .

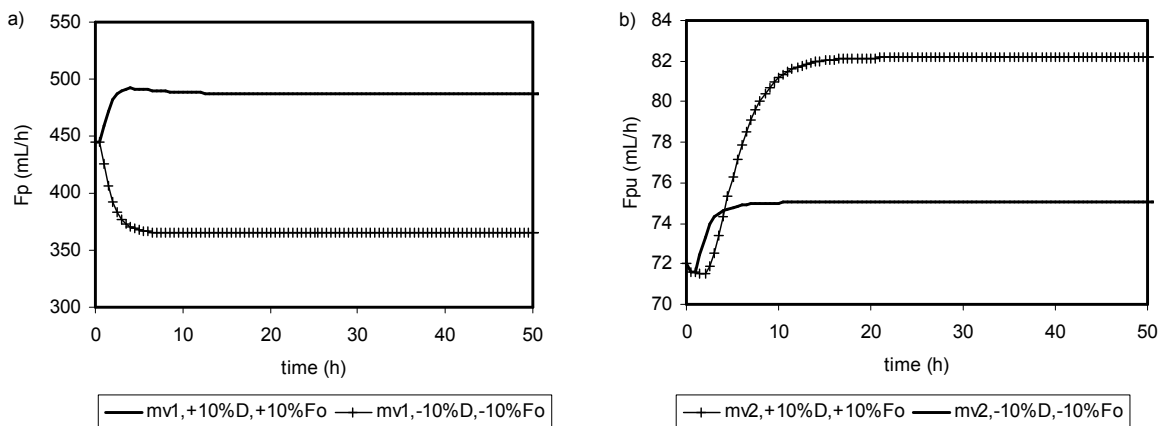


Figure 11.9. Manipulated variables in the regulator performance to disturbances of $\pm 10\%$ in the feed flow rate to the fermentor, F_o , and of $\pm 10\%$ in the dilution rate, D . a) Control loop 1, manipulated variable: Permeate flow rate, F_p . b) Control loop 2, manipulated variable: Fermentor purge flow rate, F_{pu} .

Figures 11.8 and 11.9 shows that decentralized control system had a good regulator behavior; both loop controllers rejected the load disturbances bringing the system back to the steady state.

11.4.5.2. Combined servo and regulator performances. Case A.

In this case changes were made on the setpoint value (E_f) in control loop 1 to evaluate the servo behavior in control loop 1 and the regulator behavior of the control loop 2, which receive the effects of the disturbances generated by the

successive changes in the setpoint in control loop 1. Figures 11.10 and 11.11 show the results.

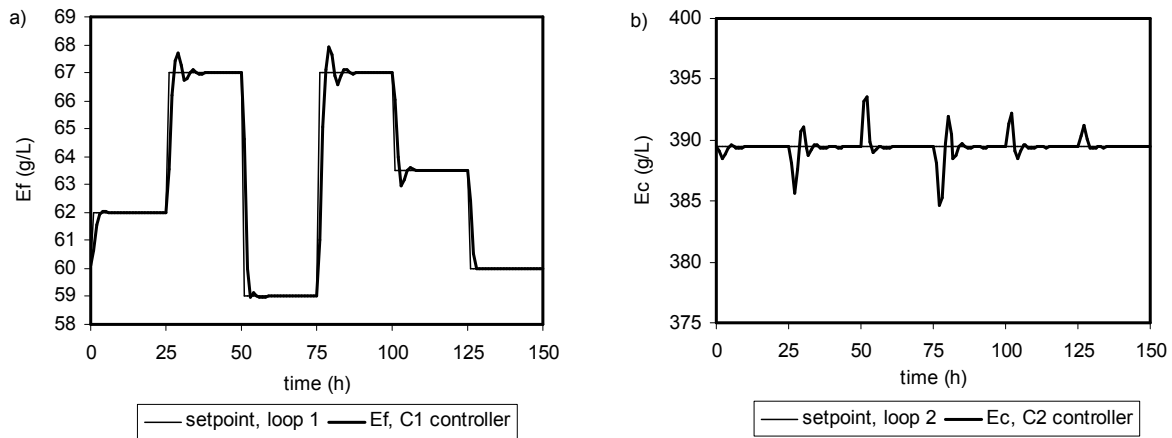


Figure 11.10. Controlled variables in the servo and regulator performance, case A: a) Servo performance: Control loop1, controlled variable: Fermentor ethanol concentration, Ef. b) Regulator performance: Control Loop 2, controlled variable: Ethanol concentration in the condensed stream, Ec.

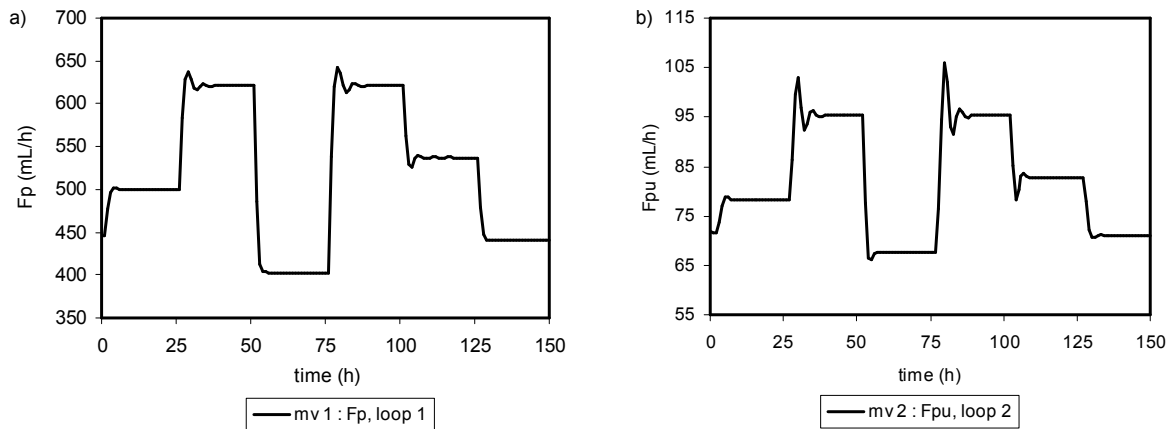


Figure 11.11. Manipulated variables in the servo and regulator performance, case A: a) Control loop 1, manipulated variable: Permeate flow rate, Fp. b) Control Loop 2, manipulated variable: Fermentor purge flow rate, Fpu.

Figures 11.10 and 11.11 show that the controllers in loops 1 and 2 have a good servo and regulator behavior. Both controllers in the decentralized control system are able to lead the process to the steady state, rejecting the different disturbances. Figure 11a shows that the response of the controlled variable has two bigger overshoots, the first one has a value of 1.05% and a decay ratio of 1/7, the second one has a value of 1.35% and a decay ratio of 1/7. In Figure 11b the

response of the controlled variable has overshoots whose values are less than or equal to 1.2%.

11.4.4.3. Combined servo and regulator performances. Case B.

In this case changes were made on the setpoint value (E_c) in the control loop 2, in order to test the servo behavior of the control loop 2 and the regulator behavior of the control loop 1, which is affected by disturbances caused by the successive changes in the setpoint in the control loop 1. Figures 11.12 and 11.13 show the results.

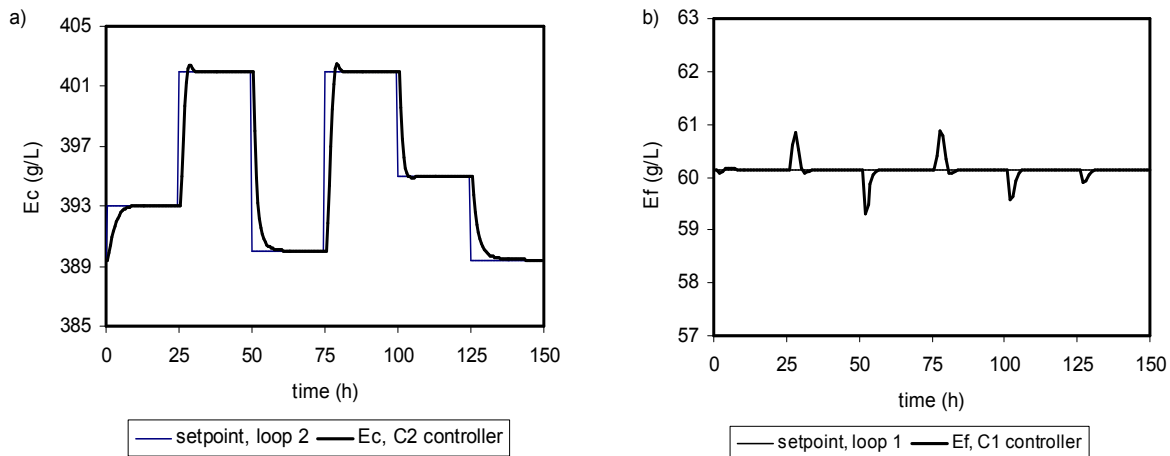


Figure 11.12. Controlled variables in the servo and regulator performance, case B: a) Servo performance: Control loop 2, controlled variable: Ethanol concentration in the condensed stream, E_c . b) Regulator performance: Control Loop 1, controlled variable: Fermentor ethanol concentration, E_f .

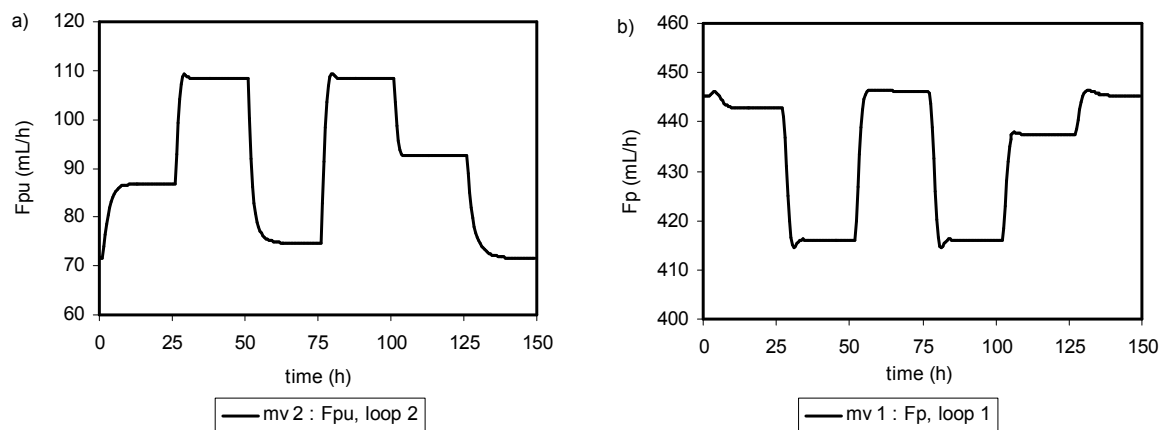


Figure 11.13. Manipulated variables in the servo and regulator performance, case B: a) Control loop 1, manipulated variable: Permeate flow rate, F_p . b) Control Loop 2, manipulated variable: Fermentor purge flow rate, F_{pu} .

Figures 11.12 and 11.13 show that the controllers in loops 1 and 2 have a good servo and regulator behavior, respectively. Both controllers in the decentralized control system, are able to lead the process to the steady state, rejecting the different disturbances. Figure 11.12a shows that the response of the controlled variable has small overshoots whose values are less than $\pm 1\%$ of the total change in the controlled variable. A similar behavior of the controlled system is shown in Figure 11.12b where the response of the controlled variable has overshoots with values less than 1.35 % of the total change in the controlled variable.

11.5. CONCLUSIONS

In the present study, a decentralized control system are proposed to be implemented in a reaction-separation process taking as study case the continuous fermentation with vacuum flashing, in which the overall system is divided in two interacting subsystems, each one with a control loop in wich the controller is designed using the MPC approach.

First it was developed an empirical MIMO model of the process based on ANN, with experimental data from a pilot plant, which was tested using different fermentor dilution rates in a working time period of 400 h. Second, it was realized a selection of variable pairing. Third, it was designed the descentralized control system developing MPC loop controllers with ANN internal models, which were trained with information given by the ANN MIMO model of the process. Fourth, the performances of the loop controllers in the decentralized control system were tested. The results show that the decentralized control structure with ANN internal models in the loop controllers and successive quadratic programming in the optimization of the objective function had good servo and regulator performances. Therefore, the decentralized control structure may provide an alternative of valuable application potential to improve the control of the ethanol concentration in the continuous extractive fermentation with vacuum flashing.

Acknowledgments

The authors acknowledge CAPES and Fapesp for financial support.

11.6. REFERENCES

1. Álvarez V.H.; Ccopa Rivera, E.; Costa, A. C.; Maciel Filho, R.; Wolf Maciel M.R.; Aznar, M. *Appl. Biochem. Biotechnol.* 2008, 148, 141-149.
2. Atala, D. I. P.; Costa, A. C.; Maciel Filho, R.; Maugeri, F. (2001). Kinetics of ethanol fermentation with high biomass concentration considering the effect of temperature. *Appl. Biochem. Biotechnol.* 91-93, 353 - 365.
3. Costa, A.C.; Meleiro, L.A.C.; Maciel Filho, R. Non-linear predictive control of an extractive alcoholic fermentation process. *Process Biochem.* 2002, 38(5), 743-750.
4. Costa, A.C.; Maciel Filho, R. Non-linear predictive control of a three-phase catalytic reactor. *Can. J. Chem Eng.* 2003, 81, 1109-1118.
5. Cardona, CA, Sánchez, OJ. (2007). Fuel ethanol production: process design trends and integration opportunities. *Bioresour. Technol.* 98, 2415-2457.
6. Fonseca, E.F. (1998). *Predictive control based in neural networks of continuous alcoholic fermentation*; (M.S. Thesis) UFPE, Recife - PE. 1998.
7. Jarzebski, A. B., Malinowski, J. J., and Goma, G. (1989). Modeling of ethanol fermentation at high yeast concentrations. *Biotechnol. Bioeng.* 34, 1225–1230.
8. Luyben, W. *Process modeling, simulation and control for chemical engineers*; Second Edition; McGraw-Hill: New York, NY, 1990, pp 287.
9. Mc Avoy, T.J., *Interaction Analysis*, ISA, Research Triangle Park, NC, 1983.
10. Milton, J.S.; Arnold, J.C. *Introduction to Probability and Statistics*; McGraw Hill, New York, NY, 1990.
11. Monbouquette, H. G. (1992). Modeling high-biomass-density cell recycle fermentors. *Biotechnol. Bioeng.* 39, 498–503.
12. Morari, M. (1983). Design of resilient processing plants-III. A general framework for the assessment of dynamic resilience. *Chem. Eng. Sci.*, 38(11), 1881-1891.

13. Ogunnaike, B.A. and Ray, W.H. *Process dynamics, modeling, and control*, New York, NY: Oxford University Press, 1994.
14. Piché, S.; Sayyar-Rodsari, B.; Johnson, D.; Gerules, M. *IEEE Contr. Syst. Mag.* 2000, 20(3), 53-62.
15. Rivera, E.C.; Atala, D. I. P.; Maugeri, F.; Costa, A. C.; Maciel Filho, R.(2010). Development of real-time state estimators for reaction-separation processes: A continuous flash fermentation as a study case. *Chem. Eng. Process.* 49, 402-409.
16. Rivera, E.C.; Rabelo, S.C.; Reis Garcia, D.; Maciel Filho, R.; Costa, A. C. *J. Chem. Technol. Biotechnol.* 2010, 85, 983 - 992.
17. Qin, S.J. and Badgwell, T.A. (2003). A Survey of Industrial Model Predictive Control Technology, *Control Eng. Pract.*, 11,733.
18. Santos, L.O.; Biegler, L.T. A tool to analyze robust stability for model predictive controllers. *J. Proc. Cont.* 1999, 9, 233-246.
19. Scattolini, R. (2009). Architectures for distributed and hierarchical Model Predictive Control – A review. *J. Proc. Cont.*,19, 723–731.
20. Seborg, D.E., Edgar, T.F., & Mellichamp, D.A. (2004). *Process Dynamics and Control*, (2nd ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.
21. Shinskey, F.G., *Process Control Systems*, 4th. ed., McGraw-Hill, New York, 1996.
22. Soeterboek, AR. (1990). Predictive Control a unified approach. PhD tesis.Technische Universiteit Delft, Rotterdam, Holand.
23. Tatjewski, P.; Lawrynczuk, M. *Int. J. Appl. Math. Comput. Sci.* 2006,16 (1), 7–26.
24. Zheng, Y; Li, S ; Wang, X. (2009) Distributed model predictive control for plant-wide hot-rolled strip laminar cooling process, *J. Proc. Cont.*, 19, 1427-1437.

CAPÍTULO 12. CONCLUSÕES GERAIS

Foi realizado com sucesso o desenvolvimento de uma abordagem de projeto de sistemas de controle preditivo com modelos empíricos de redes neurais, obtidos utilizando procedimentos de identificação de processos e dados experimentais da operação de um processo fermentativo extrativo de produção de bioetanol.

A revisão bibliográfica permitiu construir um marco de referência teórico atualizado sobre quatro conteúdos conceituais que são centrais no presente trabalho: a produção de bioetanol, as redes neurais artificiais, a modelagem e identificação de processos e o controle preditivo com modelo. Essa referência teórica foi fundamental na seleção de alternativas procedimentais na pesquisa desenvolvida.

Na produção de bioetanol foram mostradas as características e limitações no desenvolvimento de modelos teórico fenomenológicos para o processo fermentativo extrativo de produção de bioetanol com unidade de separação flash, que levam à inexistência de um modelo fenomenológico preciso para esse processo passando a utilizar no presente trabalho os volumosos conjuntos de dados experimentais existentes da operação do processo no grupo de pesquisa para desenvolver o projeto de sistemas de controle por modelagem empírica de processos.

A necessidade de utilizar o conhecimento implícito existente nos conjuntos de dados experimentais para o projeto de sistemas de controle orientou a pesquisa para dar uma ênfase no uso das redes neurais artificiais para identificação de processos, a revisão teórica permitiu utilizar técnicas para melhorar a velocidade de aprendizagem do algoritmo de retropropagação e para evitar o *overtraining* e o *overfitting* que são dois fenômenos que afetam a

capacidade de generalização das redes neurais. No Capítulo 6, a superfície de resposta construída com redes neurais foi explorada com uso de algoritmos genéticos e algoritmos de computação evolutiva obtendo o máximo global com bons resultados. Embora essas técnicas de otimização não tenham sido utilizadas depois para o projeto de controladores por existir outras técnicas menos demoradas e que permitem uma inclusão mais fácil de restrições, mostraram a boa qualidade dos modelos empíricos construídos com redes neurais.

O estudo das variáveis do processo permitiu explicar o comportamento dinâmico mostrado no período de operação experimental e a modelagem empírica com redes neurais do processo fermentativo extrativo de produção de bioetanol, possibilitou construir modelos para o monitoramento das concentrações de etanol utilizando só variáveis com determinação *on line*, com precisão similar à de modelos que utilizam variáveis com determinação *off line* no laboratório.

A modelagem empírica de processo permitiu também verificar que as redes neurais permitem obter modelos de processo mesmo em presença de ruído no sinal. No entanto, foi demonstrado que os modelos de redes neurais que podem ser utilizados como *software sensors* melhoram notoriamente a sua precisão quando é eliminado o ruído dos sinais.

O método das funções de transferência desenvolvido no presente trabalho, para a interpretação da informação contida nos modelos empíricos de redes neurais e o estudo das relações causa-efeito no processo fermentativo extrativo de produção de bioetanol, permitiu identificar e diferenciar satisfatoriamente nos dados experimentais as variáveis mais importantes e menos importantes que têm efeito sobre as concentrações de etanol, sendo que as mais importantes foram utilizadas com bons resultados como variáveis manipuláveis nos sistemas de controle e, que pode-se prescindir das menos importantes para reduzir o espaço de entradas.

Foram desenvolvidos com sucesso algoritmos baseados na integração da modelagem empírica de processos com redes neurais e as técnicas de controle baseadas em controle preditivo com modelo (MPC). Nos sistemas preditivos de controle foram utilizados modelos internos preditivos de redes neurais e procedimentos de otimização linear e não linear, para calcular as ações de controle futuras. Os controladores desenvolvidos SISO e MIMO; DMC e QDMC, possibilitaram obter bons resultados no controle das concentrações de etanol no processo fermentativo, extrativo e contínuo de produção de bioetanol.

Na implementação da abordagem de projeto de sistemas de controle para o processo de produção de bioetanol considerado como estudo de caso, foram também importantes: o uso de técnicas específicas, tais como a análise causal dos efeitos das entradas nas saídas, fazendo perturbações nas entradas uma por uma e observando as mudanças nas saídas, o estudo das características preditivas dos modelos nos estados estacionários do processo para projetar adequadamente o sistemas de controle, e o uso de vários critérios para emparelhamento de variáveis nas malhas de controle. Com o uso dessas técnicas, e uma determinação precisa dos parâmetros dos controladores foram projetados com sucesso os controladores e desenvolvido finalmente, um sistema de controle preditivo com modelo não linear descentralizado para controle simultâneo das concentrações de etanol no fermentador e de etanol condensado na corrente de saída do tanque flash.

SUGESTÕES PARA TRABALHOS FUTUROS

Como sugestões para continuação desta linha de pesquisa, propõe-se:

- Realização de ensaios experimentais com uma maior faixa de variação para as variáveis S_o , T_{ferm} , T_{flash} , e Fin , que, neste estudo, tiveram um valor aproximadamente constante, para incluir essas variáveis nos projetos de sistemas de controle preditivo.
- Desenvolvimento de controladores baseados na lógica fuzzy para o processo de produção de bioetanol com unidade de separação de produto por flash para aproveitar o conhecimento explícito obtido nos trabalhos experimentais realizados.
- Implementação na planta dos controladores preditivos ou de controladores fuzzy utilizando *software* apropriado para este propósito.
- Implementação de *software* sensors, objetivando sempre a melhor precisão no seu desempenho para o monitoramento das concentrações de etanol no processo de produção de bioetanol utilizando a fermentação extrativa continua com evaporador flash.
- Realizar estudos para o desenvolvimento de um modelo fenomenológico para o sistema de separação flash que permita eliminar a diferença existente entre os resultados experimentais e os obtidos com o uso de simuladores que utilizam modelos para o equilíbrio líquido-vapor: NRTL ou UNIFAC.

APÊNDICES

APÊNDICE 1A.

DESCRIÇÃO COMPLEMENTAR DE PROCEDIMENTOS PARA REMOÇÃO DO ETANOL DESDE O CALDO DE FERMENTAÇÃO.

APÊNDICE 1.B.

BALANÇOS DE MASSA NA FERMENTAÇÃO CONTÍNUA COM EVAPORADOR FLASH

APÊNDICE 1.C.

MODELOS PARA O EQUILÍBRIO LÍQUIDO-VAPOR: NRTL e UNIFAC

APÊNDICE 2.A.

PROCESSOS DE APRENDIZAGEM DAS REDES NEURAIIS

APÊNDICE 3.A.

CONSIDERAÇÕES MAIS AMPLAS PARA SELEÇÃO DA ESTRUTURA DO MODELO E ESTIMAÇÃO DOS PARÂMETROS DA ESTRUTURA SELECIONADA

APÊNDICE 3.B.

MODELOS NÃO LINEARES PARA IDENTIFICAÇÃO DE PROCESSOS

APÊNDICE 3.C.

OUTROS MÉTODOS QUE PODEM SER APLICADOS PARA OBTER A TOPOLOGIA ÓTIMA NOS MODELOS DE REDES NEURAIIS

APÊNDICE 3.D.

OUTROS MÉTODOS PARA MELHORAR A VELOCIDADE DE APRENDIZAGEM DO ALGORITMO DE *BACKPROPAGATION*

APÊNDICE 1.A.

DESCRIÇÃO COMPLEMENTAR DE PROCEDIMENTOS PARA REMOÇÃO DO ETANOL DESDE O CALDO DE FERMENTAÇÃO.

1) FERMENTAÇÃO - REMOÇÃO DE ETANOL POR SISTEMA DE ABSORÇÃO (gás-stripping).

Neste sistema, o produto da fermentação é enviado para uma coluna empacotada onde o etanol é removido. A técnica utiliza CO₂ produzido na fermentação como gás absorvedor.

Liu *et al.* (1990) estudaram a remoção do etanol a partir do meio de fermentação através da aplicação de absorção gasosa como um meio de prevenir a inibição por produto e de melhorar a produtividade na fermentação alcoólica. Segundo seus resultados, para uma taxa de diluição dada, com o aumento no fator de absorção: a) a concentração das células aumenta, b) o consumo de substrato aumenta, c) a concentração de etanol no caldo de fermentação diminui, e 4) a produtividade de etanol aumenta.

Pham *et al.* (1989) estudaram as características de um sistema para fermentação e absorção simultâneos do etanol produzido, segundo seus resultados, a taxa de absorção gasosa de etanol foi fortemente dependente da taxa de fluxo de gás absorvedor CO₂ e da temperatura. Com o uso de fermentação e absorção simultâneos para eliminar a inibição pelo produto, uma melhora de 27% na produtividade do processo de fermentação foi obtido, a partir de 292 g/L de glucose na alimentação usando leveduras floculantes.

2) FERMENTAÇÃO - REMOÇÃO DE ETANOL POR MEMBRANAS

Neste caso, é estudada a integração de fermentação contínua e separação por membrana na mesma unidade para melhorar o desempenho dos biorreatores.

Segundo Groot *et al.* (1992), as técnicas de separação por membrana podem ser aplicadas para a recuperação ou purificação de etanol, para a

recuperação de etanol a partir do caldo de fermentação podem ser usadas: pervaporação, perextração e osmose reversa; utilizando membranas seletivas para a remoção de etanol a partir de misturas de etanol / água . Na purificação do etanol, para quebrar o azeótropo etanol / água são úteis membranas seletivas para a remoção de água a partir de misturas de etanol / água.

2.1) PERVAPORAÇÃO

A pervaporação envolve a absorção seletiva de uma mistura líquida na membrana, a difusão através da membrana, e a dessorção em uma fase vapor sobre o lado do permeado (MULDER *et al.*, 1983). Segundo Peng *et al.* (2010), geralmente, funcionamento a vácuo é aplicado para o lado do permeado devido à mudança de fase que acontece. A pervaporação é considerada uma técnica promissora de separação para a purificação de etanol a partir de caldos de fermentação durante a produção de bioetanol (CHOVAU *et al.*, 2011), devido à sua simplicidade e não-toxicidade para microorganismos de fermentação utilizados. Os fatores desejados quando se separa etanol a partir de caldo de fermentação por pervaporação são: estabilidade da membrana, alta seletividade e alto fluxo (WU *et al.*, 2004).

2.2) PEREXTRAÇÃO

Na perextração, a extração seletiva de etanol a partir do caldo de fermentação é realizada utilizando uma membrana que tem o caldo de fermentação de um lado e um solvente do outro lado, o álcool se move para o lado solvente. Isso elimina a alta energia calorífica necessária para separar por evaporação o álcool do caldo de fermentação.

2.3) OSMOSE

Na osmose normal, a água doce atravessa a membrana para diluir uma solução concentrada. Na osmose reversa, a pressão é exercida sobre o lado com a solução concentrada para forçar que as moléculas de água atravessem a membrana para o lado de água doce. A osmose reversa pode ser utilizada para

purificar etanol hidratado e obter etanol anidro, separando a água que é uma impureza indesejável.

Diversos estudos foram desenvolvidos na literatura integrando fermentação com remoção de etanol por membranas, entre eles temos os citados a seguir.

Cho e Hwang (1991) estudaram a integração de fermentação contínua e separação por pervaporação e relatam que em relação à fermentação contínua convencional, o desempenho do sistema proposto resultou em densidades elevadas de células de levedura, a redução da inibição por etanol, maior tempo de residência do substrato, maior consumo de glicose, recuperação de etanol limpo e concentrado e um aumento de 10-20% na produtividade de etanol.

Groot *et al.* (1992) estudaram um sistema onde a recuperação de etanol por pervaporação foi acoplada à fermentação da glicose por *Saccharomyces cerevisiae*. A pervaporação foi realizada com módulos de membrana de fibra oca a base de silicone comercial. Configurações com retenção de células por microfiltração foram utilizadas a fim de otimizar a produtividade. Os autores do estudo manifestam que nos sistemas com pervaporação uma alimentação contendo 360 g/L de glicose pode ser convertida quase que completamente. A produtividade foi de 14 g/L.h em um sistema com apenas pervaporação, e poderia ser aumentada para 43 g/L.h no sistema com reciclo por microfiltração.

O'Brien e Craig, Jr., (1996) relatam os resultados de um processo, no qual a fermentação com levedura convencional foi acoplada a uma unidade de pervaporação de membrana de placa plana, para recuperar continuamente um fluxo de etanol enriquecido a partir do caldo de fermentação. O processo utilizou um fluxo de alimentação de dextrose concentrada e produziu um permeado de 20%- 23% (w/w) de etanol, mantendo um nível de etanol 4-6%, em um fermentador de tipo tanque-agitado. Durante a operação contínua alcançou produtividades de 4,9 – 7,8 g/L.h

Kaseno *et al.* (1999) propõem um novo método de fermentação de longo prazo com água residual mínima, e avaliam o efeito da remoção de etanol por

pervaporação na fermentação de etanol. Os autores relatam que glicose e fermento *Saccharomyces cerevisiae* imobilizado foram utilizados para a fermentação. Um módulo de uma membrana porosa hidrofóbica feita de polipropileno foi utilizado para o processo de pervaporação. A taxa de alimentação foi igual à soma da taxa de produção e a taxa de drenagem do caldo de fermentação. Quando a concentração de etanol no caldo permaneceu praticamente constante em 50 g / L, a taxa aparente de produção de etanol foi de 2 vezes maior do que aquela na fermentação sem pervaporação. A eficiência de conversão de glicose a etanol foi de 96,3% do rendimento teórico. A quantidade de águas residuárias foi 38,5% de aquela descarregada, desde o processo em batelada convencional.

Wu et al. (2004), pesquisaram e compararam o comportamento da pervaporação do caldo de fermentação com aqueles de misturas de etanol. O etanol foi produzido por *Sacharomyces cerevisiae* utilizando glicose de grau técnico, e recuperado por pervaporação através de uma membrana de polidimetilsiloxano. Relatam que o caldo de fermentação real melhorou o desempenho em pervaporação da membrana. Embora ambos, o caldo de fermentação e a suspensão de etanol, contivessem células, só o caldo de fermentação sempre resultou em maior fluxo de etanol e seletividade em comparação com o uso da solução modelo. Os autores explicam que o metabolismo das células ativas foi fator decisivo para a melhoria da transferência de etanol através da membrana.

Matsumura e Märkl (1986) utilizaram perextração para a eliminação da inibição de etanol na fermentação contínua do etanol, com fibras ocas para um rim artificial como membrana permeável, e avaliaram a sua capacidade de extrair etanol usando vários solventes orgânicos. Com o uso de tri-n-butylphosphate como solvente extrator, uma alimentação de 500 g / L de glicose foi fermentada com sucesso por células de levedura imobilizada. Durante esta fermentação

contínua, foi mantida uma alta produtividade de etanol de 48 g /L.h e a exigência de solvente por glicose consumida foi de 6 L-solvente/kg-glicose.

Christen *et al.* (1990) utilizaram um sistema de membrana líquida suportada que consiste em uma folha de Teflon porosa como suporte, embebido com isotridecanol, para a extração de etanol durante uma fermentação semicontínua. O método evita a dispersão do solvente de extração na cultura. Os autores relataram que a remoção de etanol levou à diminuição da inibição e houve um ganho na conversão de 452 g de glicose / L contra 293 g de glicose / L sem extração. Ao mesmo tempo, a produtividade volumétrica de etanol foi aumentada 2,5 vezes, devido a uma melhoria da viabilidade da levedura, enquanto o rendimento de conversão do substrato foi mantido acima de 95% do seu valor teórico.

Mulder *et al.* (1983) relataram o uso de ultrafiltração em que, em uma primeira etapa, como um sistema de reciclagem de células, as células rejeitadas foram devolvidas ao fermentador e o etanol foi removido continuamente do fermentador com uma variação na concentração de 5% para 40% em peso. A purificação da mistura etanol-água obtida após a ultrafiltração foi realizada por um processo de pervaporação de dois estágios, para finalmente obter uma concentração de etanol de cerca de 99% em peso.

3) FERMENTAÇÃO - REMOÇÃO DE ETANOL POR EXTRAÇÃO LÍQUIDA

A remoção por extração líquida utiliza o contato direto entre um solvente e o caldo de fermentação para recuperar seletivamente o produto. Neste caso, o solvente para a extração líquida não deve ser tóxico para os microorganismos (leveduras) utilizados na fermentação. Segundo Minier e Goma (1982) o estudo do efeito inibitório de álcoois alifáticos de diferente número de carbonos nas cadeias mostra que não se observa crescimento celular em presença de álcoois que têm entre 2 e 12 carbonos. Este efeito é suprimido quando o número de carbonos é de 12 ou maior.

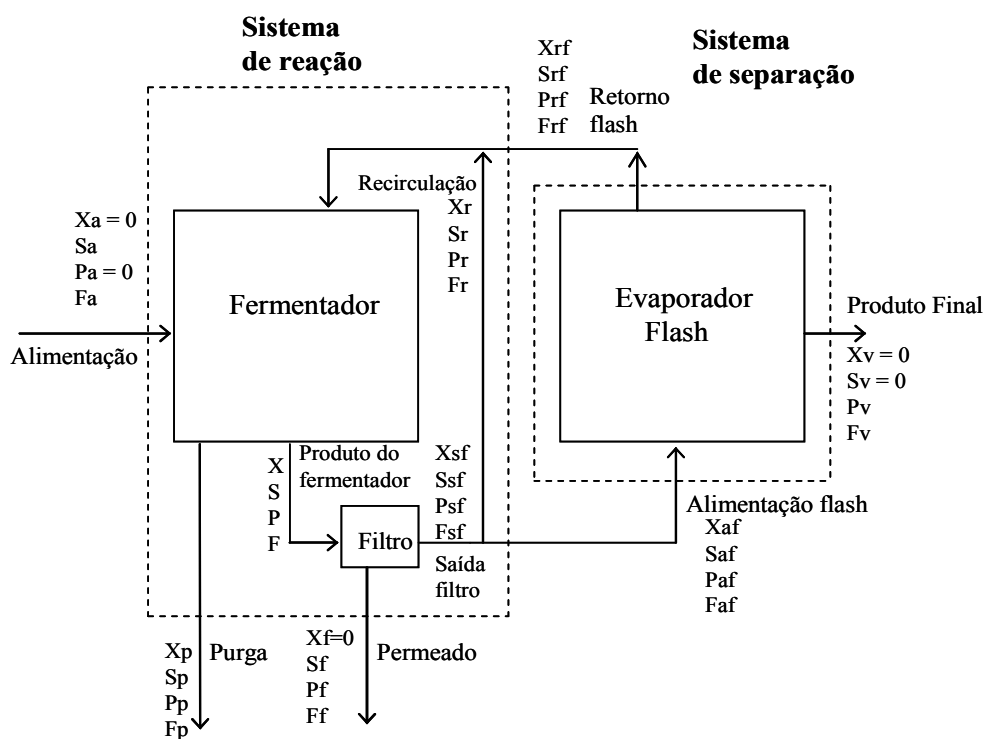
Daugulis (1994) avaliaram a capacidade de um sistema de fermentação extrativa de fluxo contínuo com reciclagem de células, utilizando álcool oleil (cis-9-octadecen-1-ol) como solvente extratante. Os autores relataram que a produtividade volumétrica em modo extrativo foi mais do que o dobro da obtida com a fermentação convencional contínua, sem efeitos deletérios na viabilidade celular, taxa de consumo específico de glicose ou na produção de etanol.

Dale *et al.* (1995) propuseram um sistema de remoção de etanol, onde um reator CSTR é acoplado a um sistema de recuperação de solvente etanol para se obter um processo contínuo de baixa energia de produção de etanol a partir de amido ou de biomassa. Nesta abordagem foram combinados os conceitos de fermentação, absorção gasosa e extração pelo solvente, para evitar os efeitos tóxicos do solvente sobre os microorganismos utilizados na fermentação. O etanol é recuperado do gás absorvedor utilizando um solvente absorvedor hidrofóbico, após o qual o gás é devolvido ao bioreator. O etanol pode então, ser removido do solvente.

APÊNDICE 1.B.

1. BALANÇOS DE MASSA NA FERMENTAÇÃO CONTÍNUA COM EVAPORADOR FLASH

Esta é uma importante etapa porque é possível integrar a cinética do processo fermentativo no funcionamento geral da unidade de produção. Para o caso da fermentação contínua com evaporador flash, se tomou como referência o sistema proposto por Atala (2006). A Figura 1.1B mostra o diagrama de blocos que representa os dois subsistemas que integram o processo. Os balanços de massas para biomassa, substrato, e produto são diretamente obtidos a partir do diagrama da Figura 1.1B.



Maiúsculo

X = Biomassa
S = Substrato
P = Produto
F = Fluxo ou Vazão

Minúsculo

f = Filtrado
p = Purga fermentador
rf = Retorno flash
af = Alimentação flash
a = Alimentação fermentador
v = Vapor
r = Recirculação
sf = Saída filtro

Figura 1.1B. Diagrama de blocos do processo de fermentação contínua com separação do produto por Flash

1.1. Balanço de Massa no Fermentador

1.1.1. Balanço de Biomassa

$$\frac{d}{dt}(X.V) = F_a.X_a + F_{rf}.X_{rf} - (F_p.X_p + F_f.X_f + F_{af}.X_{af}) + r_x.V \quad (1B.1)$$

Onde: $X_a = X_f = 0$, $X_p = X$

X representa a biomassa total, em qualquer tempo t .

$X = X_v + X_d$ e $r_v = r_x - r_d$,

X_v é a biomassa viável e X_d é a biomassa morta, r_v é taxa de formação de biomassa viável, r_x taxa formação de biomassa total, r_d taxa de morte celular.

O volume do sistema V é constante, então tem-se:

$$\frac{d}{dt}X = \frac{F_{rf}}{V}.X_{rf} - \left(\frac{F_p}{V}.X + \frac{F_{af}}{V}.X_{af}\right) + r_x \quad (1B.2)$$

1.1.2. Balanço de Produto

$$\frac{d}{dt}(P.V) = F_a.P_a + F_{rf}.P_{rf} - (F_p.P_p + F_f.P_f + F_{af}.P_{af}) + r_p.V \quad (1B.3)$$

Como:

$P_a = 0$, e $P_p = P_f = P_{af} = P$;

$$\frac{d}{dt}P = \frac{F_{rf}}{V}.P_{rf} - \frac{P}{V}(F_p + F_f + F_{af}) + r_p$$

Considerando correções pelo volume ocupado pelas células e pelo etanol intracelular Atala (2006) propõe a equação:

$$\frac{d}{dt}P = \frac{\left[F_{rf}.P_{rf} - P.(F_p + F_f + F_{af}) + r_p.V \right] + \left[\frac{P.V}{\rho} . (1 - \gamma) . \left(\frac{d}{dt}X \right) \right]}{\left[\frac{V}{\rho} . (\gamma.X - X + \rho) \right]} \quad (1B.4)$$

Onde:

ρ : constante que expressa a relação entre massa seca e massa úmida.

γ : constante que expressa a relação entre etanol intracelular e extracelular.

1.1.3. Balanço de Substrato

$$\frac{d}{dt}(S.V) = F_a.S_a + F_{rf}.S_{rf} - (F_p.S_p + F_f.S_f + F_{af}.S_{af}) - r_s.V \quad (1B.5)$$

Como: $S_p = S_f = S_{af} = S$;

$$\frac{d}{dt}S = \frac{F_a}{V}.S_a + \frac{F_{rf}}{V}.S_{rf} - \frac{S}{V}(F_p + F_f + F_{af}) - r_s$$

Considerando o volume ocupado pelas células (ATALA, 2006):

$$\frac{d}{dt}S = \frac{[F_a.S_a + F_{rf}.S_{rf} - S.(F_p + F_f + F_{af}) - r_s.V] + \left[\frac{V.S}{\rho} \left(\frac{d}{dt}X \right) \right]}{\left[\frac{V}{\rho} . (\rho - X) \right]} \quad (1B.6)$$

1.2. Balanço de Massa no Tanque Flash

1.2.1. Balanço de Biomassa

$$\frac{d}{dt}(X_{rf}.V_f) = F_{af}.X_{af} + (F_{rf}.X_{rf} - F_v.X_v) + r_x.V_f \quad (1B.7)$$

$X_v = 0$;

$$\frac{d}{dt}X_{rf} = \frac{F_{af}}{V_f}.X_{af} - \frac{F_{rf}}{V_f}.X_{rf} + r_x$$

1.2.2. Balanço de Produto

$$\frac{d}{dt}(P_{rf}.V_f) = F_{af}.P_{af} - (F_v.P_v + F_{rf}.P_{rf}) + r_p.V_f \quad (1B.8)$$

Como: $P_{af} = P$;

$$\frac{d}{dt}P_{rf} = \frac{F_{af}}{V_f}.P - \frac{1}{V_f}(F_v.P_v + F_{rf}.P_{rf}) + r_p$$

Considerando o volume ocupado pelas células no reator e a concentração de etanol intracelular (ATALA, 2006):

$$\frac{d}{dt} Prf = \frac{[Faf.P - (Fv.Pv + Frf.Prf) + rp.Vf] + \left[\frac{Prf.Vrf}{\rho} \cdot (1 - \gamma) \cdot \left(\frac{d}{dt} Xrf \right) \right]}{\left[\frac{Vf}{\rho} \cdot (\gamma.Xrf - Xrf + \rho) \right]} \quad (1B.9)$$

1.2.3. Balanço de Substrato

$$\frac{d}{dt} (Srf.Vf) = Faf.Saf - (Fv.Sv + Frf.Srf) - rs.Vf \quad (1B.10)$$

Como: Saf=S, Sv=0;

$$\frac{d}{dt} Srf = \frac{Faf}{Vf} . S - \frac{Frf}{Vf} Srf - rs$$

Considerando o volume ocupado pelas células (ATALA, 2006):

$$\frac{d}{dt} Srf = \frac{[Faf.S - Frf.Srf - rs.Vf] + \left[\frac{Vf.Srf}{\rho} \cdot \left(\frac{d}{dt} Xrf \right) \right]}{\left[\frac{Vf}{\rho} \cdot (\rho - Xrf) \right]} \quad (1B.11)$$

No caso das equações (2.27), (2.29), (2.32) e (2.34) a biomassa total é substituída pela biomassa viável, $X=Xrf=Xv$, devido ao fato de que é preciso distinguir o efeito das células viáveis e das células mortas no consumo de substrato e na formação de produto. A atividade das células viáveis resulta em consumo de açúcar e formação de produto, as células mortas não têm efeito em dS/dt e dP/dt (BOUVILLE, 2008).

1.3. Balanço de Massa no Filtro

1.3.1. Balanço de Biomassa (X)

$$Xsf.Fsf = X.F - Xf.Ff \quad (1B.12)$$

Onde:

$$Xf = 0, \text{ e } Xsf = Xr = Xaf$$

1.3.2. Balanço de Substrato (S)

$$S_{sf} \cdot F_{sf} = S \cdot F - S_f \cdot F_f \quad (1B.13)$$

Sendo:

$$S_{sf} = S_r = S_{af} = S$$

1.3.3. Balanço Produto (P)

$$P_{sf} \cdot F_{sf} = P \cdot F - P_f \cdot F_f \quad (1B.14)$$

Sendo:

$$P_{sf} = P_r = P_{af} = P;$$

APÊNDICE 1.C.

MODELOS PARA O EQUILÍBRIO LÍQUIDO-VAPOR: NRTL e UNIFAC

1. Modelo NRTL

O modelo NRTL (non-random, two liquid) foi desenvolvido por Renon e Prausnitz (1968) baseados na teoria de dois líquidos de Scott, que assume que uma mistura binária tem uma estrutura com distribuições de moléculas, nas quais cada molécula considera-se estar rodeada por moléculas de ambos líquidos, em proporções determinadas pela energia de interação entre as moléculas. O modelo é aplicável a sistemas parcialmente miscíveis.

Para o caso binário, a energia em excesso de Gibbs no modelo NRTL está dada pela seguinte equação:

$$\frac{g^E}{R \cdot T} = x_1 \cdot x_2 \cdot \left(\frac{\tau_{21} \cdot G_{21}}{x_1 + x_2 \cdot G_{21}} + \frac{\tau_{12} \cdot G_{12}}{x_2 + x_1 \cdot G_{12}} \right) \quad (1C.1)$$

Onde:

$$\tau_{12} = \frac{g_{12} - g_{22}}{R \cdot T} \quad \tau_{21} = \frac{g_{21} - g_{11}}{R \cdot T} \quad (1C.2)$$

$$G_{12} = \exp(-\alpha_{12} \cdot \tau_{12}) \quad G_{21} = \exp(-\alpha_{12} \cdot \tau_{21}) \quad (1C.3)$$

Os parâmetros g_{ij} representam os parâmetros de energia característicos das interações i e j . O parâmetro α_{12} está relacionado com a não randomicidade (ou a não aleatoriedade) na mistura, quer dizer, que os componentes na mistura não se distribuem aleatoriamente, mas que seguem um padrão ditado pela composição local. Quando α_{12} é zero, a mistura é completamente aleatória, se aproximando ao equacionamento de Margules de dois sufixos.

Os coeficientes de atividade podem ser expressos como:

$$\gamma_1 = \exp \left[(x_2)^2 \cdot \left[\tau_{21} \cdot \left(\frac{G_{21}}{x_1 + x_2 \cdot G_{21}} \right)^2 + \frac{\tau_{12} \cdot G_{12}}{(x_2 + x_1 \cdot G_{12})^2} \right] \right] \quad (1C.4)$$

$$\gamma_2 = \exp \left[(x_1)^2 \cdot \left[\tau_{12} \cdot \left(\frac{G_{12}}{x_2 + x_1 \cdot G_{12}} \right)^2 + \frac{\tau_{21} \cdot G_{21}}{(x_1 + x_2 \cdot G_{21})^2} \right] \right] \quad (1C.5)$$

Para um sistema multicomponente, a energia de excesso de Gibbs é expressa como:

$$\frac{g^E}{RT} = \sum_{i=1}^m x_i \frac{\sum_{j=1}^m \tau_{ji} G_{ji} x_j}{\sum_{i=1}^m G_{ii} x_{ii}} \quad (1C.6)$$

Onde:

$$\tau_{ij} = \frac{g_{ji} - g_{ii}}{RT}$$

$$G_{ji} = \exp(-\alpha_{ji} \tau_{ji}) \quad (\alpha_{ji} = \alpha_{ij})$$

O coeficiente de atividade para o componente i na mistura está dado pela equação:

$$\ln \gamma_i = \frac{\sum_{j=1}^m \tau_{ji} G_{ji} x_j}{\sum_{i=1}^m G_{ii} x_{ii}} + \sum_{j=1}^m \frac{x_j G_{ij}}{G_{ij} x_{ij}} \left[\tau_{ij} - \frac{\sum_{n=1}^m x_n \tau_{nj} G_{nj}}{\sum_{l=1}^m G_{ij} x_l} \right] \quad (1C.7)$$

O equacionamento NRTL fornece uma boa representação dos dados experimentais, embora sejam necessários dados de boa qualidade para estimar os parâmetros das equações. Maiores detalhes são dados em vários textos, como por exemplo Raal (1998), Walas (1985), Praunisz *et al.*(1980) and Praunisz *et al.* (1986).

2. Modelo UNIFAC (Uniquac Functional-group Activity Coefficients)

Desenvolvido por Fredenslund *et al.* (1975), tem em conta as interações intermoleculares e a forma e tamanho da molécula de cada componente do sistema (SALCEDO, 2009). Utiliza o conceito de contribuição de grupos, esse conceito baseia-se na ideia de que as propriedades de uma mistura poder ser calculadas aproximadamente considerando a soma das contribuições individuais dos grupos funcionais existentes nas moléculas presentes nessa mistura.

Nesse modelo, os coeficientes de atividade γ_i , podem ser obtidos utilizando a seguinte equação:

$$\ln(\gamma_i) = \ln(\gamma_i^C) + \ln(\gamma_i^R) \quad (1C.8)$$

Os termos da direita da equação expressam as contribuições combinatorial e residual, respectivamente.

No modelo UNIFAC modificado proposto por Peres e Macedo (1997), a contribuição combinatorial usada é a mesma que a proposta por Larsen *et al.*(1987), que é definida como:

$$\ln(\gamma_i^C) = \ln\left(\frac{\varphi_i}{x_i}\right) + 1 - \frac{\varphi_i}{x_i} \quad (1C.9)$$

Onde φ_i é a fração de volume molecular modificado do componente i :

$$\varphi_i = \frac{x_i (R_i)^{2/3}}{\sum_j x_j (R_j)^{2/3}} \quad (1C.10)$$

Onde R_i é o parâmetro de volume do componente i , e x_i é a fração molar.

A contribuição residual é a original proposta por Fredenslund *et al.*(1975):

$$\ln(\gamma_i^R) = \sum_k \nu_k^i (\ln \Gamma_k - \ln \Gamma_k^i) \quad (1C.11)$$

Os coeficientes de atividade Γ_k são calculados com:

$$\ln \Gamma_k = Q_k \left[1 - \ln \left(\sum_{j=1}^{ng} \theta_j \psi_{j,k} \right) - \sum_{j=1}^{ng} \frac{\theta_j \psi_{j,k}^j}{\sum_{m=1}^{ng} \theta_m \psi_m^j} \right] \quad (1C.12)$$

Onde ng é o número de grupos, e ψ_k^l são parâmetros de interação entre grupos, e dependem da temperatura segundo a equação:

$$\psi_k^l = \exp \left[\frac{-a_{k,l}}{T} \right] \quad (1C.13)$$

As frações de superfície para grupos, θ_k , são derivadas das frações molares X_k pelas equações:

$$\theta_k = \frac{X_k Q_k}{\sum_{m=1}^{ng} X_m Q_m} \quad (1C.14)$$

$$X_l = \frac{\sum_j x_j \nu_{l,j}}{\sum_j x_j \sum_k \nu_{k,j}} \quad (1C.15)$$

Segundo Vidal (1997), os métodos que dependem diretamente dos dados experimentais devem ser os métodos de escolha, quando possível, e o NRTL é um método que requer determinação prévia dos parâmetros a partir de dados experimentais, o problema acontece quando temos uma mistura contendo três ou mais componentes, e nós temos o acesso a dados para alguns dos componentes binários, mas não para todos eles. Os métodos de contribuição de grupos (UNIFAC ou ASOG) são uma outra alternativa para compensar a falta de dados experimentais.

APÊNDICE 2.A.

1. PROCESSOS DE APRENDIZAGEM DAS REDES NEURAIS

O processo de aprendizagem das redes neurais *feedforward* é um processo iterativo no qual são apresentados exemplos e são aplicados ajustes em seus pesos sinápticos e bias. O propósito dos ajustes dos pesos é minimizar a diferença entre o vetor de saída calculado pela rede e o vetor de saída desejado (RUMELHART *et al.*, 1986a e 1986b).

Haykin (2007) define a aprendizagem como um processo pelo qual os parâmetros livres de uma rede neural são adaptados através de um processo de estimulação pelo ambiente, no qual a rede está inserida. O tipo de aprendizagem é determinado pela maneira pela qual a modificação dos parâmetros ocorre.

O *algoritmo de aprendizagem* é um conjunto preestabelecido de regras bem definidas cuja função é modificar os pesos sinápticos da rede de uma forma ordenada para alcançar um objetivo de aprendizagem.

1.1. REGRAS BÁSICAS DE APRENDIZAGEM

1.1.1. APRENDIZAGEM POR CORREÇÃO DE ERRO

O erro é calculado quando se compara o sinal de saída do neurônio com uma resposta desejada, e logo são aplicados ajustes corretivos aos pesos sinápticos do neurônio para aproximar passo a passo o sinal de saída da resposta desejada. Este objetivo é alcançado minimizando-se uma função de custo ou índice de desempenho, $I(k)$, definido em termos do sinal do erro $e_j(k)$ como:

$$I(k) = \frac{1}{2} e_j^2(k)$$

O argumento se refere à iteração k no processo de aproximação da saída à resposta desejada. Os ajustes passo a passo dos pesos sinápticos dos neurônios são feitos de acordo com a regra de aprendizagem normalmente referida como regra delta ou regra de Widrow-Hoff, que pode ser formulada como:

O ajuste feito em um peso sináptico de um neurônio j é proporcional ao produto do sinal de erro $e_j(k)$ pelo sinal de entrada $x_i(k)$.

Utilizando a regra delta, o valor atualizado do peso sináptico $w_{ji}(k+1)$ é calculado fazendo o ajuste no valor actual do peso sináptico $w_{ji}(k)$ por:

$$w_{ji}(k+1) = w_{ji}(k) + \eta e_j(k) x_i(k)$$

Onde η é uma constante positiva que determina a taxa de aprendizado quando avançamos em um passo no processo de aprendizagem. O sub-índice i adota valores inteiros que identificam os diferentes sinais de entrada ao neurônio.

Os ajustes dos pesos continuam até o sistema atingir um estado estável. Neste ponto, o processo é encerrado.

1.1.2. APRENDIZAGEM BASEADA EM MEMÓRIA

Este tipo de aprendizagem armazena experiências passadas em uma grande memória de exemplos de entrada-saída classificados corretamente: $(x_i, d_i), i = 1, 2, \dots, N$, onde x_i representa um vetor de entrada e d_i representa a resposta desejada correspondente.

Dado um vetor de teste, x_{teste} , a rede responde com base nos exemplos de treinamento que se encontram numa vizinhança local de x_{teste} .

São exemplos de aprendizagem baseada em memória para problemas de classificação de padrões os conhecidos como: a *regra do vizinho mais próximo* e sua variante, o *classificador pelos k vizinhos mais próximos*.

1.1.3. APRENDIZAGEM HEBBIANA

A regra de aprendizado de Hebb (HEBB, 1949) é reescrita descrevendo o sincronismo da ativação de dois neurônios interconectados como a causa do

fortalecimento ou enfraquecimento do peso da sinapse (HAYKIN, 2007; BRAGA *et al.* 2000):

- a) Se dois neurônios, em ambos os lados, de uma sinapse (conexão) são ativados simultaneamente, então a força daquela sinapse é aumentada.
- b) Se dois neurônios, em ambos os lados, de uma sinapse são ativados não simultaneamente, então aquela sinapse é seletivamente enfraquecida ou eliminada.

A regra de Hebb pode ser expressa pela equação:

$$\Delta w_{ij} = \eta x_i y_j \quad (2A.1)$$

Onde: w_{ij} é o peso da sinapse entre o neurônio i e o neurônio j , η é uma constante que determina a taxa de aprendizado, x_i é a entrada do neurônio j desde o neurônio i , y_j é a saída do neurônio j .

Segundo Brown *et al.* (1990), os mecanismos fundamentais que caracterizam uma sinapse hebbiana são:

1. Mecanismo dependente do tempo. As modificações em uma sinapse hebbiana dependem do tempo exato de ocorrência dos sinais pré-sinápticos e pós-sinápticos.
2. Mecanismo local. A modificação na sinapse utiliza somente informação dos estados dos neurônios ligados pela conexão.
3. Mecanismo interativo. A aprendizagem hebbiana depende da interação entre os sinais pré-sináptico e pós-sináptico, não se pode fazer uma previsão a partir de apenas uma dessas duas atividades.
4. Mecanismo conjuncional ou correlativo. São condições para uma modificação da eficiência sináptica: a conjunção (a ocorrência simultânea) dos sinais pré-sináptico e pós-sináptico e a correlação temporal entre os sinais pré-sináptico e pós-sináptico.

1.1.4. APRENDIZAGEM COMPETITIVA

Os neurônios de saída de uma rede neural competem entre si para se tornar ativos. Enquanto que, em uma rede neural baseada na aprendizagem hebbiana, vários neurônios de saída podem estar ativos simultaneamente, na aprendizagem competitiva somente um único neurônio de saída está ativo em um determinado instante.

1.1.5. APRENDIZAGEM DE BOLTZMANN

O algoritmo de aprendizagem de Boltzmann está baseado na distribuição de Boltzmann-Gibbs do equilíbrio termal. Uma rede neural projetada com base nesse algoritmo é denominada uma máquina de Boltzmann, nela os neurônios constituem uma estrutura recorrente e operam de uma maneira binária, eles estão ou em um estado “ligado” representado por +1, ou em um estado “desligado” representado por -1 (HAYKIN, 2007).

Os neurônios de uma máquina de Boltzmann se dividem em dois grupos: os visíveis e os ocultos. Os neurônios visíveis estão conectados ao estímulo externo ou provêm o output da rede e os neurônios ocultos que sempre operam livremente.

A rede opera em dois modos. O modo chamado de *modo preso*, no qual os neurônios visíveis estão fixados a estados específicos determinados pelo ambiente. O outro modo é o *modo de operação livre*, no qual os neurônios visíveis são liberados; então, todos os neurônios estão livres para assumir qualquer estado determinado pela regra de probabilidade de transição de estado.

1.1.5.1. Função de energia

Na máquina de Boltzmann, cada estado global da rede é determinado pelos estados particulares ocupados pelos neurônios. A cada estado global pode ser atribuído um número único chamado de "energia" daquele estado. Com os pressupostos corretos, os neurônios podem agir de forma a minimizar a energia

global. Se alguns dos neurônios são externamente forçados ou "presos" em determinados estados para representar uma determinada entrada, o sistema irá então encontrar a configuração de energia mínima que seja compatível com essa entrada (ACKLEY *et al.*, 1985). A Função de energia é representada pela equação (BOSE e LIANG, 1996; HAYKIN, 2007):

$$E = -\frac{1}{2} \sum_i \sum_j w_{ji} y_i y_j, \quad i \neq j \quad (2A.2)$$

É assumida simetria na matriz de pesos de conexão ($w_{ij} = w_{ji}$).

1.1.5.2. Probabilidade de transição de estado

A probabilidade de transição de estado do neurônio y_i é uma função da variação de energia ΔE e do parâmetro T de "temperatura".

$$P(y_i \rightarrow -y_i) = \frac{1}{1 + \exp(-\Delta E_i / T)} \quad (2A.3)$$

Apesar de que uma simetria de pesos de conexão é assumida na definição de energia que é usada na definição de probabilidade, parece que na prática a assimetria de peso pode ser tolerada em grandes redes (GALLAND e HINTON, 1989).

1.1.5.3. Regra de aprendizagem de Boltzman

De acordo com esta regra, os pesos sinápticos de uma máquina de Boltzmann são ajustados utilizando-se apenas observações disponíveis localmente sob dois diferentes modos de operação: 1) *modo preso* e 2) *modo de operação livre*. A variação Δw_{ji} aplicada ao peso sináptico w_{ji} do neurônio i para o neurônio j é (BOSE e LIANG, 1996):

$$\Delta w_{ji} = \varepsilon (p_{ji} - \hat{p}_{ji}), \quad i \neq j \quad (2A.4)$$

Onde ε o tamanho do passo, p_{ji} é a probabilidade média de que os dois neurônios i e j estejam simultaneamente em estado "ligado", quando a rede opera em modo

preso, enquanto que $\hat{p}_{ji}$ é a probabilidade média de que os dois neurônios i e j estão ambos em estado “ligado”, quando a rede está em modo de operação livre.

Segundo Haykin (2007), a regra para ajustar o peso sináptico entre dois neurônios é independente da condição de visíveis ou ocultos dos neurônios.

1.2. PARADIGMAS DE APRENDIZAGEM

1.2.1. APRENDIZADO SUPERVISIONADO

Segundo Nascimento e Yoneyama (2008), de acordo com a participação do supervisor no processo de aprendizado o aprendizado pode ser:

- a) Aprendizado com *supervisão muito forte*, quando o supervisor fornece diretamente para a rede neural os valores dos pesos. Exemplo: redes de Hopfield.
- b) Aprendizado com *supervisão forte*, quando o supervisor fornece para a rede neural um conjunto de treinamento, ou seja, um conjunto de entradas e suas respectivas saídas desejadas. O objetivo é então que a rede ajuste os seus pesos de forma a produzir, para cada entrada do conjunto de treinamento, a saída desejada fornecida pelo supervisor. Exemplo: redes de tipo *feedforward* treinadas com algoritmo de *Back-Propagation*.
- c) Aprendizado com *supervisão fraca*, quando o supervisor faz apenas o papel de um crítico, fornece uma avaliação grosseira da saída da rede neural (exemplo: certo ou errado, sucesso ou fracasso), em vez de fornecer a saída desejada. Nesta categoria, está o aprendizado por reforço. Nesse caso, existe um objetivo a ser alcançado. Durante o processo de aprendizagem, a rede realiza uma ação e recebe um sinal de reforço do ambiente que permite avaliar a qualidade de sua ação. O aprendizado é realizado através da interação contínua como o ambiente. O sistema em aprendizagem alcança o objetivo retraindo seletivamente as ações que levam a sinais de reforço de melhor qualidade.

1.2.2. APRENDIZADO NÃO SUPERVISIONADO (ou auto-organizado).

Segundo Haykin (2007), nesse caso, não há professor externo ou um crítico para supervisionar o processo de aprendizado. Em vez disso, são dadas

condições para realizar uma medida da qualidade da representação que a rede deve aprender, e os parâmetros livres da rede são otimizados em relação a esta medida. Zurada, (1992) citado por Nascimento e Yoneyama (2008), propõe a seguinte analogia. Uma rede neural sendo treinada usando aprendizado supervisionado corresponde a um estudante respondendo as questões colocadas pelo tutor e comparando suas respostas com as respostas corretas dadas pelo tutor. O caso de aprendizado não supervisionado corresponde ao estudante assistindo a uma aula gravada em vídeo pelo tutor que, no entanto, está ausente para responder as perguntas. O tutor fornece os métodos (a regra de aprendizado) e, as questões (os vetores de entrada usados durante o aprendizado), mas não as respostas (os vetores de saída correspondentes).

APÊNDICE 3.A.

CONSIDERAÇÕES MAIS AMPLAS PARA SELEÇÃO DA ESTRUTURA DO MODELO E ESTIMAÇÃO DOS PARÂMETROS DA ESTRUTURA SELECIONADA

1. SELEÇÃO DA ESTRUTURA DO MODELO

Segundo Ongunaike e Ray (1994) para o caso linear na determinação de modelos de resposta dinâmica com procedimentos de identificação de processos, os modelos são obtidos utilizando dados experimentais da resposta do processo a funções de entrada específicas: degrau, impulso, pulso (retangular ou arbitrário), ondas senoidais, ruído “branco”, ou sequências binárias pseudo-aleatórias, sendo típicos candidatos de modelos de resposta dinâmica do processo em tempo contínuo:

1) Primeira ordem mais tempo morto.

$$g(s) = \frac{Ke^{-\alpha s}}{\tau s + 1} \quad (3A.1)$$

2) Segunda ordem mais tempo morto.

$$g(s) = \frac{Ke^{-\alpha s}}{(\tau_1 s + 1)(\tau_2 s + 1)} \quad (3A.2)$$

3) Simples zero, dois pólos, mais tempo morto

$$g(s) = \frac{K(\xi s + 1)e^{-\alpha s}}{(\tau_1 s + 1)(\tau_2 s + 1)} \quad (3A.3)$$

Seborg *et al.* (2004) menciona que, os modelos empíricos que podem ser desenvolvidos, utilizando identificação de processos, a partir de dados de processo, são modelos de: estado estacionário, dinâmicos, de tempo discreto, de redes neurais artificiais e modelos mais complicados. Entre os modelos mais complicados mencionam: modelos MIMO, modelos de processos estocásticos e modelos de tempo discreto não linear.

2. ESTIMAÇÃO DOS PARÂMETROS DA ESTRUTURA SELECIONADA

Segundo Ljung (1996), os princípios e algoritmos para ajustar modelos a dados são independentes da estrutura do modelo e, igualmente aplicáveis para modelos lineares ARMAX ou modelos de redes neurais artificiais. Portanto, se pode formular um procedimento geral para obter os parâmetros do modelo.

Seja θ o vetor de parâmetros desconhecidos, o procedimento para encontrar os melhores valores de θ é:

1) Formar a sequência de erros de predição,

$$\varepsilon(t, \theta) = y(t) - \hat{y}(t|\theta), \quad t = 1, 2, \dots, N. \quad (3A.4)$$

2) Filtrar os erros de predição usando um filtro linear $L(q)$,

$$\varepsilon_F(t, \theta) = L(q)\varepsilon(t, \theta) \quad (3A.5)$$

Onde q denota o operador de deslocamento, $qu(t) = u(t+1)$.

3) Escolher uma função $\ell(\cdot)$ positiva, valorizada escalar, para medir o “tamanho” ou “norma” do erro de predição:

$$\ell(\varepsilon_F(t, \theta))$$

4) Minimizar a soma dessas normas para encontrar os melhores valores de θ :

$$\hat{\theta}_N = \arg \min_{\theta} V_N(\theta, Z^N) \quad (3A.6)$$

Onde:

$$V_N(\theta, Z^N) = \frac{1}{N} \sum_{t=1}^N \ell(\varepsilon_F(t, \theta)) \quad (3A.7)$$

Um caso particular é obtido se, por simplicidade, para a minimização da equação 3A.6 é assumido o criterio quadrático para o erro de predição e colocado o filtro $L = 1$:

$$V_N(\theta) = \frac{1}{2N} \sum_{t=1}^N |y(t) - \hat{y}(t|\theta)|^2 \quad (3A.8)$$

Se o modelo $\hat{y}(t|\theta)$ é linear nos parâmetros θ , a minimização é feita pelo algoritmo de mínimos quadrados linear, se é não linear nos parâmetros θ , pelo algoritmo de mínimos quadrados não linear (DENNIS e SCHNABEL, 1983).

No mesmo sentido, segundo Seborg *et al.*, 2004, para modelos de estado estacionário polinomiais lineares em variáveis e parâmetros, e não lineares em variáveis, mas lineares nos parâmetros, a solução ao problema de estimação de parâmetros é obtida por mínimos quadrados linear, minimizando a soma dos quadrados dos desvios entre o valor de cada dado e o estimado pelo modelo. Se o modelo empírico é polinomial não linear nos parâmetros, a solução é obtida por regressão não linear em lugar da regressão linear.

Uma outra alternativa para resolver o problema dos mínimos quadrados não linear é determinar os valores dos parâmetros é o uso de algoritmos iterativos: Gauss Newton, Levenverg-Marquardt ou Gradiente Descendente (LJUNG,1996).

APÊNDICE 3.B.

MODELOS NÃO LINEARES PARA IDENTIFICAÇÃO DE PROCESSOS

1. MODELOS POLINOMIAIS NARMAX

Os modelos NARMAX (do inglês *Non linear Auto-Regressive with Moving Average and exogenous inputs*), são estruturas não lineares autorregressivas com média móvel e entrada exógena (BILLINGS e CHEN, 1989; HABER e UNBENHAUEN, 1990; AGUIRRE, *et al.*, 1998).

O modelo NARMAX para um sistema SISO é (AL SEYAB, 2006):

$$y(k) = F[y(k-1), \dots, y(k-n), u(k-1), \dots, u(k-m), e(k), \dots, e(k-q+1)] \quad (3B.1)$$

Onde, F é uma função não linear que pode ser polinomial, dada por uma rede neural, ou outra, k representa o instante de tempo, y é a saída controlada, u é a entrada manipulada, e é o ruído, n é o número de saídas passadas utilizadas, m é o número de entradas passadas utilizadas, e q é o número de entradas de ruído atual e passadas utilizadas.

A principal dificuldade com o NARMAX ainda para sistemas SISO é a determinação de uma estrutura apropriada para o modelo que melhor represente a dinâmica do processo, as dificuldades são ainda mais agravadas para sistemas de entrada múltipla e saída múltipla (MIMO) (SAHA *et al.* 2004; AL SEYAB, 2006).

2. MODELOS DE VOLTERRA

Estes modelos podem ser descritos pela seguinte relação entrada-saída (AL SEYAB, 2006).

$$y(k) = y_0 + \sum_{j=0}^{\infty} a_j u(k-j) + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} b_{ij} u(k-i) u(k-j) + \sum_{l=0}^{\infty} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} c_{lij} u(k-l) u(k-i) u(k-j) + \dots \quad (3B.2)$$

Onde a , b e c são os parâmetros do modelo. Os modelos de Volterra podem ser utilizados para modelar sistemas não lineares. No entanto, são difíceis de usar para modelagem de sistemas MIMO, devido ao fato do número de parâmetros que precisa ser estimado, o qual pode ser proibitivamente grande para um sistema multivariável (SAHA *et al.* 2004).

3. MODELOS DE WIENER- HAMERSTEIN

Os modelos de Hammerstein e Wiener têm uma estrutura especial que facilita a sua aplicação para NMPC (*Non linear Model Predictive Control*) (CERVANTES *et al.*, 2003; SENTONI *et al.*, 1996; SAHA *et al.*, 2004). Os modelos de Wiener são particularmente úteis para representar as não linearidades do processo sem introduzir as complicações associadas com operadores não lineares gerais. Al Seyab (2006) descreve sistemas dinâmicos utilizando um elemento dinâmico linear seguido em série por um elemento não linear estático. Os modelos de Hammerstein contêm os mesmos elementos na ordem inversa (PATWARDHAN *et al.*, 1997). Na identificação de sistemas não lineares, o sistema é frequentemente representado por blocos ligados em série. A Figura 3B.1 mostra o esquema de identificação que corresponde a esses modelos (CRAMA e SCHOUKENS, 2002).

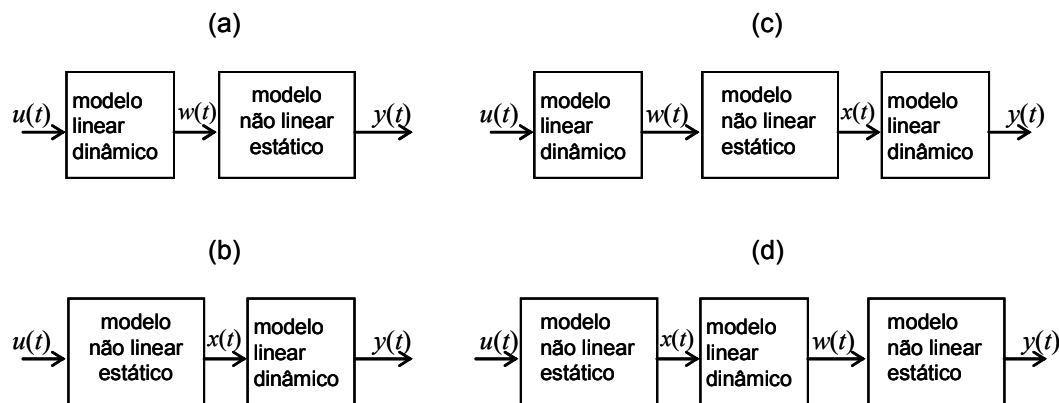


Figura 3B.1. Estrutura básica de modelos não lineares (a) Modelo de Wiener (b) Modelo de Hammerstein (c) Modelo de Wiener-Hammerstein (d) Modelo de Hammerstein–Wiener.

4. "WAVELETS"

Segundo Liu (2001), as funções *Wavelet* fornecem uma base útil para aproximação localizada de funções com qualquer grau de regularidade em diferentes escalas e com uma precisão desejada.

A representação da série *wavelet* de uma função $f(x)$ unidimensional é dada por:

$$f(x) = \sum_{k \in N} a_{j_0 k} \phi_{j_0 k}(x) + \sum_{j \geq j_0} \sum_{k \in N} b_{jk} \psi_{jk}(x) \quad (3B.3)$$

Onde: $a_{j_0 k}$ e b_{jk} , são coeficientes da *wavelet*, $\phi_{j_0 k}(x)$ é chamada de função de escala, $\psi_{jk}(x)$ é a função que define a *wavelet*:

$$\psi_{jk}(x) = 2^{j/2} \psi(2^j x - k) \quad (3B.4)$$

Para o caso n-dimensional:

$$x = [x_1, x_2, \dots, x_n],$$

Para $j \geq j_0, j \in N$,

$$k = [k_1, k_2, \dots, k_n] \in N^n, \text{ sendo } N \text{ números inteiros:}$$

$$\Phi_{jk}(x) = 2^{jn/2} \Phi(2^j x_1 - k_1, 2^j x_2 - k_2, \dots, 2^j x_n - k_n) \quad (3B.5)$$

$$\Psi_{jk}^i(x) = 2^{jn/2} \Psi(2^j x_1 - k_1, 2^j x_2 - k_2, \dots, 2^j x_n - k_n) \quad (3B.6)$$

Para $f(x) \in L_2(\mathbb{R}^n)$, a representação da série *wavelet* n-dimensional da função $f(x)$ é:

$$f(x) = \sum_{k \in N^n} a_{j_0 k} \Phi_{j_0 k}(x) + \sum_{j \geq j_0} \sum_{k \in N^n} \sum_{i=1}^{2^n-1} b_{jk}^i \Psi_{jk}^i(x) \quad (3B.7)$$

Onde: $a_{j_0 k}$ e b_{jk}^i , são coeficientes da *wavelet* (JUDITSKI, et al., 1994; LIU, 2001).

Baseadas nas *wavelet* e nas redes neurais *feedforward* foram introduzidas as chamadas *wavelet based networks* ou simplesmente *wavelet networks* para a

identificação de sistemas não lineares (ZHANG e BENVENISTE, 1992; COCA e BILLINGS, 1997; CLAUMANN, 2003; WEI *et al.*, 2010).

5. OUTROS MODELOS EMPÍRICOS COM TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL

5.1. MODELOS FUZZY

A modelagem Fuzzy é uma técnica alternativa para representar sistemas complexos não lineares. Os sistemas fuzzy têm a habilidade para tratar informações de natureza quantitativa e qualitativa, permitindo o uso de informações fornecidas por especialistas (operadores de processos) e, quando baseados, em regras têm a capacidade de aproximação universal, a capacidade para aproximar qualquer função contínua real definida sobre um domínio compacto (fechado e limitado) a qualquer grau de precisão desejada (KOSKO, 1994; ZENG e SINGH, 1995; SIMÕES e SHAW, 2007; LUO *et al.*, 2008).

O atualmente denominado Modelo Fuzzy Lingüístico foi desenvolvido com base nas ideias originais de Zadeh (1973). Alguns conceitos básicos no modelo *fuzzy* (ZADEH, 1973; SIMÕES e SHAW, 2007) são os seguintes:

a) Variável linguística.

É uma variável cujos valores são sentenças em uma linguagem natural ou artificial. Assim, se ‘não alto’, ‘pouco alto’, ‘alto’, ‘muito alto’, ‘muito muito alto’, etc, são valores de altura, então a altura é uma variável linguística.

b) Enunciados condicionais difusos.

São expressões da forma: **Se** x é A **então** y é B.

Onde: o antecedente é x é A, o conseqüente é y é B.

A e B têm um significado *fuzzy*, por exemplo:

Se x é ‘muito pequeno’ **então** y é ‘grande’.

Os valores linguísticos ‘muito pequeno’ e ‘grande’ são etiquetas de subconjuntos *fuzzy*.

Os enunciados condicionais difusos são as regras *fuzzy*.

A obtenção de um conjunto adequado de regras para o modelo *fuzzy* não é tarefa simples pela natureza linguística das variáveis. Várias contribuições foram feitas relacionadas com o uso de técnicas para gerar automaticamente conjuntos de regras *fuzzy*, para casos onde não existe conhecimento explícito expresso por um especialista, aplicações onde o especialista não existe, ou em casos de sistemas complexos. No entanto, as regras *fuzzy* geradas podem não possuir um significado lingüístico claro (YAGER e FILEV, 1994; NOZAKI *et al.* 1997; CHIU, 1997; GUILLAUME, 2001).

c) *Universo de discurso*

O universo de discurso de uma variável representa o intervalo numérico de todos os possíveis valores reais que uma variável específica pode assumir.

d) *Funções de pertinência*

Um subconjunto *fuzzy* A de um universo de discurso U é caracterizado por uma função de pertinência $\mu_A: U \rightarrow [0,1]$, a qual associa cada elemento k de U com um número $\mu_A(k)$ no intervalo $[0,1]$, que representa o grau de pertinência de k no subconjunto *fuzzy* A . Uma função de pertinência portanto, atribui valores de pertinência *fuzzy* para valores discretos de uma variável linguística. Algumas das mais utilizadas são a: triangular, trapezoidal e gaussiana. Uma explicação ampla dos conjuntos *fuzzy*, funções de pertinência, as noções de inclusão, união, interseção, complemento, etc., e das propriedades dessas noções no contexto dos conjuntos *fuzzy* é feita em Zadeh (1965), Yager (1995) e Pedrycz e Gomide (1998).

e) *Configuração básica de um sistema de inferência fuzzy*

O diagrama de blocos de um sistema *fuzzy* é apresentado na Figura 3B.2:

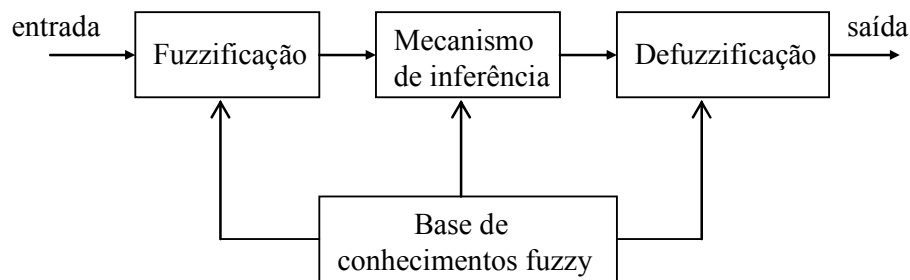


Figura 3B.2. Configuração básica de um sistema *fuzzy*

f) Fuzzificação

Procedimento que converte uma entrada *crisp* em uma variável linguística com um valor linguístico determinado usando as funções de pertinência armazenadas na base de conhecimentos *fuzzy*.

g) Mecanismo de inferência

Usando os enunciados condicionais difusos do tipo **Se x é A então y é B**, e métodos de inferência baseados em lógica *fuzzy*, transforma entradas *fuzzy* em saídas *fuzzy*. Os métodos de inferência mais conhecidos são o Min-Max (MAMDANI, 1974; MIZUMOTO, 1991) e o Produto-Soma (MIZUMOTO, 1991; MIZUMOTO, 1995).

h) Defuzzificação

Procedimento que converte a saída *fuzzy* do mecanismo de inferência em saída *crisp* (não *fuzzy*), usando métodos de defuzzificação.

Os seguintes métodos muito utilizados são descritos com detalhes em Simões e Shaw (2007): método Centro da Área (também chamado de Centro de Gravidade), método Centro do máximo e método Média do Máximo.

i) Base de conhecimentos

Compreende a base de dados (dados, definições e funções de pertinência) e a base de regras linguísticas *fuzzy*.

5.1.1. Identificação de sistemas *fuzzy*

Para Simões e Shaw (2007) a identificação de sistemas *fuzzy* significa que há a identificação da estrutura e dos parâmetros de um modelo *fuzzy*, de modo que, este modelo tenha um comportamento de entrada e saída que aproxime o do sistema dinâmico que está sendo modelado. Essa operação culmina na obtenção de uma relação *fuzzy*, baseada em N amostras, que se comporta como a imagem *fuzzy* do sistema real.

Dois importantes modelos fuzzy baseados em regras para identificação de sistemas são:

- 1) Modelo *fuzzy* Linguístico ou Modelo Mamdani: Ambos antecedente e consequente são proposições difusas.
- 2) Modelo *fuzzy* de Takagi-Sugeno: O antecedente é uma proposição *fuzzy*, o consequente é uma função crisp (não *fuzzy*).

5.1.2. Modelo Linguístico ou Modelo Mamdani

A identificação de um sistema consiste na identificação de um modelo difuso linguístico, no qual tanto o antecedente como o consequente são proposições difusas, designadas pelo Modelo de Mamdani. Este modelo linguístico difuso consiste num conjunto de regras descrito de forma genérica por:

$$R_i : \text{IF } x_i \text{ é } A_i \text{ THEN } y \text{ é } B_i, \text{ com } i=1, 2, \dots, M \quad (3B.8)$$

Onde R_i representa a regra i e M o número de regras. A variável antecedente da regra é representada por x_i e a saída difusa do sistema por y .

O conjunto de regras é obtido a partir dos dados de treino, então a partir do conjunto de regras, a resposta do modelo para uma entrada x pode ser obtida efetuando os seguintes passos (MAMDANI, 1974; MIZUMOTO, 1991; IGREJAS, 2008):

1. Para cada regra i determinar o peso do antecedente, através de:

$$\alpha_i = \mu_{A_{i1}}(x_1) \wedge \mu_{A_{i2}}(x_2) \wedge \dots \wedge \mu_{A_{im}}(x_m), \quad \text{com } i=1, 2, \dots, M \quad (3B.9)$$

2. Para cada regra derivar o conjunto difuso de saída, B_i^* , usando um operador T-norma, por exemplo: operador **and** (interseção).

$$\mu_{B_i^*}(x_1) = \alpha_i \wedge \mu_{B_i}(x_1) \quad (3B.10)$$

3. Agregar os conjuntos difusos de saída resultantes no passo 2, através da utilização de um operador T- conorma, por exemplo: operador **or** (união):

$$\mu_{B^*}(x_1) = \max_{i=1,2,\dots,k} (\mu_{B_i^*}(y)) \quad (3B.11)$$

4. Por último, transformar o conjunto difuso resultante em um valor numérico representativo (*crisp value*) usando um dos métodos de defuzzificação, por exemplo, o Centro de Gravidade:

$$VN_{y_l} = \frac{\sum_{l=1}^M \mu_{B^*}(y_l) y_{jl}}{\sum_{l=1}^M \mu_{B^*}(y_l)} \quad (3B.12)$$

5.1.3. Modelo Takagi-Sugeno

A estrutura de identificação difusa de Takagi- Sugeno tem uma forma diferente nos termos consequentes de cada regra. Neste caso, o termo

consequente da regra é agora uma função combinação linear das variáveis antecedentes, sendo que uma regra apresenta a seguinte forma:

$$R_i: \text{IF } x_1 \text{ é } A_1 \text{ e } \dots \text{ e } x_k \text{ é } A_k \text{ THEN } y = p_0 + p_1 x_1 + \dots + p_k x_k \quad (3B.13)$$

Onde:

y é a variável da consequência cujo valor é inferido.

$x_1, \dots, x_k$ são variáveis da premissa que estão também na parte da consequência.

$A_1, \dots, A_k$ são conjuntos fuzzy com funções de pertinência lineares.

Segundo Takagi e Sugeno (1984), o algoritmo raciocínio para inferir o valor de y é o seguinte:

1) Para cada regra R_i , y^i é calculado por:

$$y^i = p_0^i + p_1^i x_1 + \dots + p_k^i x_k \quad (3B.14)$$

2) O valor verdadeiro da proposição $y = y^i$ é calculado pela equação:

$$|y = y^i| = A_1^i(x_1) \wedge \dots \wedge A_k^i(x_k) \quad (3B.15)$$

Onde:

$|*|$ significa o valor verdadeiro da proposição $*$, e $\wedge$ representa a operação *min*, e $A(x)$ representa o grau de pertinência de x .

3) A saída final y inferida das n implicações é obtida como a média ponderada de todas as y^i com os pesos dados pelos valores verdadeiros $|y = y^i|$:

$$y = \frac{\sum_{i=1}^n (|y = y^i|) y^i}{\sum_{i=1}^n (|y = y^i|)} \quad (3B.16)$$

Substituindo:

$$y = \frac{\sum_{i=1}^n (A_1^i(x_1) \wedge \dots \wedge A_n^i(x_n))(p_0^i + p_1^i x_1 + \dots + p_n^i x_n)}{\sum_{i=1}^n (A_1^i(x_1) \wedge \dots \wedge A_n^i(x_n))} \quad (3B.17)$$

Onde n , é o número de regras.

Para identificação de sistemas, portanto, é preciso determinar o seguinte:

- 1) As variáveis $x_1, \dots, x_n$ das premissas das implicações (regras).
- 2) As funções de pertinência $A_1, \dots, A_n$ dos conjuntos fuzzy nas premissas.
- 3) Os parâmetros nas conseqüências $p_0, \dots, p_n$

O modelo de Takagi e Sugeno é linear nos parâmetros das funções dos consequentes das regras, de forma que esses parâmetros podem ser estimados utilizando algoritmos de Mínimos Quadrados (*Least Squares*) (TAKAGI E SUGENO, 1985; PASSINO e YURKOVICH, 1998; LJUNG, 1999; ROSS, 2004) e filtro de Kalman (JANG, 1993; ANGELOV e FILEV, 2004). Na literatura são encontradas propostas para melhorar componentes e procedimentos da modelagem *fuzzy* utilizando o modelo Takagi e Sugeno, por exemplo: otimização da estrutura das regras para modelado *fuzzy* de sistemas de alto número de dimensões (JIN, 2000), otimização de parâmetros (MÄNLE, 2001), otimização da partição do espaço de entrada (HÖPNER e KLAWONN, 2004).

5.2. MODELOS NEURO-FUZZY

Os sistemas neuro-fuzzy surgem como resultado da combinação dos sistemas *fuzzy* e as redes neurais, portanto têm também a capacidade de aproximação universal de funções contínuas e reais, definidas sobre um domínio compacto. Um sistema neuro-fuzzy é um sistema *fuzzy* que é treinado (geralmente) por um algoritmo derivado da teoria das redes neurais.

5.2.1. Sistema Neuro-Fuzzy ANFIS

O Sistema de Inferência Adaptativo Neuro-Fuzzy (do inglês, *Adaptive Neuro Fuzzy Inference Systems*, ANFIS) é proposto na literatura como um modelo fundamental que pode incorporar a experiência humana na forma de regras if – then, bem como adaptar-se através da aprendizagem. Esta arquitetura tem revelado um alto desempenho em identificação de sistemas não lineares (JANG, 1997; EFE e KAYNAK, 2000; BABUŠKA e VERBRUGEN, 2003).

A rede de ANFIS está organizada em duas partes como as regras fuzzy, cada uma das partes com várias camadas, a primeira parte é a parte do antecedente e a segunda parte é a parte da conclusão, as camadas estão conectadas entre si, na forma de uma rede neural multicamada.

Uma estrutura básica de ANFIS é mostrada na Figura. 3B.3 (JANG, 1993) com duas entradas, duas funções de pertinência para cada entrada e cinco camadas.

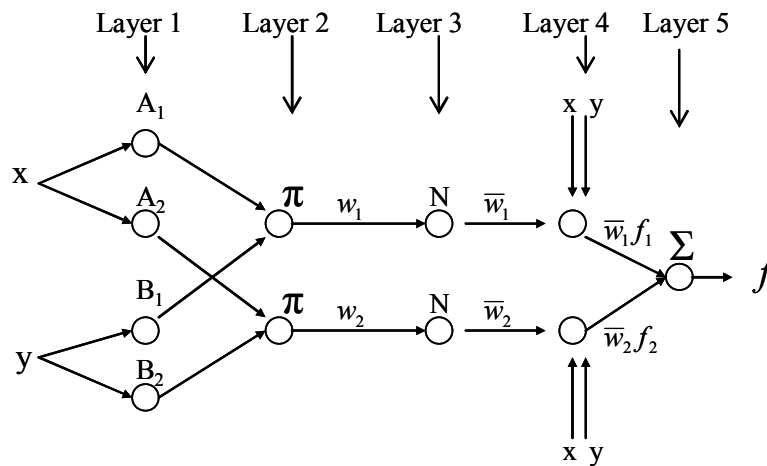


Figura 3B.3. Estrutura básica do ANFIS

As camadas executam as seguintes tarefas (JANG, 1993; GHOMSHEH *et al.* 2007).

Camada 1: Fuzzifica as entradas utilizando funções de pertinência.

$$\text{Exemplo: } \mu_{A_i}(x) = \exp\left\{-\left[\left(\frac{x-c_i}{a_i}\right)^2\right]^{b_i}\right\} \quad \mu_{A_i}(x) = \exp\left\{-\left(\frac{x-c_i}{a_i}\right)^2\right\} \quad (3B.18)$$

Onde: a_i, b_i, c_i , são os parâmetros da função de pertinência.

Camada 2: Executa operações *fuzzy* com a parte do antecedente das regras *fuzzy*. Cada nó nesta camada multiplica os sinais de entrada e envia o produto para fora:

$$w_i = \mu_{A_i}(x) \times \mu_{B_i}(y), \quad i=1,2 \quad (3B.19)$$

Outros operadores T-norma que realizam operações AND podem ser usados como a função de nó nesta camada.

Camada 3: Normaliza as saídas da camada 2.

$$\bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i=1,2 \quad (3B.20)$$

Camada 4: Opera com a saída $\bar{w}_i$ da camada 3 e a parte do consequente das regras fuzzy.

$$f_1 = p_1x + q_1y + r_1 \quad (3B.21)$$

$$f_2 = p_2x + q_2y + r_2 \quad (3B.22)$$

$$\bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i) \quad i=1,2. \quad (3B.23)$$

Onde: $p_1, p_2, q_1, q_2, r_1, r_2$, são parâmetros do consequente.

Camada 5: Computa a saída do sistema somando as saídas da camada 4.

$$f = \frac{w_1 f_1 + w_2 f_2}{w_1 + w_2} = \bar{w}_1 f_1 + \bar{w}_2 f_2 \quad (3B.24)$$

Na literatura vários métodos foram propostos para obter um conjunto ótimo de regras e para treinar redes neuro-fuzzy:

Mascioli *et al.* (1997) propuseram utilizar os modelos min-max e ANFIS para determinar um ótimo conjunto de regras IF – THEN em redes neurofuzzy.

Jang e Mizutani (1996) propuseram a aplicação do método Levenberg – Marquardt para aprendizagem de redes ANFIS.

Jang (1993) propôs utilizar uma regra de aprendizagem híbrida baseada em o algoritmo gradiente descendente e a estimativa de quadrados mínimos (*least squares estimate*, LSE) para atualizar os parâmetros da estrutura ANFIS.

Ghomsheh, *et al.* (2007) utilizaram *particle swarm optimization* (PSO) para atualizar os parâmetros da parte do antecedente e da parte do consequente na estrutura ANFIS, em um problema de identificação de um modelo não linear.

Segundo Babuska (2003), a desvantagem da modelagem neuro-fuzzy é que as técnicas para a construção e ajuste de modelos fuzzy são bastante complexas, e seu uso requer habilidades e conhecimentos específicos. Por isso, provavelmente a modelagem neuro-fuzzy de sistemas complexos permanecerá sempre como uma abordagem iterativa. Por outro lado, do ponto de vista da complexidade computacional, das abordagens que têm a propriedade de aproximação universal de funções (sistemas fuzzy, sistemas neuro-fuzzy e redes neurais), as redes neurais são a abordagem comparativamente mais simples.

APÊNDICE 3.C.

OUTROS MÉTODOS QUE PODEM SER APLICADOS PARA OBTER A TOPOLOGIA ÓTIMA NOS MODELOS DE REDES NEURAIS

1. Procedimentos estatísticos

Anders e Korn (1999) sugeriram que a análise estatística deve ser uma parte integrante da modelagem com redes neurais, e propuseram o uso de testes de hipótese para a seleção de modelos testando a significância dos parâmetros. Eles consideraram como apropriados para esse propósito os testes de Wald ou os testes LM (*Lagrange Multiplier*) porque as redes neurais *perceptron* multicamada são modelos de regressão não linear, e esses testes são utilizados para os casos de problemas que não correspondem a modelos de regressão linear clássicos; neste tipo de problemas o modelo é não linear nos parâmetros, as restrições são não lineares, ou os erros não estão distribuídos normalmente. Os fundamentos de esses testes podem-se encontrar no livro de Kennedy (2003).

2. Construção de redes

Nestas técnicas começa-se com uma rede neural pequena para a tarefa desejada, logo se adiciona um neurônio na camada escondida com pesos para as conexões com as entradas e saídas. O processo de treinamento é reiniciado e são obtidos os resultados, continua-se adicionando mais neurônios até atingir as especificações de projeto. O mais conhecido exemplo desta abordagem é correlação em cascata (*cascade correlation*) (FLAHMN e LEBIERE, 1990).

O procedimento começa com uma rede neural mínima que tem apenas algumas entradas e neurônios de saída, mas não neurônios escondidos. A rede é treinada, em seguida se adicionam os neurônios escondidos à rede um por um com as suas conexões sinápticas, obtendo-se assim uma estrutura multicamada.

Quando um novo neurônio é adicionado, os pesos sinápticos do lado das entradas do neurônio são congelados, somente os pesos sinápticos do lado de

saída do neurônio são treinados repetidamente. O procedimento de adição de novos neurônios é continuado na maneira descrita até que um desempenho satisfatório é atingido.

Os problemas importantes para o sucesso das técnicas de construção de redes são, quando parar a adição de novos neurônios e quando e onde adicionar um novo neurônio. O tratamento errado destas questões pode levar a *overtraining* e forte desaceleração no processo de construção da rede (REITERMANOVA, 2008).

3. Poda de redes (*pruning*)

O procedimento usual é: 1) Treinar uma rede multicamada grande, supostamente de tamanho excessivo para o problema, 2) Examinar os pesos das conexões para avaliar a relevância relativa dos pesos ou dos neurônios, 3) Eliminar os pesos e os neurônios menos importantes ou relevantes (podar a rede), 4) Retreinar a rede podada, 5) Repetir os passos 2- 4 até atingir o objetivo desejado.

Os métodos de poda de redes usam diferentes estratégias para medir, detectar e remover os elementos sem importância com base na medida de relevância escolhida, uma das medidas mais comuns da relevância para pesos é chamada de saliência do peso. Saliência é a sensibilidade da função de erro à remoção ou mudança de um peso simples da rede (MOZER e SMOLENSKY, 1989; LeCun *et al.*, 1990). Uma estratégia simples é eliminar parâmetros com pequena saliência, que são aqueles cuja eliminação terá o menor efeito no erro de treinamento. No entanto, segundo LeCun *et al.* (1993) estas estratégias requerem um fino ajuste dos coeficientes de poda (*pruning*) para evitar efeitos catastróficos.

Alguns métodos ilustrativos da poda de redes neurais são os seguintes:

A) *Dano Cerebral Ótimo (OBD, Optimal Brain Damage)*

Foi proposto por LeCun *et al.* (1990) e utiliza a segunda derivada da função objetivo com respeito aos parâmetros para computar as saliências. Define a saliência de um parâmetro como a mudança na função objetivo causada pela eliminação desse parâmetro.

A função objetivo (E) é aproximada por uma série de Taylor, com o objetivo de encontrar um conjunto de parâmetros cuja deleção causará o menor aumento de E (δE). Os autores, reconhecendo que o problema é praticamente insolúvel no caso geral, introduzem várias aproximações simplificadoras. Para o cálculo da matriz Hessiana utilizam só os elementos da diagonal da matriz, e assumem também que a função de custo é quase quadrática, de modo que a equação da série de Taylor se reduz a:

$$\delta E = \frac{1}{2} \sum_i h_{ii} \delta u_i^2 \quad (3C.1)$$

δu_i é um componente da perturbação δU do vetor dos parâmetros que produz uma modificação δE na função objetivo E .

h_{ii} é um elemento da diagonal da matriz Hessiana H .

Segundo Hassibi e Stork (1993) e Ragg *et al.* (1996), o método Dano Cerebral Ótimo considera que a matriz Hessiana é diagonal, o que leva a excluir pesos errados algumas vezes. Assim, os resultados não são melhores do que os obtidos com o método Cirurgião Cerebral Ótimo.

B) *Cirurgião Cerebral Ótimo (OBS, Optimal Brain Surgeon).*

Foi introduzido por Hassibi e Stork (1993). Este método usa a informação de todas as derivadas de segunda ordem da função de erro para fazer a poda da rede (excluir pesos não importantes de uma rede treinada). A função objetivo E é aproximada por uma série de Taylor, então a mudança em E (δE) produzida por uma perturbação nos pesos (δw) e dada pela equação:

$$\delta E = \left(\frac{\partial E}{\partial w} \right)^t \cdot \delta w + \frac{1}{2} \delta w^t \cdot H \cdot \delta w + O(\|\delta w\|^3) \quad (3C.2)$$

O objetivo é identificar um conjunto de pesos cuja eliminação da rede cause o menor δE , para isso fazemos as seguintes aproximações mencionadas por LeCun *et al.*, (1990):

1) Aproximação extrema: assumimos que após o processo de treinamento haver convergido, os parâmetros têm valores que correspondem a um mínimo na superfície de erro, então nesse caso o termo $(\partial E / \partial w)^t . \delta w$ pode ser igualado a zero.

2) Aproximação quadrática: assumimos que a superfície de erro em torno de um mínimo é aproximadamente quadrática, então, os termos de ordem mais alta da equação (3C.2) poder ser desconsiderados.

Com estas suposições a função objetivo E, é aproximada por:

$$\delta E = \frac{1}{2} \delta w^t . H . \delta w \quad (3C.3)$$

O objetivo do método Cirurgião Cerebral Ótimo é minimizar δE fixando em zero um dos pesos sinápticos, a eliminação desse peso é expressa pela condição:

$$\delta w_i + w_i = 0 \quad (3C.4)$$

Ou mais geralmente:

$$I_i^t . \delta w + w_i = 0 \quad (3C.5)$$

Onde I_i^t é o vetor unitario no espaço de pesos, correspondente como ao escalar w_i .

O problema de minimização da função dada na Equação 3C.3 com a restrição dada pela Equação 3C.5 é resolvido utilizando o método dos multiplicadores de Lagrange. A modificação ótima do vetor de pesos w é:

$$\delta w = - \frac{w_i}{(H^{-1})_{ii}} H^{-1} . I_i \quad (3C.6)$$

O valor ótimo do lagrangiano L_i para o peso w_i é:

$$L_i = \frac{1}{2} \frac{w_i^2}{(H^{-1})_{ii}} \quad (3C.7)$$

L_i é chamada de saliência do peso w_i , a qual é proporcional a w_i^2 .

No método Cirurgião Cerebral Ótimo, a suposição de que a matriz Hessiana é diagonal não é utilizada, de forma que os autores propõem um procedimento recursivo para facilitar o cálculo da inversa da matriz Hessiana (H^{-1}) a partir de dados de treinamento e informação estrutural da rede, o procedimento é detalhado por Hassibi e Stork (1993) e Haykin (2007).

No método, primeiro se calcula a inversa da matriz Hessiana H^{-1} , depois se encontra o peso w_i , que tem a menor saliência, se elimina o peso sináptico w_i , se atualizam todos os pesos sinápticos da rede aplicando a equação (3C.6), e se volta a calcular um novo valor da inversa da matriz Hessiana H^{-1} . A computação iterativa é finalizada quando mais nenhum peso pode ser eliminado da rede, sem um grande aumento em E (nesse ponto pode ser desejável retreinar a rede).

Os métodos Dano Cerebral Ótimo e Cirurgião Cerebral Ótimo estão direcionados à poda de pesos. A principal fraqueza que apresentam é que a sua eficiência computacional é relativamente baixa (REITERMANOVA, 2008).

O método Cirurgião Cerebral Ótimo tem inspirado o desenvolvimento de outros métodos, tais como: Cirurgião Cerebral Ótimo Geralizado (STAHLBERGER e RIEDMILLER, 1997), Cirurgião Cerebral de Unidade-Ótima (STAHLBERGER e RIEDMILLER, 1997), Cirurgião Cerebral Flexível-Ótimo (ATTIK *et al.*, 2004), e Cirurgião Cerebral Flexível-Ótimo Geralizado (ATTIK *et al.*, 2004).

As variantes do método do Cirurgião Cerebral Ótimo têm menos eficiência em redes grandes, porque são severamente influenciadas pelo ruído acumulado gerado pelo cálculo da matriz Hessiana. Eles requerem apenas um treinamento para toda a poda. Uma alternativa é voltar a aplicar estas técnicas várias vezes para resolver este problema (ATTIK *et al.*, 2005).

C) Métodos de Hagiwara

Hagiwara (1994) propôs dois métodos para detectar e remover neurônios e pesos, os quais requerem menor tempo computacional que os utilizados em o Dano Cerebral Ótimo (*Optimal Brain Damage*) ou Cirurgião Cerebral Ótimo (*Optimal Brain Surgeon*). As seguintes equações são utilizadas na aplicação dos métodos propostos:

1) Método do consumo de energia (*Consuming energy*)

$$E_i = \sum_p \sum_j y_i w_{ij} y_j \quad (3C.8)$$

2) Método da potência dos pesos (*Weights power*)

$$E_i = \sum_p \sum_j (w_{ij})^2 \quad (3C.9)$$

Nas equações 1) e 2), p é um índice sobre todos os padrões de treinamento, i é um índice sobre todas os neurônios em uma camada oculta fixa, y_i é o seu output, j é um índice sobre todos os neurônios na próxima camada, e w_{ij} é o peso sináptico entre o i -ésimo e o j -ésimo neurônio, y_j é o output do j -ésimo neurônio na próxima camada.

As duas estratégias reduzem massivamente o tamanho da rede e levam a uma melhor generalização. Segundo Hagiwara (1994), o método da potência dos pesos tem o melhor desempenho na remoção de neurônios e pesos. No entanto, segundo Hassibi *et al.* (1993) os métodos de poda baseados nas magnitudes dos pesos têm que ser aplicados com precauções porque podem também remover partes importantes das rede.

4. Uso de algoritmos evolutivos

Os algoritmos evolutivos, principalmente Algoritmos Genéticos - AG, foram utilizados de diversas maneiras para encontrar uma topologia ótima para o modelo de rede neural.

Arifovic e Gençay (2001) propuseram uma metodologia para seleção de modelos de redes *feedforward* baseada em algoritmos genéticos, o método basicamente busca o ótimo global no modelo da rede *feedforward* evoluindo sobre: o tipo de entradas, número de neurônios na camada oculta e a estrutura de conexão entre as camadas de entrada e saída.

Admuthe *et al.* (2009) usaram algoritmos genéticos (AG) para estabelecer o número ótimo de neurônios em camadas, funções de transferência, taxa de aprendizado, momentum e número de épocas para um problema dado. No processo, o AG evolui com informação de várias estruturas, diferentes parâmetros e níveis de ajuste de cada estrutura. O AG decide o ajuste utilizando a rede neural como função objetivo.

Tahmasebi e Hezarkhani (2010) utilizaram AG para otimizar os parâmetros: número de entradas, número de neurônios nas camadas ocultas, número de camadas ocultas, o momentum e a taxa de aprendizado. O critério utilizado para avaliar o ajuste da solução potencial foi o mse (*mean squared error*).

Aghbashlo *et al.* (2011) em uma primeira etapa utilizaram ANN *feed forward* multicamada treinadas por *back propagation* para desenvolver modelos utilizando uma variável de saída e quatro variáveis de entrada, os fatores considerados para a topologia das redes foram: o número de neurônios, o coeficiente de momentum, o tamanho de passo, o número de épocas de treinamento e o número de corridas de treinamento (*training runs*). Em uma segunda etapa, criaram um modelo de superfície de resposta preditiva, utilizando RMS, a variável de saída foi a predição da rede e as variáveis de entrada foram os parâmetros de treinamento da rede. Em uma terceira etapa utilizaram algoritmos genéticos para encontrar a topologia ótima da rede neural no modelo de superfície de resposta.

Outro tipo de algoritmos bio-inspirados como os de inteligência de enxame (*swarm intelligence*) também foram propostos para otimizar a topologia das redes neurais. Nestas propostas: **Inteligência** é a capacidade de um sistema para adaptar o seu comportamento para atingir seus objetivos em uma variedade de ambientes (FOGEL, 1995) e **Swarm** é a população de elementos interagentes que é hábil para otimizar algum objetivo global através da busca cooperativa de um espaço (KENNEDY e EBERHART, 2001).

Swarm Intelligence reúne uma gama de algoritmos que simulam a dinâmica de populações de pequenos indivíduos para solucionar problemas com sofisticado grau de complexidade, de forma extremamente eficiente. Entre elas temos:

a) *Ant Colony Optimization* (ACO), proposta por Dorigo *et al.* (1996). O algoritmo simula o comportamento natural das formigas, incluindo mecanismos de cooperação e adaptação, com os quais os agentes são dirigidos para otimização local e global.

b) *A técnica de otimização baseada em partículas* (*Particle Swarm Optimization-PSO*), proposta por Kennedy and Eberhart (1995). O algoritmo de otimização é baseado na população de indivíduos capazes de interagir entre si e com o meio ambiente. Os indivíduos são representados por pontos, denominados de partículas. Uma partícula irá se mover em uma direção, que é função da posição atual da partícula, velocidade da partícula, da posição da partícula que levou a seu melhor desempenho até o momento, e do melhor desempenho global do sistema até o momento.

Conforth e Meng (2008) utilizaram *Ant Colony Optimization* para otimizar a topologia da rede neural e *Particle Swarm Optimization*, para ajustar os pesos de conexão da rede nos modelos de ANN com topologia selecionada com ACO, em um método desenvolvido para treinar ANNs.

APÊNDICE 3.D.

OUTROS MÉTODOS PARA MELHORAR A VELOCIDADE DE APRENDIZAGEM DO ALGORITMO DE *BACKPROPAGATION*

1. Métodos de gradiente conjugado

O métodos do gradiente conjugado pertencem à classe dos métodos de otimização de segunda ordem, os quais operam segundo uma aproximação quadrática da superfície de erro em torno do ponto concorrente (HAYKIN, 2007). Os métodos de gradiente conjugado constituem excelentes métodos para treinar eficientemente redes neurais, devido à sua simplicidade, eficiência numérica, e os seus requerimentos de memória muito baixos (LIVIERIS e PINTELAS, 2012). Estes métodos geram uma seqüência de pesos usando a fórmula iterativa:

$$w_{k+1} = w_k + \eta_k d_k, k = 0, 1, \dots, \quad (3D.1)$$

$$s_{k-1} = w_k - w_{k-1} \quad (3D.2)$$

Onde k é a iteração (época) atual, $w_0 \in \mathbb{R}^n$ é um ponto inicial dado, $\eta_k > 0$ é a taxa de aprendizagem, e d_k é uma direção de busca definida por :

$$d_k = \begin{cases} -g_0 & \text{se } k = 0 \\ -g_k + \beta_k d_{k-1}, & \text{se } k \geq 1 \end{cases} \quad (3D.3)$$

Onde $g_k = \nabla E(w_k)$ é o gradiente da função de erro E em w_k , e β_k é um escalar que determina a direção de busca sem conhecimento explícito da matriz hessiana (HAYKIN, 2007). Na literatura, as diferentes propostas para o cálculo do escalar β_k dão origem aos distintos métodos de gradiente conjugado, assim temos:

Método de Hestenes-Stiefel (HS), (Hestenes e Stiefel, 1952).

$$\beta_k^{HS} = \frac{g_k^t y_{k-1}}{y_{k-1}^t d_{k-1}} \quad (3D.4)$$

Método de Fletcher-Reeves (FR), (Fletcher e Reeves, 1964).

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \quad (3D.5)$$

Método de Polak-Ribière (PR), (Polak e Ribière, 1969).

$$\beta_k^{PR} = \frac{g_k^t y_{k-1}}{\|g_{k-1}\|^2} \quad (3D.6)$$

Onde: y_{k-1} é definido por: $y_{k-1} = g_k - g_{k-1}$, além disso, $s_{k-1} = w_k - w_{k-1}$, e $\|\cdot\|$ denota a norma euclidiana.

O método de Polak-Ribière é considerado um dos mais eficientes métodos de gradiente conjugado. No entanto, apesar das práticas vantagens deste método, seu grande inconveniente é não ser globalmente convergente para funções gerais (LIVIERIS e PINTELAS, 2012). Muitos esforços têm sido dedicados a investigar as propriedades de convergência global de diversos métodos de gradiente conjugado e para desenvolver métodos novos de gradiente conjugado, globalmente convergentes para funções gerais e computacionalmente superiores aos métodos clássicos, se distinguem os seguintes propósitos:

a) Propor modificações no parâmetro β_k que melhorem a convergência dos métodos clássicos.

Uma análise feita por Powell (1984) revelou que quando $\beta_k^{PR} < 0$ e as direções d_k e d_{k-1} são quase opostas, o método de Polak-Ribière pode rodar indefinidamente sem convergir, portanto Powell sugeriu restringir β_k para valores não negativos, deixando $\beta_k = \max\{\beta_k^{PR}, 0\}$. Assim, cada vez que $\beta_k = 0$, o procedimento iterativo do método é reiniciado, conduzindo a uma convergência mais lenta do que seria alcançada se não houve reinício (HAGER e ZHANG, 2006). No entanto, esta opção assegura convergência global no algoritmo resultante (GILBERT e NOCEDAL, 1992; GRIPPO e LUCIDI, 1997; HAYKIN, 2007).

Hager and Zhang (2006) propuseram um novo parâmetro β_k e um novo método de gradiente conjugado, nomeado de método CG-DESCENT, onde os parâmetros são dados por:

$$\bar{\beta}_k^N = \max\{\beta_k^N, \phi_k\} \quad (3D.7)$$

$$\phi_k = \frac{-1}{\|d_k\| \min\{\phi, \|g_k\|\}} \quad (3D.8)$$

$$\beta_k^N = \frac{1}{d_k^t y_k} \left(y_k - 2d_k \frac{\|y_k\|^2}{d_k^t y_k} \right)^t g_{k+1} \quad (3D.9)$$

Onde, $\phi > 0$ é uma constante. A proposta para β_k analiticamente pode-se considerar uma modificação da fórmula para β_k^{HS} do método de Hestenes-Stiefel (LIVIERIS e PINTELAS, 2012).

Yuan (2009) propôs uma modificação para o parâmetro β_k^{PR} do método de Polak-Ribière, da seguinte forma:

$$\beta_k^{DPR+} = \beta_k^{PR} - \min \left\{ \beta_k^{PR}, C \frac{\|y_{k-1}\|^2}{\|g_{k-1}\|^4} g_k^t d_{k-1} \right\} \quad (3D.10)$$

Onde, C é um parâmetro que controla o peso relativo entre as propriedades de conjugação e descida. Segundo Livieris e Pintelas (2012), este método é globalmente convergente para funções gerais.

b) Utilizar novas equações da secante para desenvolver outros métodos de gradiente conjugado.

Nesta alternativa, as pesquisas buscam formular novas condições de conjugação, modificar a condição usual da secante e propor novos parâmetros β_k para melhorar o desempenho dos métodos de gradiente conjugado.

A condição de conjugação dos métodos de gradiente conjugado não lineares é dada por (DAI e LIAO, 2001; NARUSHIMA e YABE, 2011):

$$d_k^T y_{k-1} = 0 \quad (3D.11)$$

Perry (1978) estendeu a condição de conjugação utilizando duas características importantes dos métodos quase-Newton :

a) A condição da secante

$$\bar{H}_k s_{k-1} = y_{k-1}, \text{ e} \quad (3D.12)$$

b) A direção de busca

$$\bar{H}_k d_k = -g_k \quad (3D.13)$$

Onde $\bar{H}_k$ é uma aproximação simétrica à Hessiana.

Utilizando as Equações 3D.11, 3D.12 e 3D.13, são obtidas as seguintes relações:

$$d_k^T y_{k-1} = d_k^T (\bar{H}_k s_{k-1}) = (\bar{H}_k d_k)^T s_{k-1} = -g_k^T s_{k-1} \quad (3D.14)$$

Assim, a condição de conjugação de Perry, é definida por (WU e CHEN, 2010; NARUSHIMA e YABE, 2011):

$$d_k^T y_{k-1} = -g_k^T s_{k-1} \quad (3D.15)$$

Substituindo 3D.15 em 3D.3 obtemos o parâmetro β_k proposto por Perry (LIU e XU, 2011).

$$\beta_k^P = \frac{(y_{k-1} - s_{k-1})^T g_k}{y_{k-1}^T d_{k-1}} \quad (3D.16)$$

Dai e Liao (2001) incorporaram um parâmetro não negativo t na Equação 3D.15 e propuseram a seguinte condição de conjugação:

$$d_k^T y_{k-1} = -t g_k^T s_{k-1} \quad (3D.17)$$

Se $t=0$, a Equação 3D.17 se reduz à condição usual de conjugação Equação 3D.15. Se $t = 1$, se tem a condição de Perry.

Utilizando a condição de conjugação proposta, Dai e Liao propuseram o seguinte parâmetro β_k (DAI e LIAO, 2001; NARUSHIMA e YABE, 2011):

$$\beta_k^P = \frac{g_k^t (y_{k-1} - ts_{k-1})}{d_{k-1}^t y_{k-1}} \quad (3D.18)$$

Zhang e Xu (2001) propuseram uma equação da secante modificada:

$$\bar{H}_k s_{k-1} = \hat{y}_{k-1} \quad (3D.19)$$

$$\hat{y}_{k-1} = y_{k-1} + \frac{\theta_{k-1}}{s_{k-1}^t} u \quad (3D.20)$$

Onde $u \in \mathbb{R}^n$ é qualquer vetor que satisfaça a condição $s_{k-1}^t u \neq 0$.

$$\theta_{k-1} = 6(f_{k-1} - f_k) + 3(g_k + g_{k-1})^t s_{k-1} \quad (3D.21)$$

Yabe and Takano (2004) consideraram uma extensão da equação da secante modificada, Equação 3D.20, introduzindo um parâmetro $\rho \geq 0$, como segue:

$$\bar{H}_k s_{k-1} = z_{k-1} \quad (3D.22)$$

$$z_{k-1} = y_{k-1} + \rho \frac{\theta_{k-1}}{s_{k-1}^t} u \quad (3D.23)$$

e propuseram a seguinte condição de conjugação:

$$d_k^t z_{k-1} = -t g_k^t s_{k-1} \quad (t \geq 0) \quad (3D.24)$$

Por substituição da Equação 3D.24 na Equação 3D.3 tem-se a nova fórmula proposta para o parâmetro β_k :

$$\beta_k^{YT} = \frac{g_k^t (z_{k-1} - ts_{k-1})}{d_{k-1}^t z_{k-1}} \quad (t \geq 0) \quad (3D.25)$$

Babaie-Kafaki *et al.*, (2010) fizeram uma análise do parâmetro β_k^{YT} e afirmam que para uma função geral f , é possível que o valor do denominador tome o valor zero, também para valores de $\|s_{k-1}\|$ maiores do que um ($\|s_{k-1}\| > 1$), é esperado que a equação da secante padrão seja mais precisa do que a equação da secante modificada. Portanto, eles propõem uma versão modificada da equação da secante, Equação 3D.23, como segue:

$$\overline{H}_k s_{k-1} = \tilde{y}_{k-1}^* \quad (3D.26)$$

$$\tilde{y}_{k-1}^* = y_{k-1} + \rho_{k-1} \frac{\max\{\theta_{k-1}, 0\}}{s_{k-1}^T u} u \quad (3D.27)$$

Onde ρ_{k-1} é escolhido para estar no intervalo $[0,1]$, e $u = s_{k-1}$.

c) Propor procedimentos de reinício para melhorar a convergência do método de gradiente conjugado.

É aceito que o método de gradiente conjugado é particularmente útil para resolver problemas de otimização de funções de muitas variáveis sem restrições, especialmente problemas de grande escala, porque este método não requer o armazenamento de matrizes. No entanto, a taxa de convergência do algoritmo é linear se não se utiliza um procedimento de reinício. O problema é quando e como deveriam fazer-se os reinícios. Alguns dos pesquisadores que abordam este problema são: Beale (1972); Powell (1977); Dai *et al.*, (2004); Lukšan *et al.* (2008).

d) Desenvolver métodos de gradiente conjugado para produzir direções de busca que não precisam recomeçar o algoritmo.

Usar um reinício periódico ao longo da direção do gradiente negativo pode ser útil para prevenir a perda de conjugação devido a erros numéricos, no entanto, acredita-se geralmente que o uso de reinícios regulares pode não ser uma técnica conveniente para fazer cumprir a convergência global (GRIPPO e LUCIDI, 1997). Algumas das pesquisas que podem ser consideradas nesta alternativa são os trabalhos de: Gilbert e Nocedal (1992); Zhang *et al.* (2007); Zhang e Zhou (2008); Liu e Wang, (2011).

2. Métodos quase-Newton

Para encontrar o mínimo de uma função f de n variáveis, o método de Newton pode ser aplicado, com a fórmula de iteração seguinte:

$$w_{k+1} = w_k - (G_k)^{-1} g_k \quad (3D.28)$$

e a direção de busca de Newton se determina com:

$$G_k p_k + g_k = 0 \quad (3D.29)$$

Onde: G é a matriz Hessiana de f em w_k , $G = Hf(w_k)$, e g é o gradiente de f em

$$w_k, g = \nabla f(w_k) \quad (3D.30)$$

No entanto, para uma função f geral, a Hessiana G freqüentemente é desconhecida. Para remediar isto existem os métodos chamados Quase-Newton, que constroem $(G_k)^{-1}$ iterativamente à medida que avançam em sua execução.

Nos métodos quase-Newton, a matriz Hessiana G_k é substituída por uma matriz $\bar{G}_k$ e a direção de Newton é aproximada pela chamada direção de descida quase-Newton p_k gerada pela equação:

$$\bar{G}_k p_k + g_k = 0 \quad (3D.31)$$

Entre os vários métodos quase-Newton, o método que tem sido mais bem sucedido nos estudos publicados é o método de Broyden, Fletcher, Goldfarb, and Shanno (BFGS), neste método $\bar{G}_k$ é atualizado por a seguinte equação (LI e FUKUSHIMA, 1999):

$$\bar{G}_{k+1} = G_k - \frac{G_k s_k s_k^t G_k}{s_k^t G_k s_k} + \frac{y_k y_k^t}{y_k^t s_k} \quad (3D.32)$$

Onde:

$$s_k = w_{k+1} - w_k, \text{ e } y_k = g_{k+1} - g_k, y_k^t s_k > 0. \bar{G}_{k+1} \text{ é simétrica, positiva e definida.}$$

O algoritmo BFGS requer mais computação, em cada iteração, e mais armazenamento do que os métodos de gradiente conjugado, embora convirja geralmente, em menos iterações. A dimensão da Hessiana aproximada $\bar{G}_k$ é $n \times n$, onde n é igual ao número de pesos e biases da rede. Para redes muito grandes os

métodos do gradiente conjugado são preferíveis aos métodos quase-Newton em termos computacionais, porque tem uma complexidade computacional menor do que os métodos quase-Newton (HAYKIN, 2007). Para redes menores, no entanto, o algoritmo BFGS pode ser uma alternativa eficiente.