

Universidade Estadual de Campinas Faculdade de Engenharia Química Departamento de Processos Químicos

APLICAÇÃO DE TÉCNICAS ESTATÍSTICAS MULTIVARIADAS E DE REDES NEURAIS NA MODELAGEM DE UM SISTEMA DE TRATAMENTO DE EFLUENTES INDUSTRIAIS.

Karla Patrícia Santos Oliveira Rodríguez Esquerre M.Sc. em Engenharia Química FEQ/UNICAMP

> Prof. Milton Mori Orientador

Prof. Roy Edward Bruns Co-orientador

Campinas, 14 de maio de 2003.

UNICAMP BIBLIOTECA CENTRAL SEÇÃO CIRCULANTE



BIBID 296186

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

-

CM00186319-1

R618a	Rodríguez Esquerre, Karla Patrícia Santos Oliveira Aplicação de técnicas estatísticas multivariadas e de redes neurais na modelagem de um sistema de tratamento de efluentes industriais / Karla Patrícia Santos Oliveira Rodríguez Esquerre Campinas, SP: [s.n.], 2003.
	Orientador: Milton Mori e Roy Edward Bruns. Tese (doutorado) - Universidade Estadual de Campinas, Faculdade de Engenharia Química.
	1. Modelagem de dados. 2. Águas residuais. 3. Industria de celulose. 4. Redes neurais. 5. Análise de regressão. I. Mori, Milton. II. Bruns, Roy Edward. III. Universidade Estadual de Campinas. Faculdade de Engenharia Química. IV. Título.



000 10 YOS X

Universidade Estadual de Campinas Faculdade de Engenharia Química Departamento de Processos Químicos

APLICAÇÃO DE TÉCNICAS ESTATÍSTICAS MULTIVARIADAS E DE REDES NEURAIS NA MODELAGEM DE UM SISTEMA DE TRATAMENTO DE EFLUENTES INDUSTRIAIS.

Tese de Doutorado apresentada à Faculdade de Engenharia Química como parte dos requisitos exigidos para a obtenção do título de Doutor em Engenharia Química.

Karla Patrícia Santos Oliveira Rodríguez Esquerre M.Sc. em Engenharia Química FEQ/UNICAMP

> Prof. Milton Mori Orientador

Prof. Roy Edward Bruns Co-orientador

Campinas, 14 de maio de 2003.

Tese de doutorado defendida por Karla Patrícia Santos Oliveira Rodríguez Esquerre e aprovada no dia 14 de Maio de 2003 pela banca examinadora constituída pelos doutores:

en lees

Prof. Dr. Milton Mori Faculdade de Engenharia Química Universidade Estadual de Campinas – UNICAMP

Profa. Dra. Liliane Maria Ferrareso Lona Faculdade de Engenharia Química Universidade Federal de Campinas - UNICAMP

Prof. Dr. Roberto de Campos Giordano Departamento de Engenharia Química Universidade Federal de São Carlos - UFSCar

u lau.

Prof. Dr. Roberto Guardani Departamento de Engenharia Química Escola Politécnica da Universidade de São Paulo - USP

d l

Prof. Dr. Carlos Alberto Ubirajara Gontarski Departamento de Engenharia Química Universidade Federal do Paraná - UFPR

Este exemplar corresponde à versão final da Tese de Doutorado em Engenharia Química.

refeed Orientador

Dedico este trabalho ao meu esposo - Vitaly, aos meus pais – Roseana e Alencar – aos meus irmãos – Erika, Line, Diego e Polly e à minha avó – Benedita, por todo o seu amor e apoio.

Agradecimentos

- À minha família que tanto me incentivou na escolha deste caminho.
- Ao meu admirável orientador pela confiança, apoio e amizade. Agradeço ainda por proporcionar todos os meios necessários para o desenvolvimento deste trabalho assim como para realizar várias outras atividades que contribuíram para o meu desenvolvimento profissional e pessoal.
- Ao meu co-orientador pela amizade e pelo apoio, incentivo e segurança transmitidos. Sintome realmente honrada por ter trabalhado durante os últimos cinco anos sob sua orientação.
- Ao meu esposo, Vitaly Félix Rodríguez Esquerre por todo o apoio, compreensão e auxílio quanto ao entendimento das técnicas, programas e *softwares*.
- Ao amigo Edson Guaraci Fujita pela sua presença sempre constante e amizade. Agradeço ainda pelo indispensável apoio na etapa final de conclusão deste trabalho.
- Ao professor Dale E. Seborg, pelo apoio e conhecimentos transmitidos durante o estágio desenvolvido na University of California, Santa Barbara (UCSB), de fevereiro a agosto de 2002.
- Aos amigos e estatísticos, M.Sc. Roberto C. Colacioppo, Dra. Celeste M. Díaz Cónsul e M.Sc. Carlos Contreras, por todo o auxílio quanto à análise estatística de dados.
- Aos professores da UNICAMP, em especial a profa. Liliane M. Ferrareso Lona (FEQ) e prof.
 Fernando Von Zuben (FEEC), por contribuírem com valiosas orientações e conhecimentos para elaboração deste trabalho. Agradeço ainda à Coordenadoria da Pós-graduação da FEQ, nas pessoas da profa. Telma Franco e dos estagiários Daniel e Alexandre, por todo apoio.
- Aos amigos do Laboratório de Modelagem e Simulação de Processos Químicos (LMSPQ), em especial à Anna Ritta Weber, Alexandre de Paula Peres, Fabiola Moreno e Wesley Marinho, por todo carinho, apoio e companheirismo.
- À International Paper of Brazil, antiga Champion Indústria de Papel e Celulose S/A, pelo fornecimento de dados.
- À Fundação de Amparo à Pesquisa do Estado de São Paulo FAPESP, pelo indispensável apoio financeiro.
- Aos amigos Elenise B. de Moraes, Mário E. Torres, Camile Florido, Daniela C. de Amorim,
 Ana Cláudia Tresmondi, Joseane S. Luna e Sra. Jaguaraci Barros, pela inestimável amizade.

Resumo

Um dos maiores problemas na modelagem e controle de processos de tratamento de efluentes é a construção de modelos confiáveis. Para estes tipos de processos, o desenvolvimento de modelos detalhados baseados em princípios fundamentais e em estudos cinéticos é muito difícil, caro e demanda tempo. A lagoa aerada é um exemplo típico de um processo difícil de ser modelado e controlado. Seu afluente é variável (quali e quantitativamente); a população de microoganismos varia com o tempo (em quantidade e em número de espécies); o conhecimento do processo é escasso e há poucos analisadores online.

A quantidade de matéria orgânica presente é medida através da demanda química de oxigênio (DQO) e da demanda bioquímica de oxigênio (DBO). Elas são variáveis chaves do processo, indicadoras da qualidade do efluente tratado. Tem-se considerável interesse em ter-se um modelo de entrada-saída para predição da DBO, pois é necessário um tempo de cinco dias para análise em laboratório, e um significante tempo de residência na lagoa.

Recentemente, alguns trabalhos relacionados ao uso das técnicas de redes neurais artificiais na modelagem de bioprocessos têm sido publicados. Entretanto, pouca atenção tem sido dada às características dinâmicas destes processos.

O objetivo principal deste trabalho é desenvolver um modelo que forneça predições com acurácia da DBO de entrada e saída da lagoa aerada da *International Paper of Brazil*. Modelos dinâmicos e estacionários foram desenvolvidos utilizando as técnicas de redes neurais artificiais e de regressão linear multivariada. Medidas de qualidade da água, tais como, a DBO, DQO, pH, condutividade, cor e temperatura, adicionalmente às informações do processo de produção (produção de papel e celulose), medidas durante um período aproximado de quatro anos foram usadas no desenvolvimento dos modelos.

As vantagens e desvantagens das técnicas de redes neurais e de regressão linear multiariada, em relação às suas habilidades de modelar um sistema complexo e multivariado, são verificadas e descritas detalhadamente. A redes perceptron de múltiplas camadas forneceram resultados de predição consideravelmente superiores, mesmo quando um número limitado de amostras foi utilizado para a construção dos modelos.

xiii

Abstract

One of the most difficult problem in the modeling and control of wastewater treatment processes is the construction of reliable process models. For these processes, the development of detailed models based on fundamental principles and intense kinetic studies is very difficult, expensive and time consuming. The aerated lagoon is a common example of such an environmental process. Its inflow is variable (both in quality and quantity); the population of microorganisms varies over time, both in quantity and in the number of species; process knowledge is scarce and the few online analyzers available tend to be unreliable.

The amount of organic matter present is measured as either the biochemical oxygen demand (BOD) or the chemical oxygen demand (COD). They are key process variables in water quality. It is very desirable to have a reasonably accurate input-output model for BOD prediction because there is a five-day delay in its laboratory analysis, and a significant hydraulic residence time delay in the aerated lagoon.

Recently, some papers using artificial neural networks (ANNs) in modeling biological wastewater treatment processes have been published. But surprisingly, very little attention has been paid to the dynamic characteristics of these systems.

The main objective of this study is to develop an estimation model that provides accurate predictions of the biochemical oxygen demand (BOD) of the input and output streams of an aerated lagoon at a pulp and paper mill operated by International Paper of Brazil. Steady-state and dynamic predictive models are presented based on both ANNs and linear multivariate regression techniques. Water quality measurements – BOD, COD, flow rate, pH, conductivity, color, temperature –, together with milling process information – pulp and paper production –, over a four-year period are used to develop the models.

The advantages and drawbacks of both neural networks and multivariate linear regression techniques in their ability to model complex and multivariate processes are verified and described in details. The ANN models were slightly more accurate but both models provide reasonably accurate predictions of BOD, even for a system that presents operational data limitation with a large number of missing values.

Sumário

	Dedicatória	ix
	Agradecimentos	xi
	Resumo	xiii
	Abstract	xv
	Lista de figuras	xxi
	Lista de tabelas e quadros	xxv
	Nomenclatura	xxvii
1	Introdução	1
	1.1 Justificativa da escolha do tema	1
	1.2 Objetivos do trabalho	4
	1.2.1 Objetivos gerais	4
	1.2.2 Objetivos específicos	4
2	Descrição do Processo	7
	2.1 Problemática ambiental	7
	2.2 Demanda química e bioquímica de oxigênio	9
	2.3 Sistema de tratamento de efluentes da International Paper of Brazil	10
	2.3.1 Estatística básica dos parâmetros de monitorização da IPB	12
	2.3.2 Pré-processamento de dados	13
	2.4 Considerações finais sobre o capítulo	14
3	Técnicas estatísticas multivariadas	15
	3.1 Regressão linear múltipla (MLR)	15
	3.1.1 Introdução	15
	3.1.2 Um breve histórico	15
	3.1.3 Desenvolvimento matemático	16
	3.1.4 Estimativa dos parâmetros do modelo	17

	3.1.5 Teste de hipóteses em regressão linear	20
	3.1.6 Intervalo de confiança sobre a resposta média	26
	3.1.7 Predição de novas observações	27
	3.1.8 Avaliação da eficiência do modelo de regressão	30
	3.2 Análise dos componentes principais	33
	3.2.1 Introdução	33
	3.2.2 Um breve histórico	33
	3.2.3 Desenvolvimento matemático	33
	3.2.4 Regressão de componentes principais (PCR)	36
	3.3 Mínimos quadrados parciais	37
	3.3.1 Introdução	37
	3.3.2 Um breve histórico	38
	3.3.3 Desenvolvimento matemático	38
	3.3.4 Regressão por mínimos quadrados parciais (PLSR)	41
	3.4 Considerações finais sobre o capítulo	42
		45
4	Redes neurais artificiais	45
4	Redes neurais artificiais	45 45
4	Redes neurais artificiais 4.1 Conceitos gerais 4.1.1 Introdução	45 45 45
4	Redes neurais artificiais 4.1 Conceitos gerais 4.1.1 Introdução 4.1.2 Um breve histórico	45 45 45 45
4	Redes neurais artificiais 4.1 Conceitos gerais 4.1.1 Introdução 4.1.2 Um breve histórico 4.1.3 Redes neurais biológicas	45 45 45 45 47
4	Redes neurais artificiais 4.1 Conceitos gerais 4.1.1 Introdução 4.1.2 Um breve histórico 4.1.3 Redes neurais biológicas 4.1.4 Neurônio artificial e funções de ativação	45 45 45 45 47 48
4	Redes neurais artificiais 4.1 Conceitos gerais 4.1.1 Introdução 4.1.2 Um breve histórico 4.1.3 Redes neurais biológicas 4.1.4 Neurônio artificial e funções de ativação 4.1.5 Arquitetura das redes neurais artificiais	45 45 45 45 47 48 51
4	Redes neurais artificiais 4.1 Conceitos gerais 4.1.1 Introdução 4.1.2 Um breve histórico 4.1.3 Redes neurais biológicas 4.1.4 Neurônio artificial e funções de ativação 4.1.5 Arquitetura das redes neurais artificiais 4.1.6 Função erro	45 45 45 45 47 48 51 52
4	Redes neurais artificiais 4.1 Conceitos gerais 4.1.1 Introdução 4.1.2 Um breve histórico 4.1.3 Redes neurais biológicas 4.1.4 Neurônio artificial e funções de ativação 4.1.5 Arquitetura das redes neurais artificiais 4.1.6 Função erro 4.1.7 Algoritmo backpropagation	45 45 45 47 48 51 52 54
4	Redes neurais artificiais 4.1 Conceitos gerais 4.1.1 Introdução 4.1.2 Um breve histórico 4.1.3 Redes neurais biológicas 4.1.4 Neurônio artificial e funções de ativação 4.1.5 Arquitetura das redes neurais artificiais 4.1.6 Função erro 4.1.7 Algoritmo backpropagation 4.2 Redes neurais perceptron de múltiplas camadas (MLP)	45 45 45 47 48 51 52 54 57
4	Redes neurais artificiais 4.1 Conceitos gerais 4.1.1 Introdução 4.1.2 Um breve histórico 4.1.3 Redes neurais biológicas 4.1.4 Neurônio artificial e funções de ativação 4.1.5 Arquitetura das redes neurais artificiais 4.1.6 Função erro 4.1.7 Algoritmo backpropagation 4.2 Redes neurais perceptron de múltiplas camadas (MLP)	45 45 45 47 48 51 52 54 57 57
4	Redes neurais artificiais 4.1 Conceitos gerais 4.1.1 Introdução 4.1.2 Um breve histórico 4.1.3 Redes neurais biológicas 4.1.4 Neurônio artificial e funções de ativação 4.1.5 Arquitetura das redes neurais artificiais 4.1.6 Função erro 4.1.7 Algoritmo backpropagation 4.2 Redes neurais perceptron de múltiplas camadas (MLP) 4.2.1 Introdução 4.2.2 Algoritmo backpropagation padrão	45 45 45 47 48 51 52 54 57 57
4	Redes neurais artificiais 4.1 Conceitos gerais 4.1.1 Introdução 4.1.2 Um breve histórico 4.1.3 Redes neurais biológicas 4.1.4 Neurônio artificial e funções de ativação 4.1.5 Arquitetura das redes neurais artificiais 4.1.6 Função erro 4.1.7 Algoritmo backpropagation 4.2 Redes neurais perceptron de múltiplas camadas (MLP) 4.2.1 Introdução 4.2.2 Algoritmo backpropagation padrão 4.2.3 Algoritmo de treinamento delta-bar-delta (DBD)	45 45 45 47 48 51 52 54 57 57 57 57
4	Redes neurais artificiais 4.1 Conceitos gerais 4.1.1 Introdução 4.1.2 Um breve histórico 4.1.3 Redes neurais biológicas 4.1.4 Neurônio artificial e funções de ativação 4.1.5 Arquitetura das redes neurais artificiais 4.1.6 Função erro 4.1.7 Algoritmo backpropagation 4.2.1 Introdução 4.2.1 Introdução 4.2.2 Algoritmo backpropagation padrão 4.2.3 Algoritmo de treinamento delta-bar-delta (DBD) 4.2.4 Definição da topologia da rede	45 45 45 47 48 51 52 54 57 57 57 57 59 61

	4.2.6 Avaliação da eficiência do modelo MLP	65
	4.3 Redes neurais functional-link	66
	4.3.1 Introdução	66
	4.3.2 Desenvolvimento matemático	67
	4.3.3 Algoritmo proposto por BILLINGS et al. (1989)	69
	4.3.4 Avaliação da eficiência do modelo FLN	73
	4.4 Considerações finais sobre o capítulo	73
5	Resultados e discussões	77
	5.1 Pré-processamento de dados	77
	5.1.1 Dados 1	77
	5.1.2 Dados 2	81
	5.2 Modelagem	84
	5.2.1 Dados 1 – Predição da DBO de entrada	84
	5.2.2 Dados 1 – Predição da DBO de saída	94
	5.2.3 Dados 2 – Predição da DBO de entrada e saída	105
	5.3 Considerações finais sobre o capítulo	109
6	Considerações finais	111
	7.1 Conclusões	111
	7.2 Perspectivas futuras	113
7	Referências bibliográficas	113
8	Publicações associadas à tese	123
	8.1 Trabalhos publicados	123
	8.2 Trabalhos apresentados	123
	8.3 Trabalho premiado	124
	8.4 Considerações finais sobre o capítulo	125
	Anexo 1: Gráficos de séries temporais e matriz de correlação	127
	Anexo 2: Resultados do pré-processamento de dados via PCA e PLS.	129

Lista de figuras

Figura 2.1: Desenho esquemático do sistema de tratamento de efluentes da IPB.	11
Figura 3.1: Área da função de probabilidade acumulada.	21
Figura 3.2: Região contendo observações para um modelo de regressão de duas variáveis.	28
Figura 3.3: Representação da relação entre os scores da matriz X e Y.	39
Figura 3.4: Desenho esquemático das técnicas MLR, PCR e PLSR, sendo consideradas oito variáveis independentes, uma variável dependente e uma redução de dimensionalidade de oito para três regressores	42
Figura 4.1a: O neurônio biológico.	48
Figura 4.1b: Conexões sinápticas.	48
Figura 4.2: Estrutura de um neurônio artificial.	49
Figura 4.3: Estrutura de uma rede neural, (a) com uma única camada de neurônios; (b) com múltiplas camadas de neurônios.	51
Figura 4.4: Desenho esquemático da superfície da função erro $E(w)$	53
Figura 4.5: Estrutura de uma FLN (functional-link network).	67
Figura 5.1a: Gráfico dos <i>loadings</i> e <i>scores</i> correspondentes às quatro primeiras CPs (Dados 1 – Modelagem estacionária).	78
Figura 5.1b: Gráfico dos <i>loadings</i> e <i>scores</i> correspondentes às quatro primeiras CPs (Dados 1 – Modelagem dinâmica)	78
Figura 5.2: Carta de T ² das componentes principais para o modelo estacionário – Dados 1; (a) Conjunto de modelagem e (b) Conjuntos de validação e teste.	80
Figura 5.3: Carta de T^2 das componentes principais para o modelo dinâmico – Dados 1; (a) Conjunto de modelagem e (b) Conjuntos de validação e teste.	8 1
Figura 5.4: Gráfico dos loadings e scores correspondentes às cinco primeiras CPs (Dados 2).	82
Figura 5.5: Carta de T ² das componentes principais – Dados 2; (a) Conjunto de modelagem; (b) Conjunto de validação (teste).	83
Figura 5.6: Teste de normalidade dos resíduos dos modelos para DBO de entrada – Dados 1; (a) MLR – 5 estacionário; (b) MLR – 6 dinâmico; (c) PLS – 3 estacionário; (d) PLS – 3 dinâmico.	86

Figura 5.7: Relação entre a DBO de entrada predita vs. medida (linha sólida) – Dados 1. As linhas tracejadas superior e inferior indicam o intervalo de predição de 95% para o modelo estacionário MLR – 5.	87
Figura 5.8: Gráfico de séries temporais – Dados 1; (a) DBO de entrada medida e predita para o modelo estacionário MLR – 5; (b) Resíduos – Linhas superior e inferior indicam um intervalo de confiança de 95%.	88
Figura 5.9: Relação entre a DBO de entrada predita vs. medida (linha sólida) – Dados 1. As linhas tracejadas superior e inferior indicam o intervalo de predição de 95% para o modelo estacionário MLP com dez neurônios.	92
Figura 5.10: Gráfico de séries temporais – Dados 1; (a) DBO de entrada medida e predita para o modelo estacionário MLP com dez neurônios; (b) Resíduos – Linhas superior e inferior indicam um intervalo de confiança de 95%.	93
Figura 5.11: Teste de normalidade dos resíduos dos modelos para DBO de saída – Dados 1; (a) MLR – 4 estacionário; (b) MLR – 6 dinâmico; (c) PLS – 3 estacionário; (d) PLS – 3 dinâmico.	96
Figura 5.12: Relação entre a DBO de saída predita vs. medida (linha sólida) – Dados 1. As linhas tracejadas superior e inferior indicam o intervalo de predição de 95% para o modelo dinâmico $MLR - 6$.	98
Figura 5.13: Gráfico de séries temporais – Dados 1; (a) DBO de saída medida e predita para o modelo dinâmico MLR – 6; (b) Resíduos – Linhas superior e inferior indicam um intervalo de confiança de 95%.	99
Figura 5.14: Relação entre a DBO de entrada predita vs. medida (linha sólida) – Dados 1. As linhas tracejadas superior e inferior indicam o intervalo de predição de 95% para o modelo dinâmico MLP com três neurônios.	103
Figura 5.15: Gráfico de séries temporais – Dados 1; (a) DBO de saída medida e predita para o modelo dinâmico MLP com três neurônios; (b) Resíduos – Linhas superior e inferior indicam um intervalo de confiança de 95%.	104
Figura 6.1: Desenho esquemático das ferramentas utilizadas na construção dos modelos preditivos.	112
Figura 1.1A: Gráficos de séries temporais das variáveis preditoras e preditas da lagoa aerada da IPB (Set/1996 a Jul/2000), onde cada amostra representa o valor individual ou médio das análises realizadas em um dia.	127

Lista de tabelas e quadros

Tabela 2.1: Estatística básica dos parâmetros monitorados pela IPB.	12
Tabela 3.1: Análise da variância para testar a significância de uma regressão.	23
Tabela 5.1: Resultados MLR e PLSR obtidos para o conjunto de modelagem – DBO de entrada (Dados 1).	85
Tabela 5.2: Resultados dos modelos MLR e PLSR obtidos para os conjuntos de validação e teste – DBO de entrada (Dados 1).	87
Tabela 5.3: Resultados FLN (ordem 1) obtidos para o conjunto de modelagem – DBO de entrada (Dados 1).	89
Tabela 5.4: Resultados FLN (ordem 2) obtidos para o conjunto de modelagem – DBO de entrada (Dados 1).	90
Tabela 5.5: Resultados FLN (ordem 1) obtidos para os conjuntos de validação e teste – DBO de entrada (Dados 1).	90
Tabela 5.6: Resultados FLN (ordem 2) obtidos para os conjuntos de validação e teste – DBO de entrada (Dados 1).	90
Tabela 5.7: Resultados MLP obtidos para o conjunto de modelagem – DBO de entrada (Dados 1).	92
Tabela 5.8: Resultados MLP obtidos para os conjuntos de validação e teste – DBO de entrada (Dados 1).	92
Tabela 5.9: Resultados MLR e PLSR obtidos para o conjunto de modelagem – DBO de saída (Dados 1).	95
Tabela 5.10: Resultados dos modelos MLR e PLSR obtidos para os conjuntos de validação e teste – DBO de saída (Dados 1).	97
Tabela 5.11: Resultados FLN (ordem 1) obtidos para o conjunto de modelagem – DBO de saída (Dados 1).	1 00
Tabela 5.12: Resultados FLN (ordem 2) obtidos para o conjunto de modelagem – DBO de saída (Dados 1).	100
Tabela 5.13: Resultados FLN (ordem 1) obtidos para os conjuntos de validação e teste – DBO de saída (Dados 1).	101

Tabela 5.14: Resultados FLN (ordem 2) obtidos para os conjuntos de validação e teste – DBO de saída (Dados 1).	101
Tabela 5.15: Resultados MLP obtidos para o conjunto de modelagem – DBO de saída (Dados 1).	102
Tabela 5.16: Resultados MLP obtidos para os conjuntos de validação e teste – DBO de saída (Dados 1).	102
Tabela 5.17: Resultados MLR e PLSR obtidos para o conjunto de modelagem – DBO de entrada e saída (Dados 2).	106
Tabela 5.18: Resultados MLR e PLSR obtidos para o conjunto de validação – DBO de entrada e saída (Dados 2).	106
Tabela 5.19: Resultados FLN e MLP obtidos para o conjunto de modelagem – DBO de entrada e saída (Dados 2).	107
Tabela 5.20: Resultados FLN e MLP obtidos para o conjunto de validação – DBO de entrada e saída (Dados 2).	108
Tabela 5.21: Resultados dos melhores modelos obtidos para os conjuntos de validação e teste – DBO de entrada (Dados 1).	109
Tabela 5.22: Resultados dos melhores modelos obtidos para os conjuntos de validação e teste – DBO de saída (Dados 1).	109
Tabela 1.1A: Matriz de correlação	128
Tabela 2.1A: Variância explicada e acumulada das componentes principais e variáveis latentes (Dados 1 – Modelagem estacionária)	129
Tabela 2.2A: Variância explicada e acumulada das componentes principais e variáveis latentes (Dados 1 – Modelagem dinâmica)	129
Tabela 2.3A: Variância explicada e acumulada das componentes principais e variáveis latentes (Dados 2).	130
Tabela 2.4A: Loadings das componentes principais (Dados 1 – Modelagem estacionária)	130
Tabela 2.5A: Loadings das componentes principais (Dados 1 – Modelagem dinâmica)	131
Tabela 2.6A: Loadings das componentes principais (Dados 2).	131
Tabela 2.7A: Loadings das variáveis latentes - DBO de entrada (Dados 1 - Modelagem estacionária).	132

Tabela 2.8A: Loadings das variáveis latentes – DBO de entrada (Dados 1 – Modelagem dinâmica).	132
Tabela 2.9A: Loadings das variáveis latentes - DBO de entrada (Dados 2).	133
Tabela 2.10A: Loadings das variáveis latentes – DBO de saída (Dados 1 – Modelagem estacionária). Tabela 2.11A: Loadings das variáveis latentes – DBO de saída (Dados 1 – Modelagem dinâmica).	133 134
Tabela 2.12A: Loadings das variáveis latentes – DBO de saída (Dados 2).	134
Quadro 4.1: Expansão polinomial de grau seis.	69

Nomenclatura

GERAL

LATINAS

R ²	coeficiente de correlação
R^2 adj.	coeficiente de correlação ajustado
F	teste F proposto por BARROS et al. (1996)
р	número de parâmetros ajustados pelo modelo
n	número de neurônios intermediários

ABREVIAÇÕES

Análise da Variância.
produção de celulose [ton/dia]
chuva [mm/dia]
Congresso Brasileiro de Engenharia Química
condutividade [[µS/cm a 20°C]
cor [mg/L]
componente principal
conjunto de dados contendo um total de 1092 amostras
conjunto de dados contendo um total de 79 amostras
algoritmo delta-bar-delta
demanda bioquímica de oxigênio na entrada da lagoa aerada [mg/L]
demanda bioquímica de oxigênio na saída da lagoa aerada [mg/L]
graus de liberdade
desvio padrão
demanda química de oxigênio [mg/L]
rede functional-link
Faculdade de Engenharia Química
Instituto de Matemática, Estatística e Computação Científica
intervalo de predição

IPB	International Paper of Brazil
MLR	regressão linear múltipla
MLP	perceptron de múltiplas camadas
MQ_E	média quadrática do resíduo
MQ_R	média quadrática da regressão
MSE	erro quadrático médio
NAM	nitrogênio amoniacal [mg/L]
NN	nitrogênio nitrato [mg/L]
PAP	produção de papel [ton/dia]
PCA	análise das componentes principais
PCR	regressão das componentes principais
pН	potencial hidrogênio iônico
PLS	mínimos quadrados parciais ou projeção de variáveis latentes
PLSR	regressão por PLS
RNA	redes neurais artificiais
SQ_E	soma quadrática dos resíduos
SQ_R	soma quadrática da regressão
SQ_T	soma quadrática total
SST	sólidos suspensos totais [mg/L]
Т	temperatura [°C]
UNICAMP	Universidade Federal de Campinas
VA	variância acumulada
VAZ	vazão na entrada da lagoa aerada [m ³ /dia]
VE	variância explicada
VL	variável latente

ESPECÍFICA DO CAPÍTULO

Capítulo 3: Técnicas estatísticas multivariadas

LATINAS

e	resíduo
Ε	esperança
Е	erro na estimativa das componentes principais no espaço X
F	erro na estimativa das componentes principais no espaço Y
F	teste F (valor calculado) proposto por BARROS et al. (1996)
f	teste F (valor tabelado) proposto por BARROS <i>et al.</i> (1996)
k	número de variáveis regressoras
H ₀	hipótese nula
H_i	hipótese não nula
K	quantidade de correlação num conjunto de dados (TODESCHINE, 1997)
KL	número máximo de componentes principais (TODESCHINE, 1997)
KP	número mínimo de componentes principais (TODESCHINE, 1997)
р	número de parâmetros ajustados
Р	probabilidade
n	número de amostras
x_j	variáveis preditoras ou independentes j
у	variáveis preditas ou dependentes
\overline{y}	valor médio da variável predita
ŷ	valor estimado da variável predita
e	erro da regressão
Q	loadings das componentes principais no espaço Y
R ² adj.	coeficiente de determinação múltipla ajustado
s^2	variância de uma amostra
S	desvio padrão de uma amostra
t	teste t de student (valor tabelado)
T_0	teste t de student (valor calculado)

T^2	estatística T ² de Hotelling
Т	scores das componentes principais
v	variância
W	loadings das componentes principais

<u>GREGAS</u>

α	parâmetro comparativo ao p-valor para inclusão e exclusão de variáveis
$oldsymbol{eta}_j$	coeficiente de regressão da variável preditora j
$\hat{oldsymbol{eta}}_{j}$	coeficiente de regressão estimado da variável preditora j
$\Phi\left(z_0 ight)$	função de distribuição cumulativa de probabilidade
λ	autovalor
μ	resposta média de y no ponto x_0
σ^2	variância da população
σ	desvio padrão de uma população

Capítulo 4: Redes neurais artificiais

LATINAS

Α	matriz triangular superior
E _{med}	erro quadrado médio ou soma dos erros quadrados
E(n)	valor instantâneo da soma de erros quadrados
Etotal	soma total de erros quadrados
$e_j(n)$	sinal do erro na saída do neurônio j
$d_j(n)$	resposta desejada da variável predita no neurônio j no tempo n
$h(\cdot)$	função de ativação da rede FLN
L	camada intermediária ou de saída
n	tempo discreto
Ν	número de amostras
m	número de entradas da rede

М	número de monômios gerados por expansão polinomial
mg	grau do monônio
m _x	tamanho do vetor de entrada da rede FLN
x_i	variável de entrada i da rede
$y_j(n)$	saída do neurônio j no tempo n
W _{kj}	peso sináptico da sinápse j pertencente ao neurônio k
e	símbolo que "pertence a"
	apóstrofe, diferenciação em relação ao argumento
Z	matriz ortogonal

GREGAS

taxa de treinamento inicial
valor limite a que pode chegar a taxa de treinamento
constante multiplicativa
derivada
média ponderada do termo do gradiente na variação do peso.
gradiente local de neurônio j no tempo n
pequena variação aplicada ao peso w
erro do algoritmo proposto por BILLINGS et al. (1989)
constante aditiva
constante de amortecimento
parâmetro da taxa de aprendizagem
função de ativação do neurônio j

~

CAPÍTULO 1

INTRODUÇÃO

1.1 Justificativa da escolha do tema

Juntamente com o aumento da atividade industrial, cresce também a demanda por meios mais eficientes de tratar os resíduos não aproveitáveis gerados durante os processos de produção. O número considerável de calamidades ecológicas tem feito a população se dar conta que a prevenção da poluição é necessária para a sobrevivência da humanidade.

Um problema ambiental óbvio e fundamental de qualquer localidade é a disposição e tratamento de resíduos líquidos. É imperativo que corporações nacionais e internacionais adotem práticas contínuas de gerenciamento ambiental objetivando a prevenção da poluição nas unidades de processo que elas possuem ou operem. Quando uma indústria ou qualquer fonte geradora não consegue minimizar a geração de tais resíduos na fonte até níveis aceitáveis, uma ou mais unidades de tratamento devem ser instaladas.

Os efluentes da indústria de papel e celulose são motivos de grande preocupação, visto que podem conter produtos químicos nocivos ao meio ambiente que são introduzidos durante a produção (YANG, 1996). Esta preocupação tem se refletido na necessidade de se desenvolver sistemas de prevenção e controle para situações críticas.

Embora seja muito importante garantir a qualidade dos efluentes tratados antes de descartá-los, o controle e operação apropriados de um sistema de tratamento de efluentes não são tarefas fáceis. A lagoa aerada é um exemplo típico de um processo dificil de ser controlado. Seu efluente de entrada é variável quali e quantitativamente; a população de microorganismos varia com o tempo (em quantidade e em número de espécies); o conhecimento do processo é escasso e poucos analisadores são online. Além disso, o tempo requerido para realizar as análises laboratoriais dos parâmetros indicativos da qualidade do efluente pode levar minutos, horas ou até mesmo dias. Perturbações externas, tal como o clima, dificultam ainda mais o controle deste tipo de processo (COTE *et al.*, 1995).

A demanda bioquímica de oxigênio (DBO) é um dos principais parâmetros utilizados para indicar a qualidade ou o nível de degradação de um efluente ou um corpo receptor. Devido ao tempo de cinco dias necessário para realizar sua análise em laboratório, as medidas da DBO chegam muito tarde, dificultando a atuação no processo e a consequente manutenção de seus valores em níveis apropriados. Evidencia-se portanto a necessidade de construção de modelos de predição para estimativa imediata deste parâmetro de qualidade.

A modelagem tradicionalmente utilizada em bioprocessos é baseada em equações de balanço e taxas de crescimento biológico, consumo de substrato e formação de produtos. Entretanto, como as reações microbiológicas em união com as interações ambientais são em geral não lineares, variantes com o tempo e de natureza complexa (HAMODA *et al.*, 1999; LEE e PARK, 1999), tem-se comprovado que estes métodos são em geral ineficientes (HAMODA *et al.*, 1991; LEE e PARK, 1999; COTE *et al.*, 1995; STEYER *et al.*, 1997).

As redes neurais têm apresentado bons resultados como ferramentas de modelagem para sistemas cujos mecanismos internos não são suficientemente conhecidos (MORRIS *et al.*, 1994), como é o caso do tratamento de resíduos (HÄCK *et al.*, 1996 e PU *et al.*, 1995). Elas possuem a propriedade de, uma vez apresentadas a um conjunto de dados de entrada e saída, "aprender" as relações que estiverem implícitas entre os dados contidos nesse conjunto.

As redes neurais perceptron de múltiplas camadas (MLP) têm sido utilizadas com sucesso na modelagem de sistemas biológicos de tratamento de efluentes, como pode ser observado nos trabalhos de COTE et al. (1995), GONTARSKI et al. (2000a), HÄCK e KOHNE (1996), HAMODA et al. (1999), LEE e PARK (1999), OLIVEIRA-ESQUERRE et al. (2002; 2003a; 2003c), PU e HUNG (1995), WILCOX et al. (1995) e ZHAO et al. (1997). Entretanto, as principais desvantagens de sua utilização é que elas requerem uma quantidade considerável de dados para a construção dos modelos e o seu mapeamento não-linear pode levar a mínimos locais. Uma estrutura neural que tem sido pouco explorada em bioprocessos é a rede *functional-link* (CHEN e BILLINGS, 1992; HARADA et al., 2002). Nesta, uma expansão polinomial das entradas da rede é realizada e os termos resultantes são combinados linearmente para a formação do modelo.

Quanto maior o número de variáveis que representam um sistema, maior o número de parâmetros a serem ajustados pela rede e conseqüentemente maior o tempo de convergência. Assim, várias técnicas de compressão de dados têm sido pesquisadas, tais como a análise de componentes principais (PCA) e mínimos quadrados parciais (PLS). As técnicas PCA e PLS,

2

além de serem capazes de reduzir a dimensionalidade das variáveis preditoras, são utilizadas na ortogonalização das mesmas. Através do uso de dados ortogonais como entrada da rede é possível otimizar o seu tamanho e consequentemente diminuir o número de parâmetros a serem ajustados (KANJILAL, 1995). Os méritos do uso de RNA em combinação com a técnica estatística de PCA podem ser observados nos trabalhos de CANCILLA e FAND (1996), DUTTA *et al.* (1998), HOLCOLM e MORARI (1992), KOMPANY-ZARED *et al.* (1999), OLIVEIRA-ESQUERRE *et al.* (2002), PONTON e KLEMES (1993) e ZUPAN e GASTEIGER (1993).

Alguns métodos estatísticos clássicos de regressão podem também ser utilizados para construção de modelos de predição entrada-saída, que são em geral resolvidos através de redes neurais artificiais (BAUGHMAN, 1995). Tais métodos podem ser utilizados para aproximar relações complexas sobre regiões limitadas das variáveis preditoras (BAFFI *et al.*, 1999). Esta afirmação é baseada na suposição de que relações não lineares podem ser aproximadas localmente por modelos lineares. A técnica de regressão por mínimos quadrados parciais (PLSR) tem-se mostrado uma poderosa ferramenta na solução de problemas onde os dados são ruidosos e altamente correlacionados e ainda, quando apenas um número limitado de amostras está disponível para a construção dos modelos (BAFFI *et al.*, 1999; OLIVEIRA-ESQUERRE, 2003b). Para métodos estatísticos convencionais de regressão, como regressão linear múltipla – MLR, o uso de variáveis ortogonais também é desejável visto que reduz o erro de modelagem.

As pesquisas realizadas nos últimos anos têm contribuído significativamente para o melhor entendimento da natureza do aprendizado das redes neurais e suas propriedades matemáticas (UTOJO e BAKSHI, 1995), e das relações existentes entre estas e técnicas estatísticas clássicas de regressão (POGGIO e GIROSI, 1990; UTOJO e BAKSHI, 1995). Entretanto, os esforços para análise comparativa entre estas duas técnicas têm sido direcionados para casos aplicados à modelagem de funções analíticas (BAKSI e CATTERJEE, 1998; BAKSI e UTOJO, 1999; UTOJO e BAKSHI, 1995) ou modelagem de dados simulados de processos (MELEIRO, 2002). Uma revisão unificada das redes neurais e técnicas de regressão multivariadas aplicadas à construção de modelos de predição para casos reais, mas especificamente, para bioprocessos, é difícil de ser encontrada. Diante desta perspectiva, e afirmando que não se deve querer aplicar tais tecnologias indiscriminadamente, deseja-se estabelecer alguns critérios para que se possa avaliar melhor quando podem ser obtidos resultados satisfatórios com as redes neurais ou através de métodos estatísticos clássicos.

3

Nesta tese pretende-se realizar uma análise comparativa da performance dos modelos construídos utilizando as técnicas de redes neurais artificiais e de regressão linear, para estimar a DBO de entrada e saída da lagoa aerada da *International Paper of Brazil*. Uma descrição detalhada do processo a ser modelado e das técnicas de modelagem e de pré-tratamento de dados utilizadas é apresentada.

1.2Objetivos do trabalho

1.2.1 Objetivos gerais

Destacam-se os principais objetivos gerais do trabalho desenvolvido como sendo:

- a) Aprender a utilizar as técnicas de regressão linear multivariada e de redes neurais artificiais;
- b) Avaliar as vantagens e desvantagens de uso das técnicas clássicas de regressão e de redes neurais na modelagem de um caso real;
- c) Estabelecer uma visão crítica sobre a aplicação de técnicas estatísticas de redução de dimensionalidade e ortogonalização para auxiliar o mapeamento das redes neurais;
- d) Realizar uma revisão unificada e detalhada das redes neurais e técnicas de regressão multivariadas aplicadas à construção de modelos de predição para um caso real (este é o maior aporte desta tese), e
- e) Avaliar de maneira crítica a utilização de redes neurais e de técnicas clássicas de regressão a processos biológicos de tratamento de efluentes, indicando possíveis soluções aos problemas encontrados.

1.2.2 Objetivos específicos

O objetivo específico deste trabalho é fornecer subsídios à *International Paper of Brazil* para prever situações futuras do despejo com base no parâmetro de qualidade da demanda bioquímica de oxigênio, com o propósito de:

- a) Auxiliar a monitorização da qualidade do efluente tratado, e
- b) Aumentar o nível de controle sobre o processo.

Este trabalho visou o desenvolvimento de modelos de predição da DBO de entrada da lagoa aerada, adicionalmente aos modelos predição da DBO de saída da mesma. Desta forma, os modelos poderão ser utilizados de forma preventiva, auxiliando no ajuste de parâmetros controlados (como a carga de nutrientes e a taxa de aeração da lagoa), proporcionando assim meios de manter os valores da DBO de saída em níveis desejados.

Desta forma, a proposta de trabalho envolve os seguintes itens:

- a) Técnicas: Regressão linear multivariada e redes neurais artificiais.
- b) Processo: Lagoa aerada para tratamento de efluentes industriais da International Paper of Brazil.
- c) Variáveis a serem preditas: DBO de entrada e saída da lagoa aerada.
- d) Dados: Dados históricos fornecidos pela indústria.
- e) Programas e Softwares:
- Neurosolutions for Excel, para preparação e aleatorização dos dados.
- Minitab e Statistica, para construção dos modelos de regressão linear multivariada.
- Compilador Matlab, para utilização do programa desenvolvido por HENRIQUE (1999) para treinamento e geração da rede *functional-link*; construção de programas para estimativa e análise de parâmetros dos modelos.
- Neurosolutions, para treinamento e geração da rede neural perceptron de múltiplas camadas.
- Microcal Origin 6.0, para análise de dados e construção dos gráficos.

CAPÍTULO 2 DESCRIÇÃO DO PROCESSO

2.1 Problemática ambiental

A atividade industrial como um todo, sendo uma atividade essencial na sociedade, gera os mais altos índices de poluição. Os poluentes industriais que mais preocupam são os orgânicos, que marcam presença na maior parte das atividades industriais e domésticas. As características ecológicas de uma água receptora de despejos são afetadas pela destruição de qualquer grupo de microorganismos e a quantidade e concentração dos despejos de uma determinada indústria variam dentro de amplos limites.

Em condições normais, um rio é capaz de receber uma carga orgânica apreciável, eliminando-a gradativamente mediante ações naturais que se processam ao longo de vários quilômetros do seu percurso, entretanto, quando o curso de água fica sem oxigênio dissolvido suficiente, ocorre a morte dos organismos aeróbios e a aplicação desta água fica praticamente impossibilitada para diversos usos e finalidades.

No território brasileiro, as águas doces, salobras e salinas são protegidas pela legislação brasileira, de acordo com a portaria 36/GM de 18 de junho de 1986 (CONSELHO NACIONAL DO MEIO AMBIENTE, 1990). No âmbito do estado de São Paulo se estabelecem padrões para os efluentes de qualquer fonte poluidora quanto a certas características tais como pH, temperatura, materiais sedimentáveis, demanda bioquímica de oxigênio de cinco dias (DBO₅) e concentrações de substâncias tóxicas. O estabelecimento de tais padrões obrigam as indústrias e quaisquer geradores de águas residuárias a possuírem sistemas de tratamento de efluentes líquidos, de forma a atender às exigências legais; o método de tratamento está relacionado intimamente com as características do efluente.

Basicamente, no chamado tratamento primário são empregadas operações unitárias como peneiramento, sedimentação e filtração para a remoção de sólidos sobrenadantes ou sedimentáveis normalmente encontrados em efluentes líquidos. No tratamento secundário são usados processos químicos ou biológicos para eliminar grande parte da carga de materiais orgânicos. No tratamento terciário usam-se combinações de operações e processos unitários para

7

remover outros constituintes que não foram removidos no tratamento secundário, tais como nitrogênio e fósforo (METCALF & EDDY, 1991). Os efluentes inorgânicos são tratados preferencialmente por processos físico-químicos. Os orgânicos podem receber tratamento por processo físico-químico ou biológico. Os processos biológicos são adequados quando o efluente orgânico apresenta características biodegradáveis.

Os processos biológicos que ocorrem na presença de oxigênio livre, denominados aeróbios, são bem conhecidos e são os mais utilizados. A implantação de sistemas com processos aeróbios oferece, portanto, poucos riscos de investimento, pois têm custo baixo de construção e apresentam operação e manutenção simples e econômicas. No entanto, são sistemas que ocupam áreas consideráveis; os equipamentos necessários, como os aeradores, demandam energia, e o lodo, gerado como subproduto, ainda é passível de decomposição por ação biológica. O objetivo do tratamento biológico de efluentes industriais ou domésticos é coagular e remover sólidos coloidais que não sedimentam, e ainda estabilizar a matéria orgânica. Esta redução da concentração da matéria orgânica é usualmente medida como DBO (demanda bioquímica de oxigênio), TOC (carbono orgânico total), ou DQO (demanda química de oxigênio) (METCALF & EDDY, 1991).

Outros processos de decomposição biológica ocorrem na ausência de oxigênio livre e se denominam como processos anaeróbios. Quando o tratamento é realizado em lagoas anaeróbias pode-se ter problemas estéticos devido à formação de espuma na superfície da água; já, quando realizado em bio-reatores fechados, o processo pode apresentar as vantagens de menor área de ocupação, produção de lodo em menor quantidade e já estabilizado, e produção do gás metano como subproduto, que pode ser utilizado como fonte de energia,

O processo em estudo neste trabalho consiste de uma lagoa aerada. Em termos de redução de DBO, a eficiência de uma lagoa aerada pode ser superior a 95%. Dentre os principais fatores que influenciam na atividade tanto das lagoas aeradas, assim como das lagoas de estabilização, têm-se:

- a) Fatores incontroláveis: temperatura, agitação (ventos), evaporação e chuvas;
- b) Fatores parcialmente controláveis: permeabilidade do fundo, nutrientes e carga poluidora;

c) Fatores relacionados ao projeto de construção da lagoa: localização, número de unidades, disposição das unidades, dimensões, período de detenção e detalhes construtivos (OLIVEIRA, 2000).

Estes fatores exercem influência determinante na eficiência do processo. Um estudo consistente destes fatores deve ser realizado assim como diversos parâmetros devem ser monitorados e controlados de modo a garantir o bom funcionamento das lagoas. A escolha dos parâmetros de monitoração deve ser feita baseada no conhecimento do processo aeróbio. Eles devem ser suficientes para informar a realidade do comportamento da biomassa e como esta reage ante variações das características da água residuária sob tratamento. Por fim, devem ser de facilmente medidas, uma vez satisfeita a exigência anterior.

Os dois principais parâmetros para avaliação do grau de depuração de recursos hídricos, ou da qualidade de um efluente, são a demanda bioquímica de oxigênio (DBO) e a demanda química de oxigênio (DQO). Uma descrição detalhada destes parâmetros é apresentada na Seção 2.2, que segue.

2.2 Demanda química e bioquímica de oxigênio

A DBO baseia-se no princípio de que a matéria orgânica, quando metabolizável, tende a ser metabolizada pelos organismos aeróbios presentes, consumindo o oxigênio presente na água. A DBO pode ser definida como a quantidade de oxigênio requerida pelas bactérias para estabilizar a matéria orgânica "decomponível" sob condições aeróbicas. É essencialmente um bioensaio envolvendo a medida de oxigênio consumido por organismos vivos (principalmente bactérias) enquanto utilizam a matéria orgânica presente no despejo, em condições tão similares quanto possível daquelas que ocorrem na natureza.

As reações oxidativas envolvidas nos testes de DBO são resultado de atividades biológicas e a taxa pela qual as reações processam-se é governada pelo número populacional e pela temperatura que é conduzida a 20°C, que representa um valor médio da temperatura da maioria dos corpos de água. A velocidade dos processos metabólicos a 20°C e as condições do teste são tais que o tempo deve ser medido em dias. Teoricamente um tempo infinito é requerido para a oxidação completa da matéria orgânica, mas para propósitos práticos, a reação é

considerada completa em 20 dias. Porém é adotado um período de incubação de 5 dias, quando já ocorreu de 70 a 80% da oxidação.

A DQO é uma medida equivalente, em oxigênio, na porção de matéria orgânica na amostra que permite a oxidação por um oxidante químico forte, geralmente dicromato, em meio fortemente ácido sob aquecimento. O teste é baseado no fato de que todos os compostos orgânicos, com poucas exceções, podem ser oxidados pela ação de um agente oxidante forte em meio ácido. Uma das limitações entretanto é o fato de que o teste não diferencia matéria orgânica biodegradável e matéria orgânica não biodegradável, a primeira determinada pelo teste de DBO. A vantagem é o tempo de teste, realizado em aproximadamente duas horas, enquanto o teste de DBO requer no mínimo 5 dias (período de incubação).

Não é possível estabelecer relações fixas entre as medidas de DBO e DQO, até que uma determinada amostra seja caracterizada por ambos os parâmetros. De um modo geral, tem-se que se a amostra é constituída de compostos que são oxidados por ambos processos (DBO e DQO) a DQO pode substituir a DBO ou a DQO pode ser usada como indicação da diluição necessária para análise da DBO. Já, se a amostra é caracterizada pela predominância de material oxidável quimicamente, porém não bioquimicamente, a DQO será maior que a DBO. Por outro lado, despejos de destilarias e refinarias têm alta DBO e baixa DQO. Uma descrição detalhada da DBO e DQO, incluindo seus métodos de medição em laboratório, é apresentada no site http://www.hach.com/cs/cod_bod.htm.

2.3 Sistema de tratamento de efluentes da International Paper of Brazil

A International Paper of Brazil (IPB), antiga Champion Papel e Celulose Ltda, está localizada na Rodovia SP 340, Km 171, Mogi Guaçu SP. Esta produz celulose branqueada e papel. Todos os efluentes coletados no seu processo são levados para o seu sistema de tratamento, que consiste de um tratamento primário seguido de um tratamento biológico. O tratamento biológico é realizado em uma lagoa aerada de aproximadamente 169.000 m³ e complementado em um conjunto de cinco lagoas de estabilização cujo volume total gira em torno de 1.500.000 m³. Um desenho esquemático do processo de tratamento da IPB é apresentado na Figura 2.1.



Figura 2.1: Desenho esquemático do sistema de tratamento de efluentes da IPB.

Para o controle do seu sistema de tratamento de efluentes, mais especificamente, de suas lagoas aerada e de estabilização, a IPB monitora diariamente nove parâmetros, sendo eles: demanda química de oxigênio, demanda bioquímica de oxigênio, pH, sólidos suspensos totais, cor, temperatura, condutividade, vazão e chuva. As concentrações de nitrogênio (amoniacal e nitrato), fósforo e sulfato são medidas semanalmente.

Como mencionado anteriormente, o monitoramento e controle de uma lagoa aerada é em geral bastante complicado tendo em vista que, seu afluente e a população de microorganismos é variável e o conhecimento do processo de degradação ainda é escasso. A dificuldade em realizar medidas online limita ainda mais o controle efetivo da qualidade do efluente (HARREMOËS et al.; 1993, LEE e PARK, 1999). As análises laboratoriais requerem minutos, horas ou até mesmo dias para serem realizadas, de acordo com os diferentes parâmetros analisados, por exemplo: sólidos suspensos (trinta minutos), DQO (duas horas) e DBO (cinco dias).

2.3.1 Estatística básica dos parâmetros de monitorização da IPB

Adicionalmente aos dez parâmetros de monitorização da lagoa aerada, dados diários de produção de papel e celulose foram disponibilizados pela IPB para a realização desta pesquisa. Os dados são referentes ao período de junho de 1996 a setembro de 2000.

Gráficos de séries temporais e a matriz de variância dos parâmetros disponibilizados são apresentados no Anexo 1. Tabela 2.1 são apresentadas as estatísticas básicas dos parâmetros disponibilizados pela empresa.

Parâmetro	Média	DP	Mínimo	Máximo	Skewness	Curtose	* Lacunas [%]
DBO _{in}	245	46,2	41	449	0,08	1,56	6,24
DBO _{out}	85,1	25,4	16	1 8 7	0,15	0,43	5,96
DQO	561	104	136	925	-0,16	1,05	6,24
CEL	886	155	0	1112	-3,31	14,2	7,85
CHUVA	4,83	1.5	0	175	5,26	5	17,9
COR	467	123	41	1317	0,50	3,41	3,57
COND	1529	377	37 9	5810	2,68	18	3,92
NAM	2,45	1,76	0	20	2,42	16,4	54,2
NN	1,44	0.88	0.03	7.38	2,42	11,0	80,5
PAP	1043	94,1	382,4	1304	-1,57	5,78	6,45
pН	7,45	1.21	0,85	12.53	1,79	4,17	3,71
SST	1 49	85,9	12	591	1,58	4,09	60,4
Т	45,4	3,07	28	50.5	-2,31	8,57	32,6
VAZ	67392	115 8 2	4474	47 8 50	-1,54	4,93	0

Tabela 2.1: Estatística básica dos parâmetros monitorados pela IPB.

* Relação entre o número de amostras que contém lacunas e a quantidade total de amostras disponibilizada pela empresa (1427 amostras).

Através do coeficiente de assimetria (*skewness*) e da forma como os dados se distribuem em torno da média (curtose) é possível analisar quão próxima da gaussiana apresenta a distribuição de cada parâmetro. Os valores de *skewness* estimados indicam que apenas a DQO, a DBO de entrada e saída e a COR apresentam distribuições simétricas. Os valores da curtose indicam que todos os parâmetros apresentam cauda mais pesada que a normal, com exceção da DBO de entrada e saída e da DQO, que apresentam distribuição praticamente normal. A DQO apresenta-se como o parâmetro mais correlacionado com a DBO de entrada e saída da lagoa.

Os dados originais disponibilizados cobrem um período de 1427 dias consecutivos, aproximadamente quatro anos amostrados. Entretanto, a extrema incidência de lacunas existente no banco de dados é um problema relevante, principalmente nos casos dos sólidos suspensos totais e nitrogênio amoniacal e nitrato, em que a proporção de lacunas é superior à quantidade de dados medidos no período em estudo. Esta situação dificulta ainda mais a modelagem da lagoa, fator este que deve ser levado em consideração na análise da performance dos modelos.

2.3.2 Pré-processamento de dados

De forma a tentar-se minimizar a perda de informação com a exclusão das amostras que contêm lacunas, dois conjuntos de dados foram construídos para predição da DBO de entrada e saída. O conjunto Dados 1 contem apenas as variáveis medidas com mais freqüência, ou seja, VAZ, DQO, COND, COR, pH, T, PAP e CEL, enquanto o conjunto Dados 2 contém dados de todas as doze variáveis disponibilizadas pela IPB. A variável CHUVA não foi incluída no conjunto Dados 1 por apresentar um número considerável de valores zero. O banco de dados inicial contendo 1427 amostras foi então reduzido aos conjuntos Dados 1 e 2 contendo 782 e 79 amostras, respectivamente.

Para construção dos modelos dinâmicos foram consideradas medidas das variáveis preditoras realizadas no dia anterior, ou seja, no tempo (t-1), onde t representa o dia em que a amostra foi coletada. As lacunas existentes nas amostras medidas no tempo (t-1) foram preenchidas através de interpolação linear.

Construídos os dois conjuntos de dados, o software Neurosolutions foi utilizado para escalonamento dos dados de 0,2 a 0,8 e aleatorização dos mesmos. Aproximadamente 80% dos dados de cada conjunto foram destinados à construção dos modelos e os 20% restantes à validação dos mesmos (veja Seção 4.2.3).

De forma a melhor avaliar a performance dos modelos de predição, e tendo em vista a necessidade de construção de no mínimo um conjunto de teste para os modelos perceptron de

múltiplas camadas - MLP, foi verificada a necessidade de formação de pelo menos um conjunto de dados para teste.

Dois novos conjuntos contendo 156 amostras foram então formados para testar a performance dos modelos construídos com o conjunto Dados 1, fazendo-se uso de interpolação linear para preencher as lacunas das amostras inicialmente excluídas, que até então não tinham sido utilizadas. Para o conjunto Dados 2 não foi possível a formação de novos conjuntos tendo em vista a existência de lacunas consecutivas de até sete dias para algumas variáveis.

2.4 Considerações finais sobre o capítulo

O objetivo principal deste capítulo foi apresentar a problemática que envolve a necessidade de monitorização e controle do sistema de tratamento de efluentes, bem como a necessidade de uso de modelos de predição para a estimativa da DBO, sendo este um dos principais parâmetros indicativos de qualidade de resíduos líquidos e recursos naturais. A dificuldade na obtenção de um banco de dados sem lacunas é evidenciada quando realizada a análise dos dados disponibilizados pela empresa. Vale ressaltar que as variáveis de entrada e saída da lagoa são relacionadas considerando o tempo de residência do efluente na mesma.

Dois conjuntos de dados foram formados para a construção dos modelos de predição. Para formação do primeiro, Dados 1, tentou-se maximizar o número de amostras utilizadas para a sua formação, havendo portanto a exclusão das variáveis raramente medidas e com uma grande quantidade de zeros. O segundo conjunto de dados foi formado de maneira a conter informações das doze variáveis inicialmente disponibilizadas pela empresa, sacrificando-se portanto a quantidade de amostras utilizadas. Três conjuntos foram formados para auxiliar a análise da performance dos modelos, obtidos à partir do conjunto Dados 1, já para o conjunto Dados 2 apenas um conjunto de validação (ou teste) foi formado.

CAPÍTULO 3 TÉCNICAS ESTATÍSTICAS MULTIVARIADAS

3.1 Regressão linear múltipla (MLR)

3.1.1 Introdução

O modelo de regressão que envolve mais de um regressor (variável independente) é chamado um modelo de regressão linear múltipla (ou MLR do inglês *Multiple Linear Regression*). Através deste modelo é possível observar os efeitos das variáveis regressoras sobre as respostas.

Infelizmente, o nome regressão, selecionado à partir do título do primeiro artigo F. Galton, de forma alguma reflete a importância ou profundidade de aplicação desta metodologia. O objetivo da MLR é encontrar relações entre variáveis dependentes e combinações lineares de variáveis independentes que minimizem o erro na estimativa dessas propriedades de interesse. Os coeficientes do modelo MLR são calculados utilizando o critério dos mínimos quadrados.

As limitações da regressão linear múltipla concentram-se no número de amostras que não pode ser menor do que o número de variáveis, e em realizar alguns cálculos matriciais quando as variáveis independentes possuírem uma elevada correlação entre si.

3.1.2 Um breve histórico

A técnica de regressão linear múltipla é bastante antiga. Suas aplicações são numerosas e ocorrem em quase todos os campos, incluindo, engenharia, física, economia, gerenciamento, etc.

O método de mínimos quadrados, também conhecido por análise de regressão, foi usado pela primeira vez por Sir Francis Galton, um dos pioneiros da Estatística, em um trabalho de 1885 intitulado *Regression toward mediocrity of hereditary stature*. Este método pode ser considerado a técnica estatística mais utilizada até os dias de hoje (MONTGOMERY e PECK, 1992).

15
3.1.3 Desenvolvimento matemático

Em geral, a variável dependente, ou resposta, y está relacionada com k variáveis independentes ou regressoras. O modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \in$$
(3.1)

é chamado de modelo linear de regressão múltipla com k regressores. Os β_j , j = 0, 1, ..., k, são chamados de coeficientes de regressão. Este modelo descreve um hiperplano no espaço de k dimensão das variáveis regressoras (x_j) e da variável dependente (y). O parâmetro β_j representa a mudança esperada na resposta y por unidade de mudança em x_j quando todos os demais regressores x_i $(i \neq j)$ são constantes. Se o domínio dos dados inclui $x_1 = x_2 = ... = x_k = 0$, então β_0 é a média de y quando $x_1 = x_2 = ... = x_k = 0$. Caso contrário, β_0 não tem interpretação física.

Modelos de regressão linear múltipla são geralmente usados como modelos empíricos. Ou seja, o modelo fenomenológico que relaciona $y e x_1, x_2, \ldots, x_k$ é desconhecido, mas sobre certos domínios das variáveis independentes o modelo de regressão linear é uma aproximação adequada.

Modelos que são mais complexos em estrutura que a Equação (3.1) podem geralmente ser analisados por técnicas de regressão linear múltipla. Neste caso termos quadráticos e cúbicos das variáveis dependentes podem ser adicionados.

Modelos que incluem efeitos de interação também podem ser analisados por métodos de regressão linear múltipla. Uma interação entre duas variáveis pode ser representada por um termo de produto cruzado no modelo, por exemplo, x_1x_2 .

Note que, apesar dos modelos de regressão linear poderem apresentar formas de superfícies não lineares, eles são lineares em seus parâmetros (os β 's), indiferentemente da forma da superfície que ele gera.

Na maioria dos problemas reais, os valores dos parâmetros (os coeficientes de regressão β_j) e as variâncias σ^2 não são conhecidos, e eles devem ser estimados à partir de uma amostra de dados. A análise de regressão consiste em uma coleção de ferramentas estatísticas para determinar estimativas dos parâmetros no modelo de regressão. Então, a equação de regressão, ou

modelo, é tipicamente usada na predição de observações futuras de y, ou para estimar a resposta média em um nível particular de x.

3.1.4 Estimativa dos parâmetros do modelo

Estimativa dos coeficientes de regressão por mínimos quadrados

O método de mínimos quadrados é geralmente utilizado para estimar os coeficientes de regressão no modelo de regressão simples assim como no de regressão múltipla. Supondo que n > k observações estão disponíveis, sendo que x_{ij} denota a *i*-ésima observação ou nível da variável x_j . As observações são,

$$(x_{i_1}, x_{i_2}, \dots, y_i), i = 1, 2, \dots, n \in n > k$$
 (3.2)

Cada observação $(x_{i1}, x_{i2}, ..., y_i)$ satisfaz o modelo da Equação (3.1), ou

$$y = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \epsilon_i, \quad i = 1, 2, \ldots, n$$
(3.3)

No ajuste do modelo de regressão, é conveniente expressar as operações matemáticas usando notação de matriz, o que facilita o entendimento, logo, o modelo da Equação (3.3) é um sistema de *n* equações que podem ser expressas em notação de matriz como,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{3.4}$$

onde,

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_{i} \\ \mathbf{y}_{2} \\ \vdots \\ \mathbf{y}_{n} \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & \mathbf{x}_{11} & \mathbf{x}_{12} & \dots & \mathbf{x}_{1k} \\ 1 & \mathbf{x}_{21} & \mathbf{x}_{22} & \dots & \mathbf{x}_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \mathbf{x}_{n1} & \mathbf{x}_{n2} & \dots & \mathbf{x}_{nk} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_{0} \\ \boldsymbol{\beta}_{1} \\ \vdots \\ \boldsymbol{\beta}_{k} \end{bmatrix} \quad \boldsymbol{\epsilon} = \begin{bmatrix} \boldsymbol{\epsilon}_{i} \\ \boldsymbol{\epsilon}_{2} \\ \vdots \\ \boldsymbol{\epsilon}_{n} \end{bmatrix}$$
(3.5)

Em geral, y é um vetor $(n\times 1)$ das observações, X é uma matriz $(n\times p)$ dos níveis das variáveis independentes, β é um vetor $(p\times 1)$ dos coeficientes de regressão, e \in é um vetor $(n\times 1)$ de erros aleatórios.

Deseja-se então encontrar o vetor de estimadores de mínimos quadrados, $\hat{\beta}$, que minimiza

$$\mathbf{L} = \sum_{i=1}^{n} \boldsymbol{\epsilon}_{i}^{2} = \boldsymbol{\epsilon}^{T} \boldsymbol{\epsilon} = \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)^{T} \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right), i = 1, 2, ..., n$$
(3.6)

onde $(^{T})$ representa a transposta. O estimador dos mínimos quadrados $\hat{\beta}$ é a solução para β na Equação (3.7) que segue,

$$\frac{\partial \mathbf{L}}{\partial \hat{\mathbf{\beta}}} = 0 \tag{3.7}$$

Logo, as equações resultantes que devem ser resolvidas são representadas pela Equação (3.8),

$$\mathbf{X}^T \, \mathbf{X} \, \hat{\boldsymbol{\beta}} = \mathbf{X}^T \, \mathbf{y} \tag{3.8}$$

A Equação (3.8) representa as equações normais de mínimos quadrados em forma de matriz. Para resolver as equações normais, multiplica-se ambos os lados da Equação (3.8) pelo inverso de $X^T X$. Portanto, a estimativa dos mínimos quadrados de β é

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{y}$$
(3.9)

Nota-se então que existem p = k + 1 equações normais em p = k + 1 parâmetros desconhecidos (os valores de $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k$). Escrevendo a Equação (3.8) em detalhes, tem-se

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i2} & \cdots & \sum_{i=1}^{n} x_{ik} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}^{2} & \sum_{i=1}^{n} x_{i1} x_{i2} & \cdots & \sum_{i=1}^{n} x_{i1} x_{ik} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{n} x_{ik} & \sum_{i=1}^{n} x_{ik} x_{i1} & \sum_{i=1}^{n} x_{ik} x_{i2} & \cdots & \sum_{i=1}^{n} x_{ik}^{2} \end{bmatrix} \begin{bmatrix} \hat{\beta}_{0} \\ \hat{\beta}_{1} \\ \vdots \\ \hat{\beta}_{k} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_{i} \\ \sum_{i=1}^{n} x_{i1} y_{i} \\ \vdots \\ \hat{\beta}_{k} \end{bmatrix}$$
(3.10)

É fácil verificar que $\mathbf{X}^T \mathbf{X}$ é uma matriz simétrica $(p \times p)$ e $\mathbf{X}^T \mathbf{y}$ é um vetor $(p \times 1)$. Nota-se a estrutura especial da matriz $\mathbf{X}^T \mathbf{X}$, onde os seus elementos diagonais são as somas dos quadrados dos elementos na coluna de \mathbf{X} e os elementos de $\mathbf{X}^T \mathbf{y}$ são as somas dos produtos cruzados de \mathbf{X} e as observações y_i . Às vezes, a matriz $\mathbf{X}^T \mathbf{X}$ é singular.

O modelo de regressão ajustado é então dado pela Equação (3.11), que segue

$$\hat{y} = \hat{\beta}_0 + \sum_{j=1}^k \hat{\beta}_j x_{ij}, \quad i = 1, 2, ..., n \quad \text{ou} \quad \hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$$
 (3.11)

Para medida da precisão do modelo de regressão é levada em consideração a linha ajustada, ou seja, a seguinte identidade

$$y_i - \hat{y}_i = (y_i - \overline{y}) - (\hat{y}_i - \overline{y})$$
(3.12)

O resíduo $e_i = y_i - \hat{y}_i$ é a diferença entre duas quantidades: (i) o desvio dos valores observados y_i a partir da média total \overline{y} e (ii) o desvio dos valores ajustados \hat{y}_i a partir da média total \overline{y} . O vetor ($n \times 1$) de resíduos é denotado por

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} \tag{3.13}$$

Estimativa de σ^2

A estimativa de σ^2 é obtida a partir dos resíduos. A soma quadrática dos resíduos é dada por

$$SQ_E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

= $\sum_{i=1}^{n} e_i^2$ (3.14)

A soma quadrática residual tem n - p graus de liberdade associada com ela, desde que p parâmetros sejam estimados no modelo de regressão. Logo, a média quadrática do resíduo (ou erro quadrático médio) que representa uma estimativa de σ^2 é dada por,

$$s^2 = MQ_E = \frac{SQ_E}{n - p} \tag{3.15}$$

3.1.5 Teste de hipóteses em regressão linear

Os testes de hipóteses apresentados nesta sessão são úteis na medição da adequação dos modelos. Estes testes requerem que os termos de erro \in_i no modelo de regressão sejam distribuídos normalmente e independentemente com média zero e variância σ^2 .

Antes de discutir-se sobre os testes de hipóteses, será apresentado um parâmetro de probabilidade, denominado p-valor, que auxilia na decisão de aceitação ou rejeição da hipótese nula, H₀.

P-valor no teste de hipóteses

É comum considerar o teste estatístico (e os dados) significantes quando a hipótese H₀ é rejeitada, então tem-se o p-valor como o menor nível de α para os dados serem significantes.

Para o teste de distribuição normal, tem-se que se z_0 é o valor estimado para o teste estatístico então, o p-valor é

$$P = 2 \left[1 - \Phi(z_0) \right] \text{ para testes de duas caudas (Fig. 3.1a)}$$

$$P = 1 - \Phi(z_0) \text{ para testes com cauda superior (Fig. 3.1b)}$$

$$P = \Phi(z_0) \text{ para testes com caudas inferiores (Fig. 3.1c)}$$
(3.16)

onde $\Phi(z_0)$ é a função de distribuição cumulativa de probabilidade.



Figura 3.1: Área da função de probabilidade acumulada.

Em geral utiliza-se p-valor maior que 0,05, correspondente a um intervalo de confiança de 95%, para indicar a aceitação da hipótese nula, H_0 .

O p-valor tem sido bastante utilizado na prática (MONTGOMERY et al., 2002). Como pode ser observado, este teste tem a vantagem de ser de fácil interpretação sendo seu valor estimado e fornecido por programas estatísticos, como o Minitab e o Statistica.

Teste para Significância de Regressão

O teste para significância de regressão é utilizado para determinar se existe uma relação linear entre a variável resposta e um conjunto de variáveis regressoras x_1, x_2, \ldots, x_k . As hipóteses apropriadas são

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1: \beta_j \neq 0 \quad \text{para pelo menos um } j$$
(3.17)

A rejeição de $H_0: \beta_1 = \beta_2 = ... = \beta_k = 0$ implica que pelo menos uma das variáveis regressoras $x_1, x_2, ..., x_k$ contribui significativamente para o modelo. Entretanto, a aceitação da hipótese nula pode indicar tanto uma contribuição não significativa das variáveis regressoras para o modelo assim como a presença de relações não-lineares entre as variáveis regressoras e a resposta.

Um método chamado análise de variância (ANOVA) pode ser usado para testar a significância da regressão. Este procedimento divide a variabilidade total nas variáveis respostas em componentes significantes como base para o teste. A identidade de análise de variância é,

$$\sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2 + \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$
(3.18)

Os dois componentes do lado direito da Equação 3.18 medem, respectivamente, a quantidade de variabilidade em y_i considerado pela linha de regressão e a variação residual não explicada pela linha de regressão.

A Equação (3.18) pode ser escrita como,

$$SQ_T = SQ_R + SQ_E \tag{3.19}$$

onde SQ_T é soma quadrática total de y dada por $SQ_T = \sum_{i=1}^n (y_i - \overline{y})^2$ ou $SQ_T = \sum_{i=1}^n y_i^2 - \frac{(\sum y_i)^2}{n}$; SQ_R é a soma quadrática da regressão dada por $SQ_R = \sum_{i=1}^n (\hat{y}_i - \overline{y})^2$ ou $SQ_R = \hat{\beta} \mathbf{X} \mathbf{y} - \frac{(\sum y_i)^2}{n}$ e SQ_E é o erro ou a soma quadrática dos resíduos dado por $SQ_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2$.

A soma quadrática total tem n - 1 graus de liberdade. Se a hipótese nula $H_0:\beta_1 = \beta_2 = \ldots = \beta_k = 0$ é verdadeira, então SQ_R / σ^2 é uma variável aleatória qui-quadrada com k graus de liberdade. Observa-se que o número de graus de liberdade para esta variável aleatória qui-quadrada é igual ao número de variáveis regressoras no modelo. SQ_E / σ^2 é uma

variável aleatória qui-quadrada com n - p graus de liberdade, e que SQ_E e SQ_R são independentes. O teste estatístico geralmente utilizado para $H_0: \beta_1 = \beta_2 = ... = \beta_k = 0$ é

$$F_0 = \frac{SQ_R/k}{SQ_E/(n-p)} = \frac{MQ_R}{MQ_E}$$
(3.20)

A quantidade de MQ_R é chamada de média quadrática da regressão (ou do modelo). Podese rejeitar H_0 se o valor calculado do teste estatístico, F_0 , for maior que F tabelado ($f_{\alpha, k, n-p}$). Desta forma tem-se que se a média quadrática residual for significativamente grande diz-se que existe falta de ajuste no modelo e o modelo não é adequado. O procedimento é usualmente resumido em uma tabela de análise de variância, como a Tabela 3.1.

Tabela 3.1 Análise de Variância para testar a significância de uma regressão.

Fonte de variação	Soma quadrática	Graus de liberdade	Média quadrática	F_0
Regressão	SQ_R	k	MQ_R	MQ_R / MQ_E
Erro ou resíduo	SQ_E	n-p	MQ_E	
Total	SQ_T	<i>n</i> - 1		

P-valor maior que 0,05 (intervalo de confiança de 95%) indica a aceitação da hipótese nula, ou seja, que a regressão não é significativa.

Testes sobre os coeficientes individuais de regressão

Geralmente, o conjunto das variáveis potencialmente importantes para o modelo não é conhecido, desta forma, dentre o universo de variáveis disponíveis, o subconjunto que efetivamente deve fazer parte do modelo deve ser determinado.

A aplicação de um teste de hipóteses sobre os coeficientes de regressão individuais auxilia na determinação do valor potencial de cada variável regressora no modelo de regressão. Por exemplo, o modelo pode ser mais eficiente com a inclusão de variáveis adicionais, ou talvez com a exclusão de um ou mais regressores presentes no modelo.

As hipóteses para testar a significância de qualquer coeficiente de regressão, chamado de β_{j} , são

$$H_0: \beta_j = 0$$

$$H_0: \beta_j \neq 0$$
(3.21)

Se H_0 : $\beta_j = 0$ não é rejeitado, então este indica que o regressor x_j pode ser excluído do modelo. O teste estatístico para esta hipótese é

$$T_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \tag{3.22}$$

onde C_{jj} é o elemento diagonal de $(\mathbf{X}^T \mathbf{X})^{-1}$ correspondente a β_j . A hipótese nula $H_0: \beta_j = 0$ é rejeitada se $|t_0| > t_{\alpha'_2, n-p}$. Observa-se que este é um teste parcial ou marginal, pois o coeficiente de regressão $\hat{\beta}_j$ depende de todas as variáveis regressoras x_i $(i \neq j)$ que estão no modelo. Então, este é um teste da contribuição de x_j dado os outros regressores no modelo.

Assim como no teste de significância da regressão, a análise de significância dos coeficientes de regressão pode ser realizada utilizando o p-valor, ao invés do teste T_0 . Sendo que p-valor menor 0,05 indica que o coeficiente de regressão $\hat{\beta}_j$ é significativo para o modelo. O uso da estatística p-valor oferece a vantagem de não requerer um valor tabelado para realizar o teste de hipóteses, como ocorre com a estatística T_0 .

Um processo de seleção de variáveis pode ser usado para encontrar-se o subconjunto de variáveis que efetivamente deve fazer parte do modelo. Neste processo é realizado o procedimento de inclusão ou exclusão de variáveis uma a uma até que, por algum critério estabelecido, o processo termine.

Existem vários algoritmos na seleção de variáveis, e eles não conduzem necessariamente à mesma solução. Os principais são: a seleção *stepwise*, a seleção passo-a-frente (*forward*) e a eliminação passo-a-trás (*backward*).

A seleção *stepwise*, proposto por Efroymson em 1960, é o método mais popular. Ele começa com a escolha da variável independente que melhor explica a variável dependente, ou seja, a que apresenta maior coeficiente de correlação ou de determinação múltipla, R^2 – Equação (3.32). Um teste é feito para verificar se a variável escolhida é significante. Caso seja, ela fica no modelo, e o próximo passo é escolher uma segunda variável que produza o maior aumento em R^2 quando adicionada ao modelo. Um teste é feito para verificar se a contribuição dessa nova variável é significante. Depois que a segunda variável entra no modelo, um teste é realizado para verificar se a primeira variável permanece no modelo. Caso permaneça, uma terceira variável é selecionada. Se uma terceira variável entra no modelo, testa-se para verificar se as duas primeiras continuam no modelo. Pode acontecer que uma delas, ou as duas sejam eliminadas. Tenta-se a inclusão de uma nova variável. Caso entre, tenta-se a eliminação das que já estão no modelo. O procedimento acaba quando não se consegue adicionar nem eliminar variáveis do modelo.

A seleção passo-a-frente começa da mesma forma que a *stepwise* entretanto, uma vez que uma variável foi selecionada e admitida como significante, ela fica no modelo. Desta forma, o processo continua até que o teste rejeita como significante a variável escolhida pelo critério de maior aumento de R^2 , ou seja, quando nenhuma das variáveis candidatas tem um p-valor menor que o *alfa* de inclusão.

A eliminação passo-a-trás começa pelo ajuste de todas as variáveis independentes candidatas a ficar no modelo, as variáveis são então eliminadas uma de cada vez, da mesma forma que a seleção *stepwise*. Com o processo passo-a-trás, uma vez que a variável sai do modelo, ela não entra mais. Neste caso, o procedimento termina quando nenhuma variável incluída no modelo tem um p-valor maior que o *alfa* de exclusão.

Para a seleção *stepwise*, DRAPER e SMITH (1981) sugerem o uso de $\alpha = 0,05$ (nível de 95% de confiança) e $\alpha = 0,10$ (nível de 90% de confiança) para os testes de inclusão e exclusão de variáveis, respectivamente. Esta diferença nos valores dos *alfas* tem o objetivo de dificultar a exclusão de variáveis, garantindo que pelo menos algumas variáveis preditoras sejam utilizadas na modelagem. Entretanto, DRAPER e SMITH (1981) e MONTGOMERY e PECK (1992) concordam que a escolha dos níveis de confiança deve ser de responsabilidade do analista.

Evidencia-se portanto a importância da experiência do analista pois, construir um modelo de regressão que inclua um subconjunto das variáveis potencialmente importantes envolve o seguinte paradigma: o modelo deve incluir tantos regressores possíveis de tal forma que a

informação contida nessas variáveis possa ser um bom preditor da variável dependente, ao passo que, o modelo deve incluir o menor número possível de regressores de forma a facilitar a sua interpretação e ainda minimizar o custo de coletar a informação e de manter o modelo.

3.1.6 Intervalo de confiança sobre a resposta média

Para estimar a resposta média em um ponto particular, $x_{01}, x_{02}, \ldots, x_{0k}$, define-se o vetor

$$x_{0} = \begin{bmatrix} 1 \\ x_{01} \\ x_{02} \\ \vdots \\ x_{0k} \end{bmatrix}$$
(3.23)

A resposta média neste ponto é $E(\mathbf{Y}|\mathbf{x}_0) = \boldsymbol{\mu}_{\mathbf{Y}|_{\mathbf{x}_0}} = \mathbf{x}_0^T \boldsymbol{\beta}$, que é estimado por

$$\hat{\boldsymbol{\mu}}_{\mathrm{Y}|_{\mathbf{x}_{0}}} = \mathbf{x}_{0}^{T} \hat{\boldsymbol{\beta}}, \qquad (3.24)$$

Pode-se demonstrar que a variância de $\hat{\mu}_{Y|_{x_0}}$ é (MONTIGOMERY, 1998),

$$\mathbf{V}(\hat{\boldsymbol{\mu}}_{\mathbf{Y}|_{\mathbf{x}_{0}}}) = \sigma^{2} \mathbf{x}_{0}^{T} \left(\mathbf{X}^{T} \mathbf{X} \right)^{-1} \mathbf{x}_{0}.$$
(3.25)

Um intervalo de confiança percentual de 100(1 - α) sobre $\mu_{Y|_{x_0}}$ pode ser construído a partir da estatística

$$\frac{\hat{\mu}_{\mathbf{Y}|\mathbf{x}_{0}} - \mu_{\mathbf{Y}|\mathbf{x}_{0}}}{\sqrt{\sigma^{2} \mathbf{x}_{0}^{T} (\mathbf{X}^{T} \mathbf{X})^{-1} \mathbf{x}_{0}}},$$
(3.26)

ou seja,

$$\hat{\mu}_{Y|x_{0}} - \mathbf{t}_{\alpha_{2}, n=p} \sqrt{\sigma^{2} \mathbf{x}_{0}^{T} \left(\mathbf{X}^{T} \mathbf{X} \right)^{-1} \mathbf{x}_{0}} \leq \mu_{Y|x_{0}} \leq \hat{\mu}_{Y|x_{0}} + \mathbf{t}_{\alpha_{2}, n=p} \sqrt{\sigma^{2} \mathbf{x}_{0}^{T} \left(\mathbf{X}^{T} \mathbf{X} \right)^{-1} \mathbf{x}_{0}}$$
(3.27)

Os modelos obtidos neste trabalho têm por objetivo a predição de valores individuais e não médios, logo, encontra-se maior utilidade na determinação do intervalo de confiança de predição. A definição destes intervalos é apresentada na Seção 3.1.7, que segue.

3.1.7 Predição de novas observações

Um modelo de regressão pode ser usado para predizer futuras observações da variável resposta correspondente a valores particulares das variáveis independentes, $x_{01}, x_{02}, \ldots, x_{0k}$. Se $\mathbf{x}_0^T = [x_{01}, x_{02}, \ldots, x_{0k}]$, então uma estimativa pontual da observação futura Y_0 no ponto x_{01}, x_{02}, \ldots , x_{0k} é

$$\hat{\mathbf{y}}_0 = \mathbf{x}_0^T \hat{\boldsymbol{\beta}} \,. \tag{3.28}$$

Portanto, em regressão linear múltipla, um intervalo de predição percentual de 100. $(1 - \alpha)$ para a observação futura de Y no ponto $x = x_0$ é

$$\hat{\mathbf{y}}_{0} - \mathbf{t}_{\alpha_{2}', n = p} \sqrt{\sigma^{2} \left(1 + \mathbf{x}_{0}^{T} \left(\mathbf{X}^{T} \mathbf{X} \right)^{-1} \mathbf{x}_{0} \right)} \leq \mathbf{y}_{0} \leq \hat{\mathbf{y}}_{0} + \mathbf{t}_{\alpha_{2}', n = p} \sqrt{\sigma^{2} \left(1 + \mathbf{x}_{0}^{T} \left(\mathbf{X}^{T} \mathbf{X} \right)^{-1} \mathbf{x}_{0} \right)}.$$
 (3.29)

Comparando-se o intervalo de predição, Equação (3.29), com a expressão para o intervalo de confiança, Equação (3.27), observa-se que o intervalo de predição é sempre maior que o intervalo de confiança. O intervalo de confiança expressa o erro em estimar-se a média de uma distribuição, enquanto o intervalo de predição expressa o erro em predizer-se uma observação futura a partir da distribuição do ponto x_0 . Este último deve incluir o erro em estimar-se a média neste ponto, bem como, a variabilidade inerente na variável aleatória y no mesmo valor $x = x_0$.

Na predição de novas observações e na estimativa da resposta média, dado um ponto x_{01} , x_{02}, \ldots, x_{0k} , deve-se ter o cuidado em não extrapolar além da região contendo as observações

originais. É possível que um modelo que ajusta bem uma região dos dados originais não ajuste bem fora desta região.

A Figura 3.2 ilustra a região contendo as observações para um modelo de regressão de duas variáveis. Nota-se que o ponto (x_{01}, x_{02}) está dentro dos intervalos de ambas variáveis regressoras $x_1 e x_2$, mas está fora da região que é realmente medida pelas observações originais. Logo, embora este ponto y_0 possa ser encontrado, deve-se ter em mente que esta predição é extremamente perigosa, principalmente neste caso em que o ponto (x_{01}, x_{02}) está dentro dos intervalos de ambas variáveis regressoras $x_1 e x_2$. Assim, o valor predito de uma nova observação ou a estimativa da resposta média neste ponto é uma extrapolação do modelo de regressão original.



Figura 3.2: Região contendo as observações para um modelo de regressão de duas variáveis.

Identificação de outliers através de gráficos de controle multivariados

Um problema padrão em análises multivariadas é testar a igualdade de um vetor da média desconhecida de uma população μ e um vetor da média específico μ_0 , neste caso tem-se o teste de um conjunto de validação desconhecido e o conjunto usado para construir o modelo. Este teste pode ser conduzido através de gráficos de controle multivariados, tais como, gráficos de controle dos *scores* dos componentes principais e da estatística T² de Hotelling.

Os gráficos dos *scores* das componentes principais, obtidos pelas técnicas de análise de componentes principais (PCA), têm sido utilizados para identificar *outliers* no conjunto de validação, e ainda amostras que, apesar de pertencerem ao conjunto de dados utilizados para

construir os modelos, apresentam comportamento diferente das demais, sendo identificadas como possíveis *outliers*. Entretanto, não existe um consenso quanto ao uso das primeiras componentes ou últimas componentes para a identificação destas amostras, dificultando assim as análises para sistemas de dimensões elevadas. O mesmo não acontece com os gráficos da estatística T^2 de Hotelling, onde apenas um único gráfico é obtido através das variáveis originais ou das componentes.

O desenvolvimento de um procedimento para a estimativa da estatística T^2 de Hotelling requer que as variáveis regressoras sejam normalmente distribuídas e variância normal e independentemente distribuída com média zero e variância σ^2 . Tem-se portanto

$$\mathbf{T}^{2} = n \left(\overline{\mathbf{X}} - \boldsymbol{\mu}_{0} \right)^{T} \mathbf{S}^{-1} \left(\overline{\mathbf{X}} - \boldsymbol{\mu}_{0} \right), \qquad (3.30)$$

onde $\bar{\mathbf{X}} \in \mathbf{S}^{-1}$ representam a média amostral e a matriz de covariância, respectivamente, sendo,

$$\overline{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_{i} \qquad \mathbf{e} \qquad \mathbf{S} = \frac{1}{n-1} \sum_{i=1}^{n} \left(\mathbf{X}_{i} - \overline{\mathbf{X}} \right) \left(\mathbf{X}_{i} - \overline{\mathbf{X}} \right)^{T}.$$
(3.31)

A estatística $(n-p)T^2/p(n-1)$ tem uma distribuição F com $p \in n-p$ graus de liberdade, sobre a hipótese nula $\mu = \mu_0$. O gráfico de T² apresenta apenas limite superior de controle que é dado por $F_{\alpha, p, n-p}$.

Dentre as técnicas mais importantes de decomposição da estatística T^2 está a análise de componentes principais, pois, a estimativa de T^2 através das componentes principais é menos suscetível a influências de observações aberrantes provenientes de variáveis individuais (COLACIOPPO, 2001). Assim, mais uma vez, os *scores* das componentes principais podem ser utilizados ao invés das observações originais. Na Seção 3.2, é apresentada uma descrição detalhada sobre a obtenção das componentes principais.

A maioria dos autores, inclusive JACKSON (1979; 1980; 1985), KOURTI e MACGREGOR (1996) e JOHNSON e WICHERN (1998), enfatizam aspectos de redução de dimensão e interpretação do corpo de dados colocando grande atenção às primeiras componentes principais, que retém a maioria da variabilidade. Em ORTIZ-ESTARELLES et al. (2001) e MORALES et al. (1999), pode-se encontrar aplicações desta metodologia.

Apesar de saber-se que casos de extrema não normalidade podem influenciar os resultados de T², não se conhece até que ponto a normalidade dos dados deve ser buscada. Os estudos de Monte Carlo indicam que, apesar de ser exigida a normalidade em diversas análises estatísticas, as consequências em violar-se a suposição de distribuição normal são menos severas do que se acredita (JOHNSON, 1987).

Os gráficos de T^2 podem auxiliar a detecção de um evento não usual dentro do conjunto de dados, porém não fornecem informações das razões dele ocorrer (COLACIOPPO, 2001). Desta forma, observações suspeitas (pontos influentes e possíveis *outliers*) devem ser investigadas antes de serem excluídas do modelo.

3.1.8 Avaliação da eficiência do modelo de regressão

Análises devem ser conduzidas para examinar a eficiência e adequação do modelo construído, tendo em vista as diversas suposições feitas para a sua construção, tais como, a estimativa dos parâmetros requer que a suposição de que os erros são variáveis aleatórias não correlacionadas com média zero e variância constante; testes de hipóteses e estimativa dos intervalos requerem que os erros sejam normalmente distribuídos e, além disso, assumiu-se que a ordem do modelo está correta.

Análise dos resíduos

A análise dos resíduos é frequentemente útil para checar a suposição de que os erros são distribuídos normalmente com variância constante, bem como, para determinar se a adição de um termo adicional ao modelo é útil.

Os resíduos de um modelo de regressão são $e_i = y_i - \hat{y}_i$, i = 1, 2, ..., n, onde y_i é uma observação real e \hat{y}_i é o correspondente valor ajustado a partir do modelo de regressão.

Para verificar a normalidade, pode-se construir um histograma de frequência dos resíduos ou um gráfico de probabilidade normal dos resíduos. Uma resposta numérica pode ser obtida utilizando o teste de Anderson-Darling, onde, para p-valor maior que 0,05 o resíduo é considerado normal.

Feqüentemente, é útil fazer-se o gráfico dos resíduos em sequência temporal (se conhecida); dos resíduos versus a resposta predita, e dos resíduos versus as variáveis dependentes. Estes gráficos geralmente apresentam alguns padrões que podem sugerir a presença de anormalidades no modelo. Como exemplo, tem-se que a presença de curvaturas no gráfico dos resíduos versus as variáveis dependentes podem indicar que o modelo é inadequado, ou seja, termos de ordem maior devem ser adicionados ao modelo, uma transformação sobre as variáveis x ou y (ou ambas) deve ser considerada, ou outros regressores devem ser considerados. Uma descrição detalhada sobre como analisar estes padrões podem ser encontrada em MONTGOMERY e PECK (1992).

Coeficiente de Determinação Múltipla

O coeficiente de determinação múltipla R^2 é definido como

$$R^2 = \frac{SQ_R}{SQ_T} = 1 - \frac{SQ_E}{SQ_T}$$
(3.32)

A raiz quadrada positiva de R^2 é chamada de coeficiente de correlação múltipla entre y e o conjunto de variáveis regressoras x_1, x_2, \ldots, x_k . Ou seja, R é uma medida da associação linear entre y e x_1, x_2, \ldots, x_k .

Na regressão linear simples tem-se $0 \le R^2 \le 1$. Entretanto, um valor alto de R^2 não implica necessariamente que o modelo de regressão é bom. Adicionando uma variável ao modelo o valor de R^2 sempre aumentará, indiferentemente se a variável adicionada é significante ou não (MONTGOMERY e PECKY, 1992). Então, modelos que têm altos valores de R^2 podem produzir pobres predições de novas observações ou estimativas da resposta média.

Para tentar contornar este problema, geralmente utiliza-se o coeficiente de determinação múltipla ajustado, dado por

$$R^{2} a justado = = 1 - \frac{SQ_{E}/(n-p)}{SQ_{T}/(n-1)}$$
(3.33)

O R^2 ajustado (ou R^2 adj.) reflete melhor a proporção de variabilidade explicada pelo modelo de regressão pois ele leva em conta o número de variáveis regressoras. R^2 ajustado aumentará apenas se a adição de uma variável regressora produzir uma alta redução na soma quadrada dos resíduos para compensar a perda de um grau de liberdade residual. O R^2 ajustado é útil na comparação de modelos de regressão pois o modelo com menor média quadrada residual explica mais a variabilidade em y.

Partindo-se de um modelo composto por todas as variáveis, a exclusão de uma variável pode ser feita com base no seu coeficiente de regressão. Daí, verifica-se se houve melhora ou não do R^2 ajustado. Se todos os coeficientes da regressão são estimados independentemente, a exclusão pode ocasionar pouco dano ao modelo. Entretanto, se as variáveis regressoras são altamente correlacionadas a exclusão de uma variável pode ser perigosa.

Uma diferença significativa entre R^2 e R^2 ajustado indica que o modelo foi sobre parametizado, ou seja, termos que não contribuem significativamente ao ajuste foram incluídos.

Segundo BARROS *et al.* (1996), a decisão de que modelo é mais adequado pode também ser baseada num teste *F*, comparando-se a redução na soma quadrática residual por unidade de parâmetros adicionados com a própria média quadrática do modelo contendo o maior número de parâmetros. Tem-se então a seguinte relação de interesse

$$F = \left(\frac{SQ_{E1} - SQ_{E2}}{d}\right) / MQ_{E2} , \qquad (3.34)$$

onde, os sub-índices 1 e 2 representam o menor e maior modelo, respectivamente, e *d* a diferença do número de parâmetros dos dois modelos.

Se o valor de F não for significativamente maior que $F_{k_1-k_2,(n-p)_2}$, não há ganhos em se aumentar o número de parâmetros. Vale ressaltar que a adição de parâmetros pode ser devido à inclusão de termos quadráticos, cúbicos, ou a inclusão de uma nova variável.

3.2 Análise dos componentes principais (PCA)

3.2.1 Introdução

A análise de componentes principais PCA (do inglês *Principal Component Analysis*), também chamada de transformação canônica (DUTTA *et al.*, 1998), apresenta como idéia básica encontrar combinações lineares das variáveis de um sistema, de modo a reduzir sua dimensão a uma coleção muito menor de variáveis transformadas, que ainda descreva suas características.

3.2.2 Um breve histórico

Os procedimentos gerais da técnica de componentes principais, na forma que é conhecida hoje, foram desenvolvidos por Harold Hotelling e publicados no seu trabalho pioneiro em 1933, embora esta tenha sido originada em 1901 por Karl Pearson (MARDIA *et al.*, 1979; JACKSON, 1991).

O desenvolvimento desta técnica foi um tanto fragmentada nos anos seguintes, alternando-se entre períodos de estagnação e de grande atividade. Com o advento computacional, surgiu um grande número de estudos voltados para aplicação da técnica de PCA nas mais diversas áreas por apresentarem a necessidade de compreensão e interpretação de dados de seus sistemas multivariados.

Para um melhor entendimento da técnica de PCA, na Seção 3.2.3 a seguir apresenta-se uma descrição clara do seu desenvolvimento matemático.

3.2.3 Desenvolvimento Matemático

PCA é uma técnica estatística para redução da dimensionalidade de um conjunto de dados multivariados.

Considerando-se uma matriz X, $n \times p$, onde n é o número de amostras e p o número de variáveis independentes, as componentes principais são obtidas diagonalizando a matriz de covariância X^tX, onde X^t é a transposta de X.

Os autovetores correspondentes da matriz de covariância formam os eixos das coordenadas do sistema do espaço transformado e as variâncias nos dados ao longo destes eixos são os autovalores, λ_i . ADAMS (1995) e JACKSON (1991) apresentam formas simples para cálculo dos autovetores e autovalores.

Uma transformação linear simples relaciona os valores das componentes principais T, chamadas scores, aos valores dos dados originais, X pela Equação (3.40),

$$\mathbf{X} = \mathbf{T} \mathbf{W}^T + \mathbf{E} \,. \tag{3.40}$$

Geometricamente as componentes principais (eixos do espaço transformado) são relacionadas com os eixos do espaço original pelas rotações dadas pelos *loadings* ou pesos, W. Os *loadings* e os *scores* podem ser normalizados de modo que a soma quadrática de cada um seja igual a seus apropriados autovalores e a unidade, respectivamente, ou vice e versa. E representa o erro associado ao modelo.

A transformação do conjunto de dados pelas componentes principais apresenta duas vantagens principais para o tratamento de dados, a redução da dimensionalidade e a ortogonalidade das variáveis (MCLAIN *et al.*, 2000; OLIVEIRA, 2000).

A primeira componente principal (CP) é orientada na direção que contém a maior variância dos dados. A segunda CP é ortogonal à primeira e é orientada na direção que contém a máxima quantidade de variância residual dos dados, isto é, a variância não explicada pela primeira CP. Sucessivamente, as demais componentes explicam a quantidade máxima de variância residual.

Quanto maior a porção de variância nos dados puder ser explicada pelas primeiras componentes principais, uma redução de dimensionalidade mais eficiente pode ser obtida. Esta porção da variância é representada pela razão da soma dos autovalores das componentes no espaço reduzido, (i=1, 2, ..., k) e da soma total de autovalores (j=1, 2, ..., p), ou seja,

$$V_{k} = \left(\frac{\lambda_{1} + \ldots + \lambda_{k}}{\lambda_{1} + \ldots + \lambda_{p}}\right), \tag{3.41}$$

onde k representa as primeiras k componentes principais e p é a dimensionalidade original.

Um aspecto crítico do PCA é seleção do número de componentes principais (k). Infelizmente não existe um critério universal para esta seleção (MORALES *et al.*, 1999; TODESCHINE, 1997), e existem discrepâncias entre os diferentes métodos (ORTIZ-ESTERELLES, 2001). Assim os autores têm sugerido diversas formas para seleção das componentes principais, dentre elas tem-se:

- (a) incluir apenas CPs suficientes para explicar 90% do total da variância. E se a matriz de correlação foi utilizada, excluir aquelas CPs cujos autovalores são menores que um (MALKAVAARA et al., 2000; MARDIA et al., 1979);
- (b) utilizar critérios estatísticos (GONZÁLEZ e GONZÁLEZ-ARJONA, 1995; JACKSON, 1991), e
- (c) analisar gráficos dos autovalores, SCREE plot, proposto por Cattell (1966) (JACKSON, 1991; PIELOU, 1984 e MARDIA et al., 1979).

TODESCHINE (1997) apresenta uma comparação detalhada entre métodos que estabelecem a correlação dos dados e o número de componentes principais a serem incluídas no modelo PCA. O autor propõe um novo índice de correlação (K) como uma medida da quantidade de correlação no conjunto de dados, Equações (3.42a) e (3.42b), assim como as funções linear (KL) e não-linear (KP) para estimar o número máximo e mínimo de CPs significantes, respectivamente, Equações (3.43) e (3.44).

$$K = \frac{\sum_{k=1}^{p} |VE_k - 1/p|}{2 (p - 1)/p}, \quad 0 \le K \le 1,$$
(3.42a)

onde,

$$VE_k = \frac{\lambda_k}{\sum_{k=1}^p \lambda_k}.$$
(3.42b)

$$KL = int[l + (p-1)(l - K)], \qquad (3.44)$$

$$KP = int \left[p^{(1-K)} \right]. \tag{3.45}$$

Nas Equações (3.42a) e (3.42b), VE_k é a variância explicada a partir da k-ésima componente principal; o denominador corresponde ao máximo valor alcançado a partir do numerador e é usado para escalonar os valores de 0 a 1. Pode-se observar que todos os autovalores zero são considerados nesta equação, cada um fornecendo uma contribuição de 1/p. Nas equações 3.44 e 3.45, *int* significa o mais próximo e maior número inteiro.

O autor conclui que KL e KP podem ser usados para estimar o número de CPs que fornecem informações úteis, e que são mais simples que a maioria dos critérios que vêm sendo utilizados. Neste trabalho, utiliza-se o critério proposto por TODESCHINE (1997) para estimar as componentes principais significantes para o modelo PCA.

Vale ressaltar que, nem sempre a escolha das primeiras componentes principais levam a melhor representação de um sistema, haja vista que as últimos podem conter informações importantes de algumas variáveis originais (DESPAGNE e MASSART, 1998; MARDIA *et al.*, 1979; OLIVEIRA-ESQUERRE *et al.*, 2003a).

Como citado anteriormente, além da redução da dimensionalidade, a técnica de PCA apresenta a vantagem de fornecer um novo conjunto de parâmetros não correlacionados. A propriedade de ortogonalidade das CPs é desejável de maneira a reduzir o erro da modelagem em tratamentos estatísticos, especialmente quando as CPs apresentam-se física, química ou biologicamente facilmente interpretáveis.

3.2.4 Regressão de componentes principais (PCR)

Regressão dos componentes principais (PCR do inglês *principal components regression*) consiste num caso de regressão múltipla de Y sobre as componentes principais de X. De uma forma geral, PCR é utilizada quando um grande número de variáveis independentes que apresentam extensiva colinearidade ou correlação entre elas está disponível para a formação de um modelo de predição.

Colinearidade adiciona redundância ao modelo causando instabilidade numérica na estimativa dos coeficientes de regressão. Desta forma, através do uso dos componentes principais, ou de um subconjunto formado pelas primeiras componentes, o modelo de regressão é reduzido tendo ainda informações relevantes dos dados originais.

A regressão das componentes principais é obtida pela regressão de Y sobre os *scores* de X. Os coeficientes de regressão β para cada y podem ser escritos de acordo com a Equação (3.45),

$$\hat{\boldsymbol{\beta}} = (\mathbf{T}' \ \mathbf{T})^{-1} \ \mathbf{T}^T \mathbf{Y} , \qquad (3.45)$$

onde T é a matriz dos *scores* de X. Uma descrição mais detalhada de PCR pode ser encontrada em MARDIA (1979) e MARTENS e NAES (1989).

Apesar do PCR resolver o problema de colinearidade e possibilitar a formação de um modelo de menor dimensão (permitindo redução de ruído); este apresenta o risco de informações preditivas úteis estarem sendo descartadas com as componentes principais e algum ruído permanecer nas componentes usadas na regressão (GELADI e KOWALSKI, 1986). Isto ocorre pelo fato de que para a formação das componentes não se utilizam informações sobre a relação entre **X** e a variável a ser predita, mesmo havendo uma tendência geral das componentes com maiores variâncias explicarem melhor as variáveis dependentes (MARDIA, 1979).

Modelos contendo termos quadráticos das componentes principais também podem ser formados, entretanto, quanto maior a ordem do modelo mais difíceis sua análise e interpretação se tornam.

A metodologia de análise da eficiência dos modelos PCR é a mesma da utilizada para os modelos MLR.

3.3 Mínimos quadrados parciais (PLS)

3.3.1 Introdução

O método de mínimos quadrados parciais, PLS do inglês partial least squares, é uma técnica estatística de análise e regressão multivariada geralmente utilizada para a formação de modelos de menor dimensão. Pesquisadores que trabalham com análises químicas, particularmente na área de calibração multivariada, têm utilizado intensivamente esta técnica. BEEBE *et al.* (1998), GELADI e KOWALSKI (1986) e MARTENS e NAES (1989) apresentam uma descrição detalhada desta técnica e ainda aplicações à calibração multivariada. Nos últimos anos,

PLS tem ganho popularidade como uma técnica que possibilita a redução de dimensionalidade concomitantemente à modelagem de processos industriais (BLANK e BROUN, 1993a e 1993b; CHEN et al., 1998; HOLCOMB e MORARI, 1992; KANO et al., 2000; SIMOGLOU et al., 2000; QIN e MCAVOY, 1992).

3.3.2 Um breve histórico

O método PLS, inicalmente apresentado por Herman Wold, é designado para maximizar a predição assim como o ajuste, ou seja, o método é otimizado de forma a maximizar a proporção de variância da variável dependente que é explicada por seus preditores (ou variáveis independentes). Este permite a redução de dimensionalidade do sistema em estudo semelhantemente a PCA, entretanto, a formação dos componentes, agora denomidadas de variáveis latentes – VLs, sofre influência da relação entre as variáveis de entrada e a predita.

Enquanto muitos pesquisadores estão devotos à aplicação prática do método, outros encontram censuras ao algoritmo proposto por Wold. O algoritmo proposto por Hoskuldsson (1988) e Frank (1987) apresenta-se levemente diferente e mais compacto que o proposto por Wold, sendo o seu entendimento considerado mais simples (MARTENS e NAES, 1989).

Apesar de algumas críticas, o sucesso de sua aplicação tem sido apresentado nos inúmeros artigos publicados desde a sua apresentação em 1960. S. Wold e H. Martens (1970) foram os pioneiros a utilizarem esta técnica em aplicações químicas. Em adição à calibração espectrométrica, a técnica de PLS tem sido aplicada para monitorar e controlar processos industriais, pois estes podem conter centenas de variáveis controláveis e dezenas de monitoradas.

3.3.3 Desenvolvimento Matemático

Como em qualquer outra técnica de regressão, o que se deseja com o método PLS é formar um modelo de predição das variáveis Y através das variáveis X, minimizando o erro desta estimativa. Considerando a possibilidade de redução de dimensionalidade proposta pela técnica PCA, tem-se que o espaço Y poderia ser explicado pelas primeiras componentes principais onde as últimas significariam ruídos para o modelo, entretanto não é o que acontece. Apesar das primeiras componentes principais estarem orientadas na direção de maior variância dos dados de

X, estas podem não explicar muito bem o espaço Y, tendo em vista que nenhuma informação de Y foi considerada para estimá-las.

De forma a minimizar este problema, pelo método PLS tem-se a utilização de informações de Y para a formação das componentes de X. Logo, este método apresenta vantagens sobre a técnica de PCA na redução de dimensionalidade de um sistema.

O processo de desenvolvimento do modelo PLS pode ser ilustrado pela Figura 3.3.



Figura 3.3: Representação da relação entre os scores da matriz X e Y.

O primeiro passo na modelagem via PLS consiste na utilização da técnica PCA para obtenção das componentes principais t e p de X e Y, respectivamente - Figuras 3.3 (a) e (c). Estima-se então o vetor **b** que melhor representa a relação linear entre t e p - Figura 3.3 (b). Em seguida, com base no valor estimado de **b**, realizam-se leves rotações de t e p - representadas pelas linhas tracejadas na Figura 3.3 (a) e (c). Obtém-se por fim um modelo com maior precisão para a predição de Y utilizando as componentes t e p rotacionadas.

O método de mínimos quadrados parciais, representado pela Figura 3.3, foi proposto por H. Wold e baseia-se nas propriedades do algoritmo NIPALS (*Nonlinear iterative partial least squares*). Como mencionado, neste tem-se que tanto a matriz das variáveis independentes, X como a das variáveis dependentes Y são representadas pela análise de componentes principais, ou seja:

$$\mathbf{X} = \mathbf{T} \mathbf{W}^T + \mathbf{E} \tag{3.46a}$$

$$\mathbf{Y} = \mathbf{P} \mathbf{Q}^T + \mathbf{F} \tag{3.46b}$$

Uma relação entre os dois blocos é então realizada correlacionando-se os *scores* para cada componente de cada vez, utilizando-se um modelo linear:

$$p_k = b_k t_k, \qquad (3.47a)$$

onde :

$$b_k = \frac{p_k t_k}{t_k^T t_k},\tag{3.47b}$$

para cada k = 1, 2, ..., p, componente principal.

Como mencionado anteriormente, esse modelo, entretanto, não é o melhor possível. Isto porque a análise de componentes principais é realizada em cada matriz separadamente, podendo resultar numa relação não muito satisfatória (não linear) entre os *scores* dos dois blocos (POPPI, 1993). Deve-se buscar um modelo onde as matrizes dos resíduos **E**, e principalmente **F**, sejam as menores possíveis e, ao mesmo tempo, conseguir uma relação linear entre t e p. O algoritmo de obtenção de tal modelo foi proposto por Hoskuldsson e Frank (MARDIA, 1979), considerando apenas uma variável dependente y. Em MARTENS e NAES (1989) pode-se encontrar a extensão deste algoritmo para mais de uma variável y.

Tem-se, portanto, que o espaço de variáveis independentes X pode ser transformado em um novo espaço, deixando muito do seu ruído, que afeta a estimativa do espaço Y, como resíduos. Este novo espaço, que pode apresentar menor dimensão e é representado por variáveis ortogonais T, pode então ser escrito pela Equação (3.48), que segue

$$\mathbf{T} = \mathbf{X} \mathbf{W} (\mathbf{P}^T \mathbf{W})^{-1} .$$
 (3.48)

3.3.4 Regressão por mínimos quadrados parciais (PLSR)

A regressão por mínimos quadrados parciais (PLSR) é uma boa alternativa aos clássicos métodos de regressão linear múltipla e regressão de componentes principais por ser mais robusto (GELADI e KOWALSKI, 1986).

PLSR consiste basicamente na regressão dos *scores* de X, obtidos pela Equação (3.48), sobre os scores de Y, onde os coeficientes de regressão *b* são estimados pela Equação (3.47b). Para um espaço Y unidimensional a regressão é realizada sobre o próprio Y.

Infelizmente, assim como para o caso das componentes principais, não existe um critério universal para definição do número de VLs que devem ser utilizados para a construção dos modelos, e discrepâncias entre os diferentes critérios são bastante comuns. O número ótimo de VLs é aqui identificado considerando o aumento na variância acumulada predita, que deve ser maior que 2% com a adição de uma nova VL, segundo instruções do MATLAB PLS_Tollbox 2.0. A metodologia de análise da eficiência dos modelos PLSR é a mesma da utilizada para os modelos MLR.

Embora sejam métodos lineares, PCR e PLSR podem ser usados para a modelagem de alguns tipos específicos de dados não lineares. Se a forma de relação não linear entre as variáveis de entrada e aquela a ser predita é conhecida, um modelo pode ser linearizado pela transformação apropriada das variáveis originais, ou adicionando uma ordem maior e termos cruzados para a equação de regressão (DESPAGNE e MASSART, 1998). Entretanto, existe um risco de introduzir uma quantidade significante de informação irrelevante no modelo.

Variações não lineares de PCR e PLSR também existem. Sua limitação principal é que eles são baseados na suposição que uma relação simples (por exemplo, quadrática) existe entre a resposta do modelo e os componentes, entretanto, podem apresentar resultados satisfatórios em modelagem não linear (WALCZAC e MASSART, 1996).

Para PLSR, relações não lineares entre os *scores* de X e Y devem ser encontradas. Alguns trabalhos propõem a incorporação da técnica de redes neurais em modelos PLS para auxiliar a busca destas relações (BAFFI *et al.*, 1999; QIN e MCAVOY, 1992; HOLCOMB e MORARI, 1992).

3.4 Considerações finais sobre o capítulo

O objetivo principal deste capítulo foi apresentar uma base teórica necessária para o entendimento das técnicas MLR, PCR e PLSR utilizadas neste trabalho para a construção de modelos de predição da DBO.

A técnica MLR apresenta a vantagem de utilizar variáveis originais, sendo os modelos obtidos fáceis de serem interpretados. Entretanto, quando existe colinearidade entre suas variáveis preditoras, e/ou quando um número reduzido de amostras e um número elevado de variáveis são disponibilizadas para a construção do modelo, dificuldades numéricas na estimativa nos parâmetros de modelagem são verificadas. A utilização de um método de seleção de variáveis pode minimizar estes problemas (Figura 3.4a).



Figura 3.4: Desenho esquemático das técnicas MLR, PCR e PLSR, sendo consideradas oito variáveis independentes, uma variável dependente e uma redução de dimensionalidade de oito para três regressores.

As técnicas PCR e PLSR surgem então como alternativas à MLR na construção de modelos lineares de predição, Figuras 3.4 (b) e (c), respectivamente. Apesar de serem capazes de reduzir a dimensionalidade do sistema, e eliminar o problema de colinearidade existente entre as variáveis preditoras, os modelos obtidos podem ser difíceis de serem interpretados fisicamenteInformações úteis das variáveis preditoras podem ainda ser perdidas com a exclusão das últimas componentes principais.

Neste trabalho, os resultados de predição dos MLR e PLSR são apresentados e analisados de forma detalhada. Modelos PCR foram inicialmente testados, entretanto, apresentaram performance de predição inferior aos modelos MLR e PLSR, logo, apenas alguns resultados serão apresentados aqui.

O software Minitab foi utilizado para a construção dos modelos MLR e PCR, e o software Statistica foi utilizado para a construção dos modelos PLSR assim como para a construção dos gráficos da estatística T² de Hotelling.

CAPÍTULO 4

REDES NEURAIS ARTIFICIAIS

4.1 Conceitos gerais

4.1.1 Introdução

Redes neurais artificiais (RNAs) é o nome dado a um conjunto de métodos matemáticos e algoritmos computacionais especialmente projetados para simular o processamento de informações e aquisição de conhecimento do cérebro humano.

A partir dos anos 80, iniciou-se um crescimento acelerado nos níveis das atividades de pesquisa em relação às RNAs. Enquanto muitos pesquisadores têm se esforçado de maneira a desenvolver princípios fundamentais e novos algoritmos, os estudos também têm se voltado para o campo de aplicações práticas. De fato, nos últimos anos tem ficado claro que as redes neurais oferecem um conjunto poderoso de ferramentas para a solução de problemas de reconhecimento de padrões, processamento de dados, modelagem e controle não linear, que pode ser considerado como complementar aos métodos mais convencionais.

A seguir será apresentado um breve histórico e uma descrição detalhada dos usos e aplicações das técnicas de RNAs e o desenvolvimento matemático que as envolvem.

4.1.2 Um breve histórico

O cérebro humano é a estrutura computacional mais complexa conhecida pelo homem. Sua capacidade de aprender através de estímulos externos tem inspirado muitos pesquisadores das mais diversas áreas na tentativa de criar um modelo computacional que copie as atividades do cérebro de diferentes maneiras.

As origens das RNAs datam dos anos 40 com a publicação do trabalho de McCulloch e Pitts, *A Logical Calculus of Ideas Immanent in Nervous Activity*, que descrevia um modelo de cálculo lógico de uma rede neural. Além de orientar todo o desenvolvimento posterior de redes neurais, este trabalho também ajudou os pesquisadores da Universidade da Pennsylvania a construírem o EDVAC (*Eletronic Discrete Variable Automatic Computer*) e o seu sucessor o ENIAC (*Eletronic Numerical Integrator And Computer*), entre 1943 e 1946. O trabalho de McCulloch e Pitts também inspirou outros pesquisadores como Marvin Minsky na área de inteligência artificial e Norbert Wiener na área de cibernética.

O segundo grande passo para o desenvolvimento das RNAs ocorreu em 1949 com a publicação do livro *The Organization of Behaviour* por Hebb, em que ele propôs um mecanismo específico para aprendizagem em redes neurais, baseado nas conexões sinápticas biológicas. Em 1958, ROSENBLATT publicou o trabalho: *Principles of Neurodynamics*, que apresentou a estrutura de um perceptron, uma unidade de reconhecimento de padrões que é a base do neurônio de uma rede neural.

Por volta de 1969, MINSKY e PAPERTS publicaram o livro *Perceptrons*, que mostrava as limitações de um perceptron isolado. Tal trabalho, adicionado ao alto custo computacional da época desestimulou pesquisadores e agências de financiamento a investirem em redes neurais, fazendo com que na década de 70 apenas aplicações de reconhecida eficácia, tais como o reconhecimento de imagens, fossem desenvolvidas.

Na década de 80, as pesquisas em RNAs ressurgiram devido principalmente à publicação de importantes trabalhos tais como o da energia de aproximação de Hopfield, *Neural Networks and Physical Systems with Emergent Collective Computacional Abilities*, e o algoritmo de aprendizagem de retro-propagação (*back-propagation*) para perceptrons de múltiplas camadas primeiramente proposto por Werbos, modificado várias vezes, e então popularizado por Rumelhart e McClelland após a publicação do livro de dois volumes *Parallel Distributed Processing: Explorations in the Microestrutures of Cognition*, em 1986.

Apesar do desenvolvimento de redes neurais ter sido retomado, na década de 80 os trabalhos de modelagem e simulação estavam voltados praticamente para a utilização da técnica de colocação ortogonal, devido à sua característica de necessitar um número reduzido de pontos de discretização e exigir menor carga computacional de resolução numérica dos modelos matemáticos.

A década de 90 foi caracterizada pela consolidação dos fundamentos teóricos das RNAs assim como pelo seu sucesso em aplicações práticas, ocasionado pelo aumento do poder de cálculo dos microcomputadores e das estações de trabalhos (*workstation*), e a redução do custo dos mesmos. Tentou-se compensar o período de estagnação nesta área de pesquisa.

O futuro da aplicação de redes neurais é promissor pelas suas características principais. Diz-se que elas são capazes de "aprender" relações do tipo entrada-saída de um sistema a partir de um processo de "treinamento" semelhante ao aprendizado de um cérebro humano, sem necessitar da idealização, elaboração e validação de modelos matemáticos complexos.

Uma outra característica é sua estrutura de distribuição paralela de informações. Concebida de forma semelhante ao cérebro, esse paralelismo vem atraindo a atenção de inúmeros pesquisadores de todos os setores do conhecimento, em particular os da área de informática. Já se sabe que os computadores só conseguirão aumentar sua capacidade e velocidade de cálculo adotando-se uma estrutura essencialmente paralela de fluxo de informações, como a que ocorre no tecido nervoso cerebral (GONTARSKI, 2000b).

4.1.3 Redes neurais biológicas

Como o próprio nome indica, as redes neurais artificiais são técnicas computacionais que apresentam um modelo matemático inspirado na estrutura neural de organismos inteligentes, numa tentativa de imitar o funcionamento do cérebro humano. São formadas por unidades básicas que buscam copiar a forma de ligação dos neurônios no cérebro humano (HONG-GUANG e JI-ZONG, 2000).

Para um melhor entendimento da base estrutural de uma RNA, faz-se necessário uma breve abordagem de citologia.

O sistema nervoso é formado por um conjunto extremamente complexo de células, os neurônios. O neurônio, unidade celular fundamental do cérebro, é composto por três regiões distintas, soma, dendritos e axônio, que são apresentadas na Figura 4.1a. A parte central da célula, ou soma, é a região nuclear onde está a maior parte do material intracelular e onde ocorrem os processos metabólicos. Do soma partem dois tipos de filamentos, os dendritos e o axônio. O neurônio recebe sinais (impulsos nervosos) de outros neurônios através dos dendritos (receptores) e transmite o sinal gerado através do seu corpo celular (transmissor). Vê-se então que os dendritos formam um conjunto de terminais de entrada e o axônio um longo terminal de saída.

A formação da estrutura nervosa ocorre através das conexões sinápticas, que conectam a árvore dendrital com os axônios de outras células (Figura 4.1b). Quando os impulsos nervosos chegam ao terminal sináptico, certas substâncias chamadas neurotransmissores alteram o potencial elétrico das membranas de modo a promover ou inibir a passagem do impulso nervoso. A efetividade de uma sinapse pode ser ajustada pela passagem de sinais através dela então, a sinapse pode aprender a partir das atividades em que ela participa.



Figura 4.1a: O neurônio biológico.

Figura 4.1b: Conexões sinápticas.

O cérebro humano possui cerca de 10^{11} neurônios e estima-se que o número de conexões sinápticas gira em torno de 10^{14} a 10^{15} , possibilitando a formação de uma rede muito complexa (BISHOP, 1994). Pode-se dizer que o cérebro representa um sistema dinâmico de retroalimentação, não linear, massivamente paralelo de proporções astronômicas e que, portanto, consegue realizar tarefas com uma velocidade surpreendentemente superior ao melhor computador.

4.1.4 Neurônio artificial e funções de ativação

A RNA é um instrumento computacional que tenta imitar o sistema de aprendizagem do cérebro humano. A propriedade chave dos sistemas neurais biológicos e artificiais está na sua habilidade de modificar respostas como um resultado de exposição a sinais externos. Isto é geralmente citado como aprendizagem e ocorre através da mudança da força das sinapses.

Um neurônio é uma unidade de processamento de informação que é fundamental para a operação de uma rede neural artificial. Este neurônio é representado pela Figura 4.2 como um diagrama em blocos.



Figura 4.2: Estrutura de um neurônio artificial.

A Figura 4.2 representa o modelo de McCulloc-Pitts em que o sinal x_i na entrada da sinapse *i* conectada ao neurônio *j* é multiplicado pelo peso sináptico w_{ji} (que é análogo à intensidade sináptica do neurônio biológico) e é então somado a todos os *inputs* pesados para dar o *input* total para a unidade na forma,

$$\upsilon_{j}(n) = \sum_{i=1}^{m} w_{ji}(n) x_{i}(n) + w_{j0}, \qquad (4.1)$$

onde *m* é o número total de entradas (excluindo o bias) aplicadas ao neurônio *j*; o parâmetro de deslocamento w_0 é chamado de *bias* (e corresponde à descarga limite em um neurônio biológico). Formalmente, o *bias* pode ser considerado como um caso especial de um peso de um *input* extra cujo valor x_0 é permanentemente mantido em +1. Assim, pode-se escrever a Equação (4.1) na forma,

$$\upsilon_{j}(n) = \sum_{i=0}^{m} w_{ji}(n) x_{i}(n), \qquad (4.2)$$

onde $y_0 = 1$. Como os pesos e os *bias* podem apresentar qualquer sinal, as sinápses podem ser excitadoras ou inibidoras. O *output* y(n) da unidade é então dado pela operação sobre uma função de ativação φ , assim

$$y_j(n) = \varphi_j(\upsilon_j(n)). \tag{4.3}$$

A função φ mais intensamente empregada é a função sigmoidal, definida como:

$$\varphi_j(\upsilon_j(n)) = (1 + e^{\upsilon_j(n)})^{-1}.$$
(4.4a)

Outras funções também empregadas são:

Linear
$$\varphi_j(\upsilon_j(n)) = A\upsilon_j(n) + B$$
 (4.4b)

Tangente hiperbólica
$$\varphi_{j}(\upsilon_{j}(n)) = \left(e^{\upsilon_{j}(n)} - e^{-\upsilon_{j}(n)} / e^{\upsilon_{j}(n)} + e^{-\upsilon_{j}(n)}\right)$$
(4.4c)

Limite ("threshold")
$$\varphi_j(\upsilon_j(n)) = +1$$
 Se $\upsilon_j(n) > 0$ (4.4d)
Senão, $\varphi_j(\upsilon_j(n)) = 0$

Nota-se então que este modelo simples de neurônio forma o elemento matemático básico dos modelos de redes neurais artificiais. Assim, através da conexão de vários elementos processadores é possível construir uma classe muito geral de mapeamento não linear, que pode ser aplicada aos mais variados problemas práticos. A adaptação dos valores dos pesos, de acordo com um apropriado algoritmo de treinamento, pode permitir a aprendizagem em resposta aos dados externos.

4.1.5 Arquiteturas das redes neurais artificiais

A maneira pela qual os neurônios de uma rede neural estão estruturados está intimamente ligada com o algoritmo de aprendizagem usado para treinar a rede. Nas Figuras 4.3 (a) e (b) são apresentadas as estruturas de uma rede de camada única e de uma rede de múltiplas camadas, respectivamente. Note que para tal designação a camada de entrada não é contabilizada.



Figura 4.3: Estrutura de uma rede neural, (a) com uma única camada de neurônios; (b) com múltiplas camadas de neurônios.

Duas características básicas das redes neurais também são levadas em consideração como uma tentativa de estabelecer uma classificação dos tipos existentes de redes neurais:

- a) A forma como a informação flui pela rede. Em algumas ela flui sempre em um sentido único, da entrada para a saída. Em outras a informação flui independente de sentido ou direção.
- b) A forma como os resultados são interpretados, quando a rede neural compara as suas respostas com uma desejada durante o seu treinamento, ou quando a rede neural por si própria estabelece a melhor forma de encontrar uma resposta.

As redes onde as informações avançam em apenas uma direção são denominadas de diretas (*feedforward*). São bastante comuns por sua relativa simplicidade e estabilidade. As redes de retro-propagação (*backpropagation*) são exemplos deste tipo.

Aquelas redes onde as conexões permitem que as informações fluam em qualquer sentido são chamadas de recorrentes (*feedback*). Em geral, as redes deste tipo são capazes de representar sistemas mais complexos, mas isto pode ser às custas de complicações indesejáveis para o treinamento da rede.

A forma como as redes são treinadas também é uma característica usada em sua classificação. O método mais comum para treinar uma rede é alimentá-la com dados de entrada e com os dados de saída desejados. Os primeiros resultados são produzidos pela rede e comparados com os resultados desejados gerando os erros. Estes são usados para ajustar os pesos na rede de forma que a próxima vez que o mesmo dado for usado, a rede se aproxime da resposta desejada. Esta forma de treinamento é chamada de supervisionada, e é o caso das redes de retro-propagação.

As redes neurais de aprendizagem supervisionada possuem uma grande capacidade para inferir sobre as relações entre as variáveis de entrada e saída, mesmo quando o número de neurônios é pequeno. Isto facilita para que os cálculos possam ser efetuados a baixo custo até em microcomputadores de pequeno porte. Mesmo sendo simples, a rede *feedforward* mostra ser uma ferramenta eficaz para a modelagem de processos estacionários e dinâmicos não-lineares.

Em algumas redes não supervisionadas, não se fornece o resultado desejado, mas a ela é permitido organizar os dados numa forma que ela entenda estarem ajustados. Redes deste tipo são também chamadas de auto-organizadas. Algumas redes têm uma fase inicial auto-organizada seguida de uma fase supervisionada. A discussão e o equacionamento de diversos tipos de estruturas para redes neurais podem ser encontrados em HAYKIN (1994), BAUGHMAN e LIU (1995) e KOVÁCS (1996).

4.2.6 Função erro

O sinal de erro na saída do neurônio j, na iteração n (i.e., a apresentação do n-ésimo exemplo de treinamento), é definido por:

$$e_{j}(n) = d_{j}(n) - y_{j}(n),$$
 (4.5)

onde o neurônio j é um nó de saída.
Por definição, o valor instantâneo da energia do erro para o neurônio $j \in \frac{1}{2}e_j^2(n)$. Correspondentemente, o valor instantâneo de E(n) da energia do erro é obtido somando-se os termos $\frac{1}{2}e_j^2(n)$ de todos os neurônios da camada de saída, logo,

$$E(n) = \sum_{j \in C} \frac{1}{2} e_j^2(n), \qquad (4.6)$$

onde o conjunto C corresponde todos os neurônios da camada de saída da rede. Considere então que N represente o número total de padrões (exemplos) contidos no conjunto de treinamento. A energia média do erro quadrado é obtida somando-se os E(n) para todos os n e então normalizando em relação ao tamanho do conjunto N, ou seja,

$$E_{med}(n) = \frac{1}{N} \sum_{n=1}^{N} E(n).$$
(4.7)

O objetivo do processo de aprendizagem é ajustar os parâmetros livres (pesos) da rede para minimizar a energia do erro quadrado (E_{med}) . Esta variável será denominada, no decorrer deste trabalho, como erro quadrático médio, *MSE*.

A função erro pode ser considerada geometricamente como uma superfície ajustando-se sobre o espaço dos pesos, como mostrado na Figura 4.4, e o problema de treinamento da rede corresponde à pesquisa pelo mínimo global w_4 da superfície do erro.



Figura 4.4: Desenho esquemático da superfície da função erro E(w).

Para uma rede de múltiplas camadas a função erro é altamente não linear em função dos pesos, e a pesquisa pelo mínimo geralmente procede de modo interativo, começando por alguns pontos encontrados no espaço dos pesos. Alguns algoritmos encontram os mínimos locais mais próximos, enquanto outros são capazes de escapar dos mínimos locais e oferecer a possibilidade de encontrar o mínimo global. Em geral, a superfície de erro é extremamente complexa e para muitas aplicações práticas um bom mínimo local pode ser suficiente para encontrar resultados satisfatórios.

Muitos algoritmos executam a minimização da função erro através das derivadas desta com relação aos pesos da rede. Estas derivadas formam as componentes do vetor gradiente $\nabla E(w)$ da função erro, que, em um dado ponto no espaço dos pesos, dá o gradiente da superfície do erro, como indicado na Figura 4.4.

Um dos mais importantes fatores da classe das funções de mapeamento não lineares dado pela MLP é que existe um procedimento computacional eficiente para avaliar as derivadas da função erro, baseadas na técnica de retro-propagação do erro, ou seja, no algoritmo de *backpropagation*.

4.1.7 Algoritmo backpropagation

Durante o treinamento com o algoritmo *backpropagation*, a rede opera em uma sequência de dois passos. Primeiro, um conjunto de variáveis é apresentado à camada de *input* da rede. A atividade flui através da rede, camada por camada, até que a resposta seja produzida pela camada de *output*. No segundo passo, o *output* obtido é comparado ao *output* desejado para esse conjunto particular. Se esta não estiver correta, o erro é calculado. O erro é então propagado a partir da camada de *output* até a camada de *input*, e os pesos das conexões das unidades das camadas internas vão sendo modificados conforme o erro é retro-propagado.

Assim, vê-se que a formulação matemática do algoritmo de *backpropagation* consiste num problema de atualização dos pesos em função do erro.

Para a formulação matemática deste problema é considerada uma rede contendo apenas uma camada intermediária.

Na Seção 4.2.1 introduziu-se o conceito de uma única unidade de processamento descrito pelas Equações (4.2) e (4.3). Considerando-se então um conjunto de *m outputs*, todos contendo

inputs comuns, então é obtida uma rede tendo uma camada de parâmetros adaptativos (pesos) como ilustrado na Figura 4.3 (a). As variáveis de output, denotadas por y_j , são dadas pela Equação (4.3).

O ajuste dos pesos pode ser realizado pelo algoritmo de retro-propagação. Inicialmente aplica-se uma correção $\Delta w_{ji}(n)$ ao peso sináptico $w_{ji}(n)$, que é proporcional à derivada $\partial E(n)/\partial w_{ji}(n)$. De acordo com a regra da cadeia do cálculo, tem-se,

$$\frac{\partial E(n)}{\partial w_{ji}(n)} = \frac{\partial E(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)},$$
(4.8)

A derivada $\partial E(n)/\partial w_{ji}(n)$ representa um fator de sensibilidade, determinando a direção de busca no espaço de pesos, para o peso sináptico w_{ji} .

Diferenciando ambos os lados da Equação (4.6) em relação a $e_j(n)$, da Equação (4.5) em relação a $y_j(n)$, da Equação (4.3) em relação a $v_j(n)$ e da Equação (4.2) em relação a $w_{ji}(n)$, obtém-se,

$$\frac{\partial E(n)}{\partial e_j(n)} = e_j(n), \qquad (4.9)$$

$$\frac{\partial e_j(n)}{\partial y_j(n)} = -1, \qquad (4.10)$$

$$\frac{\partial y_j(n)}{\partial v_j(n)} = \varphi_j'(v_j(n)), \qquad (4.11)$$

$$\frac{\partial v_j(n)}{\partial w_{ji}(n)} = y_j(n).$$
(4.12)

O uso da apóstrofe na Equação (4.11) significa a diferenciação em relação ao argumento. Substituindo-se as Equações (4.9) a (4.12) na Equação (4.8), obtém-se,

$$\frac{\partial E(n)}{\partial w_{ji}(n)} = -e_j(n)\varphi_j(v_j(n))y_j(n), \qquad (4.13)$$

A correção $\Delta w_{ji}(n)$ aplicada a $w_{ji}(n)$ é então definida por,

$$\Delta w_{ji} = -\eta \frac{\partial E(n)}{\partial w_{ji}(n)}, \qquad (4.14)$$

onde η é o parâmetro da taxa de aprendizagem do algoritmo de retropropagação. O uso do sinal negativo na Equação (4.14) indica a busca de uma direção para a mudança dos pesos que reduza o valor de E(n). Correspondentemente, o uso da Equação (4.13) em (4.14) produz,

$$\Delta w_{ji}(n) = \eta \delta_j(n) y_j(n), \qquad (4.15)$$

onde o gradiente local $\delta_j(n)$ é definido por,

$$\delta_{j}(n) = -\frac{\partial E(n)}{\partial \upsilon_{j}(n)} = -\frac{\partial E(n)}{\partial e_{j}(n)} \frac{\partial e_{j}(n)}{\partial y_{j}(n)} \frac{\partial y_{j}(n)}{\partial \upsilon_{j}(n)} = e_{j}(n)\varphi_{j}(\upsilon_{j}(n)).$$

$$(4.16)$$

O gradiente aponta para as modificações necessárias nos pesos sinápticos. De acordo com a Equação (4.16), o gradiente local $\delta_j(n)$ para o neurônio de saída *j* é igual ao produto do sinal do erro $e_j(n)$ corresponde para aquele neurônio pela derivada $\varphi'_j(\upsilon_j(n))$ da função de ativação associada.

4.2 Redes neurais perceptron de múltiplas camadas (MLP)

4.2.1 Introdução

Quando RNAs de uma só camada são utilizadas, o mapeamento é realizado diretamente entre as variáveis de *input* e as de *output*, ou seja, não é possível a formação de uma representação interna. Tal restrição implica que variáveis de *input* resultem em variáveis de *output* similares, o que leva o sistema à incapacidade de aprender importantes mapeamentos.

O desenvolvimento do algoritmo de treinamento *backpropagation* mostrou ser possível treinar eficientemente redes com camadas intermediárias, resultando assim no modelo de RNAs mais utilizado atualmente, as redes perceptron de múltiplas camadas (MLP) (BISHOP, 1994).

Nessas redes, cada camada tem uma função específica. A camada de *output* recebe os estímulos da camada intermediária e constrói a variável que será a resposta. As camadas intermediárias funcionam como extratoras de características, seus pesos são uma codificação das características apresentadas pelas variáveis de *input* e permitem que a rede crie sua própria representação.

Embora existam inúmeras arquiteturas de redes neurais, a arquitetura multicamadas é, sem dúvida, a mais freqüentemente encontrada na literatura. Entre as razões para a sua popularidade podemos citar sua capacidade de *aproximação universal* e sua flexibilidade para formar soluções de qualidade para uma ampla classe de problemas, a partir de um mesmo algoritmo de aprendizado.

4.2.2 Algoritmo backpropagation padrão

A estrutura neural da rede perceptron de múltiplas camadas (Figura 4.3 (b)) pode ser representada por $(\mathbf{x}(n), \mathbf{d}(n))$, com o vetor de entrada $\mathbf{x}(n)$ aplicado à camada de entrada e o vetor resposta desejada $\mathbf{d}(n)$ apresentado à camada de saída. Desta forma, o campo induzido $v_i^l(n)$ para o neurônio *j* na camada *l* pode ser representado pela Equação (4.17),

$$\upsilon_{j}^{l}(n) = \sum_{i=0}^{m=0} w_{ji}^{l}(n) y_{i}^{(l-1)}(n), \qquad (4.17)$$

onde $y_i^{(l-1(n))}$ é o sinal (função) de saída do neurônio *i* na camada anterior *l*-1, na iteração *n*, e $w_{ji}^{(l)}(n)$ é o peso sináptico do neurônio *j* na camada *l*, que é alimentado pelo neurônio *i* da camada *l*-1. Para i=0, tem-se $y_0^{(l-1)}(n) = +1$ e $w_{j0}^{(l)}(n) = b_j^{(l)}(n)$ é o bias aplicado no neurônio *j* na camada *l*. Assumindo-se uma função sigmoidal, o sinal da saída do neurônio *j* na camada *l* é,

$$y_{i}^{(l)}(n) = \varphi_{j}(\upsilon_{j}(n)),$$
 (4.18)

Considerando-se j na primeira camada intermediária, tem-se,

$$y_i^{(0)}(n) = x_j(n),$$
 (4.19)

onde $x_j(n)$ é o *j-ésimo* elemento do vetor de entrada $\mathbf{x}(n)$. Se o neurônio *j* está na camada de saída (i.e., *l=L*, onde *L* é denominado de profundidade da rede), tem-se

$$y_i^{(L)}(n) = o_j(n),$$
 (4.20)

e o erro segue a mesma forma da Equação (4.5), ou seja,

$$e_{j}(n) = d_{j}(n) - o_{j}(n),$$
 (4.21)

onde $d_j(n)$ é o *j-ésimo* elemento do vetor resposta desejado $\mathbf{d}(n)$.

Para ajuste dos pesos tem-se os gradientes locais (δs) definidos por,

$$\delta_{j}^{(l)} = e_{j}^{(L)}(n)\varphi_{j}^{'}\left(\upsilon_{j}^{(L)}(n)\right) \qquad \text{para o neurônio } j \text{ na camada } L \qquad (4.22)$$

$$\delta_{j}^{(l)} = \varphi_{j}^{'}\left(\upsilon_{j}^{(l)}(n)\right)\sum_{k}\delta_{k}^{(l+1)}(n)w_{kj}^{(l+1)}(n) \qquad \text{para o neurônio } j \text{ na camada oculta } l. \qquad (4.23)$$

onde o apóstrofe em $\varphi_{j}(\cdot)$ representa a diferenciação em relação ao argumento. O ajuste dos pesos sinápticos da rede na camada l é então realizado na forma,

$$w_{ji}^{(l)}(n+1) = w_{ji}^{(l)}(n) + \eta \delta_j^{(l)}(n) y_i^{(l-1)}(n) .$$
(4.24)

onde η é o parâmetro da taxa de aprendizagem.

Um método simples de aumentar a taxa de aprendizagem, evitando no entanto o perigo de instabilidade, é modificar a Equação (4.24) incluindo um termo de momento, como mostrado por RUMELHART *et al.* (1986),

$$w_{ji}^{(l)}(n+1) = w_{ji}^{(l)}(n) + \alpha \left[w_{ji}^{(l)}(n-1) \right] + \eta \delta_j^{(l)}(n) y_i^{(l-1)}(n)$$
(4.25)

onde α é usualmente um número positivo chamado de constante de momento. A Equação (4.25) é chamada de regra delta generalizada.

Um importante fator para os cálculos das derivadas é sua eficiência computacional. Geralmente o número de pesos é muito maior que o número de unidades de processamento, e a contribuição dominante do erro propagado ou retropropagado é proveniente da avaliação da soma dos pesos (com a avaliação da função de ativação sendo negligenciada). Entretanto, a técnica de *backpropagation* permite que todas as derivadas sejam avaliadas usando uma única direção de propagação, ou seja, a retropropagação.

Em síntese, o processo de treinamento da rede se constitui basicamente em um problema de otimização e exige a apresentação da base de dados à rede por diversas vezes, até que o conjunto de pesos encontrados satisfaça a condição desejada para o erro apurado. Desta forma, as mesmas dificuldades associadas à otimização de processos também ocorrem na fase de

aprendizado da rede, tais como problemas de convergência, existência de mínimos locais e tempo de computação indeterminado.

4.2.3 Algoritmo de treinamento delta-bar-delta (DBD)

A regra de treinamento chamada de delta-bar-delta (DBD) é um aprimoramento do algoritmo *backpropagation* que foi sugerido inicialmente por JACOBS (1988). No algoritmo *backpropagation*, mais especificamente na regra delta generalizada, tem-se que o parâmetro da taxa de aprendizagem, η , e o passo, α , são constantes fixas. No caso da regra DBD, a taxa de treinamento e o passo são adaptados de acordo com os valores passados do erro em cada neurônio. Se as atualizações do peso atual e passada apresentam o mesmo sinal, a taxa de treinamento cresce linearmente. A razão para isto é que se a modificação do peso está se movendo na mesma direção do decrescimento do erro, o decaimento deste será mais rápido se o passo for aumentado. Se as atualizações tiverem sinais diferentes, existe uma indicação de que o peso tem se movido para muito longe. Quando isso acontece, a taxa de aprendizagem decresce geometricamente para evitar divergência.

Para implementar a heurística que deverá incrementar e decrementar as taxas de treinamento para cada conexão, Jacobs empregou uma média ponderada dos componentes do gradiente como sendo,

$$\overline{\delta}(n) = (1-\lambda)\delta(n) + \lambda\delta(n-1)$$
(4.26)

e ainda,

$$\Delta \eta(n) = \begin{cases} \kappa & \overline{\delta}(n-1)\delta(n) > 0\\ -\beta \eta(n) & \overline{\delta}(n-1)\delta(n) < 0\\ 0 & \text{caso contrario} \end{cases}$$
(4.27)

onde κ é uma constante aditiva, β é uma constante multiplicativa, λ é uma constante de amortecimento, e $\delta(n)$ é o gradiente do peso descrito anteriormente.

Como mencionado anteriormente, esta regra aumenta a taxa de treinamento de forma linear, mas o decréscimo é geométrico. Assim, segundo GONTARSKI (2000), quando deseja-se apressar a convergência deve-se modificar o $\eta(0)$, ou $\eta(bond)$ (que são os valores inicial e limite para $\eta(n)$, respectivamente) e o κ para valores mais elevados. Pode-se ainda experimentar baixar o valor de β , mas este parâmetro controla sensivelmente as oscilações no método. Portanto, para eliminar oscilações no erro deve-se reduzir os valores de $\eta(0)$, $\eta(bond)$, κ e de β .

4.2.4 Definição da topologia da rede

Os fatores que contribuem para a complexidade do processo de modelagem via redes MLP variam desde as etapas de coleta e pré-processamento de dados até a busca e definição da melhor arquitetura da rede neural.

Quanto à definição da topologia da rede, os fatores de complexidade podem ser enunciados da seguinte forma (MILLER et al., 1989):

- (a) o espaço de busca é infinito e o número de neurônios e conexões entre neurônios é ilimitado;
- (b) o espaço de busca não é diferenciável, pois as mudanças no número de neurônios e conexões são discretas, além de promoverem um efeito descontínuo no comportamento da rede neural;
- (c) redes neurais com estruturas similares podem apresentar comportamento muito diferentes, enquanto que comportamentos similares podem ser produzidos por redes neurais estruturalmente muito diferentes;
- (d) o resultado do processo de busca depende da condição inicial.

Diversas topologias da rede MLP têm sido propostas (FELDMAN e BALLARD, 1982; HOPFIELD, 1982; KOHONEN, 1984; RUMELHART e McCLELLAND, 1986). Cada uma se diferencia pelo número e características dos neurônios, conexões, procedimentos de treinamento, funções de transferência e se os valores de entrada e saída são discretos ou contínuos. A otimização da topologia de uma rede neural é provavelmente o passo mais tedioso do desenvolvimento do modelo visto que o mapeamento não linear na busca do erro global mínimo geralmente é demorado, sendo que em alguns casos apenas mínimos locais são encontrados.

Número de neurônios e camadas

Uma importante etapa para obter-se um resultado satisfatório quando se trabalha com redes MLP é a de definição do número de camadas e do número de neurônios em cada uma delas. A terminologia usada para descrever a topologia das MLPs pode variar de acordo com os autores, alguns deles consideram as camadas de *input* e *output* como simples "para-choques" (DESPAGNE e MASSART, 1998). Neste trabalho, consideram-se os *inputs* e *outputs* formadores de suas respectivas camadas.

Não existe ainda nenhuma regra que esclareça satisfatoriamente a forma correta de configurar o número de camadas e de neurônios, contudo, pode-se afirmar que quanto mais complexa a rede criada, maior pode ser a sua capacidade de ajustar-se aos dados, e maior também é o tempo para treiná-la. Mas um número elevado de neurônios, ou mesmo a adição de mais camadas podem não significar um melhor resultado. Isto porque, quando o número de pesos a serem ajustados é maior que o número de dados de *input* disponíveis, é possível que venha a ocorrer o chamado *overfitting* ou *overtraining*, onde a rede representa muito bem os dados usados no treinamento, mas perde a capacidade de generalizar para outras situações (NASCIMENTO *et al.*, 2000). Assim, aliado a outros cuidados, recomenda-se sempre a estrutura mais simples possível.

Na modelagem de processos químicos, o uso de uma única camada interna tem se mostrado suficiente, visto que quando há a necessidade de modelos mais complexos o ajuste do número de neurônios da camada intermediária geralmente é suficiente.

O número de neurônios na camada de entrada é, em geral, igual ao número de variáveis de *inputs*. Entretanto, este número pode ser reduzido através do uso de técnicas estatísticas de compressão de dados como a análise de componentes principais (PCA) e mínimos quadrados parciais (PLS), descritas em detalhes nas Seções 3.1 e 3.2. Tal redução permite a eliminação de informações irrelevantes tais como ruído e redundâncias presentes na matriz de dados. O sucesso da redução do número de variáveis de *input* da rede pode resultar no aumento da velocidade de treinamento, menor memória de armazenamento, melhor habilidade de generalização do modelo,

obtenção de um modelo mais robusto com respeito aos ruídos nas medidas e representação de um modelo mais simples.

Já para a camada de saída, o número de neurônios corresponde ao número de variáveis a serem preditas, entretanto, é recomendado que cada modelo apresente uma única resposta (um neurônio), o que diminui o número de parâmetros a serem ajustados e conseqüentemente a carga computacional exigida. Uma exceção a esta regra é para situações onde se deseja predizer diversas respostas correlacionadas, como as concentrações de diferentes constituintes de uma mistura em um sistema fechado.

A definição do número inicial de variáveis de input pode ser feita de duas formas:

- (a) iniciar com um número pequeno de variáveis de *input* e ir adicionando novas variáveis até que não haja mais melhora da performance de predição da rede, ou
- (b) iniciar com todas as variáveis disponíveis e que são consideradas importantes para o modelo como *inputs* e daí ir removendo-as gradativamente até que não mais seja observada melhora na performance da rede.

No segundo caso variáveis irrelevantes podem ser eliminadas durante a modelagem já no primeiro variáveis importantes podem deixar de ser incluídas no modelo. Assim, quando as relações a serem modeladas não são bem entendidas, as técnicas estatísticas de compressão de dados PCA e PLS podem ser utilizadas (MAIER e DANDY, 2000).

Escolha da função de transferência

As funções de transferência geralmente utilizadas em processos químicos são a sigmoidal e a tangente hiperbólica (MAIER e DANDY, 2000 e OLIVEIRA, 2000) que são facilmente diferenciáveis, permitem o ajuste dos mais variados modelos não-lineares, e apresentam um comportamento aproximadamente linear no seu centro, o que torna possível, também, a modelagem de sistemas lineares.

A função (ou funções) de transferência na camada de *output* pode ser linear ou não. Em muitas situações, se o número de neurônios intermediários é suficiente, toda a modelagem é feita na camada intermediária. Um procedimento às vezes útil é trabalhar com ambos os tipos de funções de transferência (linear e não-linear) durante a otimização da topologia da rede e basear a decisão na forma dos resíduos do modelo construído com o mesmo conjunto de variáveis de *input*.

4.2.5 Número de amostras

Assim como em outro método de regressão, existem restrições com relação ao número de amostras requeridas para desenvolver um modelo de redes neurais. O número de parâmetros a serem ajustados (pesos) é um fator diretamente proporcional ao número de amostras disponível para a modelagem, desta forma, o número de amostras consistentes é geralmente um fator limitante para qualquer tipo de RNAs.

É possível obter excelentes resultados para a modelagem de sistemas utilizando um número limitado de dados durante o treinamento. Entretanto, se for tentado validar o modelo para um conjunto independente de dados, geralmente, uma significativa degradação dos resultados será observada devido ao sobre-ajuste (ou *overfitting*) dos parâmetros e consequentemente perda da habilidade de generalização do modelo.

O número de amostras disponíveis é geralmente imposto ou limitado em problemas práticos. Entretanto, apesar de não existir um critério universal para estimativa do número mínimo de amostras para o treinamento, sabe-se que a razão entre o número de amostras e o número de parâmetros ajustáveis (pesos) deve ser mantido tão alto quanto possível de maneira a tentar-se cobrir todo o domínio do problema.

Nota-se portanto que a redução do número de parâmetros ajustáveis da RNA pode ajudar na solução do problema de sobreajuste e carência de dados. Como mencionado anteriormente, uma maneira de alcançar este objetivo é a redução de dimensionalidade do conjunto de variáveis de *input* e de neurônios na camada intermediária.

Divisão de dados para treinamento, validação e teste

Um importante passo no desenvolvimento de um modelo está na divisão do conjunto de dados disponíveis em dois ou três subconjuntos (DESPAGNE e MASSART, 1998 e MAIER e DANDY, 2000):

(a) treinamento, utilizado para estimar os parâmetros do modelo;

(b) validação, utilizada para verificar a habilidade de generalização do modelo frente a amostras independentes do conjunto de treinamento, e

c) teste, utilizado para validar o modelo usando novas amostras.

Se a quantidade de dados disponíveis é pequena, o conjunto de teste ou validação pode ser eliminado. Para as MLPs o problema é complexo pois com o ajuste dos dados de treino, deve-se verificar se o número de neurônios e camadas intermediárias é suficiente, buscando minimizar o erro da predição, o número de parâmetros a serem ajustados e o tempo de treinamento. Portanto, um conjunto de monitoramento adicional, validação ou teste, é necessário para parar o treinamento da rede antes de seu sobreajuste.

Idealmente, para um número Nt de amostras de treino, o conjunto de validação e teste deve conter entre Nt/2 e Nt amostras, respectivamente (DESPAGNE e MASSART, 1998). A performance da rede não deve ser julgada pelo ajuste dos dados de treino, pois estes podem ser ajustados perfeitamente. Agora, se o conjunto de dados for grande o bastante para ser dividido em três subconjuntos, os resultados podem ser apresentados tanto pelo conjunto de monitorização (validação) como pelo conjunto de teste.

A seleção dos dados para compor estes três conjuntos deve ser feita aleatoriamente, entretanto o conjunto de treinamento deve conter a maior variação possível de forma a evitar-se a ocorrência de extrapolação na fase de predição (validação e teste). Algoritmos específicos podem ser usados para selecionar as amostras de treinamento que sejam representativas da população total e que contenham informações sobre as fontes principais de variância, tal como o algoritmo de Kennard-Stone (KENNARD e STONE, 1969). Esta seleção também pode ser realizada após a projeção das amostras em um mapeamento bidimensional com a rede de Kohonen (LOZANO *et al.*, 1995; MAJCEN *et al.*, 1995). Alguns softwares computacionais trazem seus próprios algoritmos de aleatorização, como é o caso do Neurosolutions.

4.2.6 Avaliação da eficiência do modelo MLP

Análises devem ser conduzidas para examinar a eficiência e adequação do modelo construído. O coeficiente de determinação múltipla R^2 , definido pela Equação 3.32, é em geral utilizado como indicativo da performance dos modelos de redes neurais. Para tais modelos, diferentemente do caso de modelos de regressão multivariada, o valor de R^2 nem sempre aumenta indiferentemente se a variável adicionada é significante ou não. Assim como, a inclusão de um novo neurônio intermediário ao modelo MLP nem sempre auxilia a melhorar a sua performance de predição.

Adicionalmente ao coeficiente de determinação múltipla, o erro quadrático médio, definido pela Equação (4.9) é geralmente utilizado para avaliar a performance dos modelos de redes neurais.

4.3 Redes neurais functional-link (FLNs)

4.3.1 Introdução

Uma rede neural convencional perceptron de múltiplas camadas (MLP) contém um ou mais estágios de processamento neural entre suas entradas e saídas. Estes estágios, conhecidos como camadas intermediárias, adicionam complexidade e custo de treinamento (CHEN e BILLINGS, 1992; HENRIQUES, 1999). Quando o número de variáveis de entrada do modelo e o número de amostras disponibilizadas para treinamento aumentam, o processo de definição da topologia da rede e de treinamento torna-se bastante demorado (ALKULAIBI e SORAGHAN, 1997). Entretanto, quando o número de amostras é muito reduzido, ocorre o problema de sobreajuste (OLIVEIRA-ESQUERRE *et al.*, 2002).

Uma maneira de evitar-se o treinamento não linear, inerente às redes MLP, é fazer inicialmente alguma transformação ou expansão não linear das entradas da rede e então combinar os termos resultantes linearmente. A estrutura obtida pode apresentar uma boa capacidade de aproximação não linear e ainda assim a estimação dos pesos da rede continua sendo um problema linear. PAO (1989) se referiu a esta estrutura como *functional link network* (FLN).

Uma rede neural do tipo FLN (PAO, 1989) tem estrutura similar a uma rede MLP de três camadas, exceto pelo fato de empregar termos polinomiais ou trigonométricos na camada interna. O uso desse tipo de função dentro da camada interna tem uma longa história dentro da comunidade de modelagem não linear, onde um número pequeno de termos polinomiais de baixa ordem ou os termos dominantes dentro da série de Fourier têm sido usados para introduzir não linearidades dentro do algoritmo linear convencional, o que evita o treinamento não linear (HARADA, 2001).

4.3.2 Desenvolvimento matemático

As redes FLN são algoritmos de aproximação universal; elas podem aproximar uma função não linear contínua dentro de uma precisão arbitrária, dado um número suficiente de nós na camada interna. Os termos polinomiais têm uma saída diferente de zero para o espaço de entrada total e são chamados funções base de generalização global. Nós lineares são então usados dentro das camadas de entrada e saída.

A estrutura geral de uma FLN é mostrada na Figura 4.5, onde x é o vetor de entradas da rede e $y_i(x)$ é uma das saídas. A camada interna faz uma expansão funcional das entradas da rede, transformando o espaço das entradas, de dimensão *m*, num espaço de dimensão maior, M (M > m). A camada de saída consiste em M nós, sendo cada nó, de fato, um combinador linear. A relação entrada-saída da FLN é definida pela Equação (4.28) que segue,

$$y_i(\mathbf{x}) = \sum_{j=1}^M w_{ij} h_j(\mathbf{x}), \quad 1 \le i \le m$$
(4.28)

Há várias maneiras de se fazer a expansão funcional. A mais fácil e mais usada é a expansão em polinômios. Neste caso o resultado da expansão $h_f(\mathbf{x})$ é uma série de monômios de x.



Figura 4.5: Estrutura de uma FLN (functional-link network).

Pode-se verificar, portanto, que dentre as vantagens das FLNs sobre as MLPs (*multilayer perceptron*) está o fato de que a rede resultante é linear nos pesos, os quais podem ser estimados por métodos lineares e, logo, de convergência garantida.

HENRIQUE e LIMA (1996) propuseram uma modificação na estrutura das FLNs na qual a saída descrita pela Equação (4.28) é transformada por uma função de ativação não linear inversível. A nova saída é dada pela Equação (4.29),

$$y_i(\mathbf{x}) = f_i(\sum_{j=1}^M w_{ij} h_j(\mathbf{x}))$$
(4.29)

onde $f(\cdot)$ é uma função não linear inversível como, por exemplo, a função logarítmica ou a inversa.

Segundo HENRIQUES *et al.* (1999), esta modificação aumenta a capacidade de aproximação não linear das FLNs e a estimativa dos parâmetros ainda permanece um problema linear. Neste caso, o treinamento é feito usando-se uma saída transformada, que é igual à saída original da FLN transformada pela inversa da função de ativação $f(\cdot)$. Por exemplo, se $f(\cdot)$ é uma função logaritmo neperiano, a saída transformada que é usada para o treinamento é exp(y). Desta forma, a estimação dos pesos continua a ser linear (COSTA *et al.*, 1998). Os méritos desta modificação podem ser observados nos trabalhos de COSTA *et al.* (1998; 1999; 2000), HENRIQUES (1999) e HARADA *et al.* (2002).

Há várias maneiras de se fazer a expansão funcional. A mais fácil e mais usada é a expansão em polinômios. Neste caso, o resultado da expansão $h_f(x)$ é uma série de monômios da matriz x. Os monômios gerados para uma expansão polinomial de grau máximo igual a seis são mostrados no Quadro 4.1.

O número de monômios gerados pode ser calculado pela Equação (4.30),

$$M = \frac{(m_x + m_g)!}{m_x! m_g!}$$
(4.30)

onde m_x e m_g representam o tamanho do vetor das entradas e o grau dos monômios, respectivamente.

Grau	0	1	2	3	4	5	6
	1	Zi	$Z_i Z_j$	$Z_i Z_j Z_k$	$Z_i Z_j Z_k Z_l$	$Z_i Z_j Z_k Z_l Z_m$	$Z_i Z_j Z_k Z_l Z_m Z_n$
		$i = 1, n_{z.}$	$i=1, n_z;$	$i=1, n_z;$	$i=1, n_z;$	$i=1, n_z;$	$i=1, n_z;$
			$j=i, n_{z}$	$j=i, n_z;$	$j=i, n_z;$	$j=i, n_z;$	$j=i, n_z;$
Monômios			a management of the second seco	$k=j, n_z$	$k=j, n_z;$	$k=j, n_z;$	$k=j, n_z;$
					$l=k, n_{z.}$	$l = k, n_z;$	$l=k, n_z;$
						$m = l, n_z.$	$m = l, n_z;$
	****						$n=m, n_{z.}$

Quadro 4.1: Expansão polinomial de grau seis.

Uma vez que os monômios são determinados, o estimador ortogonal proposto por BILLINGS *et al.* (1989) pode ser usado para estimar os pesos da rede (w_{ij}) e para eliminar os monômios não significativos. Neste trabalho, o autor mostra a importância da eliminação de termos não significativos, já que simplesmente aumentar a ordem e o grau de não linearidade do modelo levará quase certamente a um modelo excessivamente complexo e a um mau condicionamento numérico.

4.3.3 Algoritmo proposto por BILLINGS et al. (1989)

Considerando o caso de uma única saída (M = 1) e N pontos a ajustar, a modelagem destes pontos pela técnica FLN pode ser escrita da mesma forma que a Equação (3.4), de regressão linear múltipla, ou seja,

$$\mathbf{y} = \mathbf{H}\mathbf{w} + \boldsymbol{\epsilon} \tag{4.31}$$

onde,

$$\mathbf{y} = \begin{bmatrix} y_{i} \\ y_{2} \\ \vdots \\ y_{n} \end{bmatrix} \quad \mathbf{H} = \begin{bmatrix} \mathbf{1} \ h(x_{11}) \ h(x_{12}) \ \dots \ h(x_{1k}) \\ \mathbf{1} \ h(x_{21}) \ h(x_{22}) \ \dots \ h(x_{2k}) \\ \vdots \ \vdots \ \vdots \ \vdots \\ \mathbf{1} \ h(x_{n1}) \ h(x_{n2}) \ \dots \ h(x_{nk}) \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} w_{0} \\ w_{1} \\ \vdots \\ w_{k} \end{bmatrix} \quad \boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\xi}_{i} \\ \boldsymbol{\xi}_{2} \\ \vdots \\ \boldsymbol{\xi}_{n} \end{bmatrix}$$
(4.32)

Considerando a matriz H positiva, esta pode ser decomposta em,

$$\mathbf{H}^{T}\mathbf{H} = \mathbf{A}^{T}\mathbf{D}\mathbf{A} \tag{4.33}$$

onde A é uma matriz triangular superior com elementos diagonais unitários,

$$\mathbf{A} = \begin{bmatrix} 1 & \alpha_{12} & \dots & \alpha_{1M} \\ 0 & 1 & \dots & \alpha_{2M} \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

A Equação (4.31) pode então ser escrita como,

$$\mathbf{y} = \mathbf{Z}\mathbf{g} + \boldsymbol{\epsilon}, \quad i = 1, 2, \dots, n \tag{4.34}$$

onde,

$$\mathbf{Z} = \mathbf{H}\mathbf{A}^{-1} = \begin{bmatrix} \zeta_1(1) & \cdots & \zeta_M(1) \\ \vdots & & \vdots \\ \zeta_1(n) & & \zeta_M(N) \end{bmatrix} \quad \mathbf{e} \quad \mathbf{g} = \mathbf{A}\mathbf{W} = \begin{bmatrix} g_1 \\ \vdots \\ g_M \end{bmatrix}.$$

Observa-se portanto que a matriz Z é ortogonal, pois,

$$\mathbf{Z}^{\mathrm{T}}\mathbf{Z} = (\mathbf{H}\mathbf{A}^{-1})^{\mathrm{T}}(\mathbf{H}\mathbf{A}^{-1}) = \mathbf{D} = \mathrm{Diag}[\sum_{t=1}^{\mathrm{N}} \zeta_{1}^{2}(t), ..., \sum_{t=1}^{\mathrm{N}} \zeta_{M}^{2}(t)]$$
(4.35)

O algoritmo de estimação dos parâmetros g, é dado por, dois passos consecutivos. No primeiro passo (k=1) tem-se,

$$\alpha_{11} = 1, \qquad \zeta_1(t) = h_1(t) \quad e \qquad g_1 = \frac{\sum_{t=1}^{N} \zeta_1(t) y(t)}{\sum_{t=1}^{N} \zeta_1^2(t)}.$$
(4.36)

Para os passos seguintes (k=2,..., M), tem-se,

$$\alpha_{ik} = \frac{\sum_{t=1}^{N} \zeta_{i}(t) h_{k}(t)}{\sum_{t=1}^{N} \zeta_{i}^{2}(t)}, \qquad \zeta_{k}(t) = h_{k}(t) - \sum_{i=1}^{k-1} \alpha_{ik} \zeta_{i}(t) \quad e \qquad g_{k} = \frac{\sum_{t=1}^{N} \zeta_{k}(t) y(t)}{\sum_{t=1}^{N} \zeta_{k}^{2}(t)}$$
(4.37)

onde, $i=1,..., k-1 \in \alpha_{kk} = 1$.

Os pesos wi são calculados de trás para frente. Para o primeiro passo tem-se,

$$\mathbf{w}_{\mathbf{M}} = \mathbf{g}_{\mathbf{M}} \tag{4.38a}$$

e para o segundo passo, tem-se,

$$w_i = g_i - \sum_{k=i+1}^{M} \alpha_{ik} w_k$$
, i=M-1,...,1 (4.38b)

O estimador ortogonal pode ser usado para eliminação de termos não significativos definindo-se taxa de redução de erro devida ao termo *i*, $[err]_i$. Esta taxa dá a proporção da variança da variável dependente explicada por $\zeta_i(t)$.

Segundo o procedimento de eliminação originalmente proposto por BILLINGS *et al.* (1989), inicialmente, todos os $h_i(t)$, i=1,..., M são considerados possíveis candidatos para o cálculo de $\zeta_1(t)$ (diferentemente do algoritmo de estimação de parâmetros, onde $\zeta_1(t)=h_1(t)$). Para i=1,..., M, calcula-se,

$$\zeta_{1}^{(i)}(t) = h_{i}(t), \qquad g_{1}^{(i)} = \frac{\sum_{t=1}^{N} \zeta_{1}^{(i)}(t)y(t)}{\sum_{t=1}^{N} (\zeta_{1}^{(i)}(t))^{2}} \quad e \qquad [err]_{1}^{(i)} = \frac{(g_{1}^{(i)}(t))^{2} \sum_{t=1}^{N} (\zeta_{1}^{(i)}(t))^{2}}{\sum_{t=1}^{N} y^{2}(t)} \qquad (4.39a)$$

encontra-se o máximo de $[err]_1^{(i)}$ e chama-se $[err]_1^{(j)} = max\{[err]_1^{(i)}, 1 \le i \le M\}$. Então, seleciona-se o primeiro termo $\zeta_1(t) = \zeta_1^{(i)}(t)$, tal que $g_1(t) = g^{(j)}_1(t)$ e $[err]_1 = [err]_1^{(j)}$.

Como segundo passo tem-se que todos os $h_i(t)$, i=1,...,M, i \neq j são considerados possíveis candidatos para $\zeta_2(t)$. Para i=1,..., M, i \neq j, e calcula-se então,

encontra-se o máximo de $[err]_2^{(i)}$ e chama-se $[err]_2^{(k)} = max \{ [err]_2^{(i)}, 1 \le i \le M, i \ne j \}$. Então, seleciona-se o segundo termo $\zeta_2(t) = \zeta_2^{(k)}(t)$, tal que $\alpha_{12} = \alpha^{(k)}_{12}$, $g_2(t) = g^{(k)}_2(t)$ e $[err]_2 = [err]_2^{(k)}$.

O procedimento é terminado no passo M_s quando $1 - \sum_{i=1}^{M_s} [err]_i < tolerância desejada, para <math>M_s < M$, ou quando $M_s = M$.

O $1 - \sum_{i=1}^{M_s} [\text{err}]_i$ restante é a proporção da variança da variável dependente não explicada e $100(1 - \sum_{i=1}^{M_s} [\text{err}]_i)$ expressa esta proporção como porcentagem. O mau condicionamento pode ser evitado se, sempre que $\sum_{t=1}^{N} (\zeta_k^{(i)}(t))^2$ for menor que um dado valor pré-definido, $h_i(t)$ não for considerado como candidato no cálculo de $\zeta_k(t)$.

4.3.4 Avaliação da eficiência do modelo FLN

O método usado por HENRIQUE e LIMA (1996), e também neste trabalho, para a estimação dos parâmetros e eliminação de nós não significativos utiliza mínimos quadrados ortogonal apresentado na Seção 4.3.3.

O desempenho da rede é medido pelo coeficiente de correlação linear, R^2 , em unidades percentuais (MILTON e ARNOLD, 1990) e através do erro quadrático médio (*MSE*). Para as redes FLNs nota-se que não existe distinção entre conjunto de validação e teste, tendo em vista que apenas um único conjunto de dados (conjunto de treinamento) é utilizado no ajuste de seus parâmetros.

Uma aplicação com sucesso das FLNs requer um conjunto de funções base que possam representar a função desejada adequadamente dentro do domínio das entradas, mas sem aumentar o número de parâmetros da rede. Modificações no peso ou introdução (ou remoção) de um novo termo dentro da camada interna afetam a saída da rede globalmente e, então, não fica claro como a estrutura deve ser escolhida ou que tipo de relação é armazenada dentro da rede. O sucesso da rede depende criticamente das representações armazenadas dentro dos nós da camada interna (BROWN e HARRIS, 1994). Nos trabalhos de OLIVEIRA-ESQUERRE *et al.* (2003a; 2003c), a técnica PLS é utilizada com sucesso na redução do espaço das variáveis preditoras (matriz x) de forma a auxiliar o mapeamento linear das FLNs.

4.4 Considerações finais sobre o capítulo

O objetivo principal deste capítulo foi apresentar uma base teórica necessária para o entendimento das técnicas de redes neurais artificiais utilizadas neste trabalho para a construção de modelos de predição da DBO.

O futuro da aplicação de redes neurais é promissor pelas suas características principais. Diz-se que elas são capazes de "aprender" relações do tipo entrada-saída de um sistema a partir de um processo de "treinamento" semelhante ao aprendizado de um cérebro humano, sem necessitar da idealização, elaboração e validação de modelos matemáticos complexos.

Uma outra característica é sua estrutura de distribuição paralela de informações. Concebida de forma semelhante ao cérebro, esse paralelismo vem atraindo a atenção de inúmeros pesquisadores de todos os setores do conhecimento, em particular os da área de informática. Já se sabe que os computadores só conseguirão aumentar sua capacidade e velocidade de cálculo adotando-se uma estrutura essencialmente paralela de fluxo de informações, como a que ocorre no tecido nervoso cerebral.

A rede FLN aparece como uma alternativa mais simples que a MLP na modelagem de sistema lineares e não lineares. Neste trabalho é realizada uma análise comparativa entre os modelos construídos por ambas as técnicas. Vale ressaltar que um modelo excessivamente complexo exige não somente grande esforço computacional, como também tem valor prático limitado, já que muitas vezes o ruído também está sendo modelado. Para a maior parte dos sistemas reais, desde que os termos significativos do modelo sejam detectados, modelos relativamente pequenos são suficientes para descrever dinâmicas altamente não lineares.

O software comercial Neurosolutions Professional foi utilizado para o desenvolvimento dos modelos MLP. Já os modelos FLN foram construídos utilizando um programa desenvolvido por HENRIQUE (1999). Tal programa foi cedido pela pesquisadora Dra. Aline C Costa, do Laboratório de Otimização de Processos Químicos da Faculdade de Engenharia Química da UNICAMP.

Foram testadas várias configurações das redes MLP variando-se o número de neurônios distribuídos em uma camada intermediária, para os diferentes conjuntos de dados construídos, sendo a função sigmoidal utilizada para ativação dos neurônios intermediários e de saída.

O algoritmo de treinamento das redes MLP delta-bar-delta foi utilizado neste estudo. Como já foi dito anteriormente, a rede mais adequada não é aquela onde se observa o menor erro para o conjunto usado no treinamento, mas sim aquela que apresenta o menor erro para o conjunto de validação e teste. O procedimento "save best" do Neurosolutions foi então utilizado de forma a fazer-se com que a performance da rede MLP que estava sendo treinada fosse periodicamente avaliada por um conjunto de validação; e quando o erro deste conjunto diminuísse, os pesos da rede fossem salvos. Assim, ao final do processo, apenas um único arquivo da rede era salvo, correspondente à melhor rede obtida. Realizado o treinamento, a rede foi testada para um ou dois novos conjuntos.

Expansões polinomiais de ordem um a três foram realizadas para gerar monômios não lineares utilizados como *inputs* das redes FLNs. O estimador ortogonal proposto por BILLINGS *et* *al.* (1989) foi utilizado para calcular os pesos da rede e para eliminar os monômios que não eram significantes para explicar a variância das variáveis de saídas, DBO de entrada e saída.

De acordo com a modificação na estrutura das FLNs proposta por HENRIQUE (1999), a saída da rede é transformada em uma função não-linear, diferenciável e contínua. Esta função de ativação foi aqui escolhida através do método de tentativa e erro. As três funções utilizadas neste trabalho foram as funções de ativação sigmoidal tangente hiperbólica (tansig), sigmoidal logística (logsig) e sigmoidal logística centrada (logsic), representadas pelas Equações (4.40a)-(4.40c), respectivamente.

$$y = f(T) = \frac{1}{2} log \left(\frac{1+T}{1-T}\right) \quad (4.40a) \quad y = f(T) = log \left(\frac{T}{1-T}\right) \quad (4.40b) \quad y = f(T) = log \left(\frac{0.5+T}{0.5-T}\right) \quad (4.40c)$$

onde $T = \sum_{j=1}^{M} w_{ij} h_j(\mathbf{x})$.

CAPÍTULO 5

RESULTADOS E DISCUSSÕES

5.1Pré-processamento de dados

5.1.1 Dados 1

Inicialmente é realizado o pré-processamento de dados do conjunto Dados 1, composto por oito variáveis preditoras. Como descrito anteriormente, a técnica de PCA combina as variáveis independentes linearmente de maneira a maximizar a variância dos dados. Nas Tabelas A1 e A2 do Anexo 2 são apresentados os autovalores e as variâncias explicada e acumulada das componentes principais (CPs) obtidas, para os modelos estacionário e dinâmico, respectivamente.

Aplicando o critério de TODESCHINE (1997) para a matriz de Dados 1, são obtidos para o modelo estacionário K = 0,3218 (que indica um grau baixo de correlação), KL = 6 (representando uma variância acumulada de 91,6%) e KP = 4 (representando 75,6% de variância acumulada). Para o modelo dinâmico são obtidos, K = 0,4243 (que indica um grau baixo de correlação), KL = 9,6354 (representando uma variância acumulada de 92,0%) e KP = 5 (representando 71,0% da variância acumulada).

Considerando então as componentes que apresentam autovalor maior que 1 (MARDIA et al., 1979), as três primeiras CPs devem permanecer no modelo para o modelo estacionário, discordando do resultado obtido pelo método de Todeschine. Para o modelo dinâmico, o resultado de Todeschine está de acordo com o sugerido por Mardia.

Foi obtido portanto uma redução de dimensionalidade significativa, tendo em vista que o sistema de oito dimensões foi reduzido a quatro, para o modelo estacionário, e de dezesseis para cinco, para o modelo dinâmico.

Nas Tabelas A4 e A5 do Anexo 2 são também apresentados os *loadings* das CPs, para o modelo estacionário e dinâmico, respectivamente. Os gráficos dos *loadings* e *scores* das quatro primeiras CPs são apresentados nas Figuras 5.1a e 5.1b, para o modelo estacionário e dinâmico, respectivamente.



Figura 5.1a: Gráfico dos *loadings* e *scores* correspondentes às quatro primeiras CPs (Dados 1 – Modelagem estacionária).



Figura 5.1b: Gráfico dos *loadings* e *scores* correspondentes às quatro primeiras CPs (Dados 1 - Modelagem dinâmica).

Nenhum significado físico das componentes é claramente observado, ocasionando maior dificuldade de interpretação. Observa-se entretanto que, apesar da DQO de entrada ser a variável mais correlacionada com a DBO de saída, sua contribuição somente é importante para a segunda componente. Uma interpretação mais detalhada do significado das componentes pode ser realizada com o auxílio de um ou mais especialistas do processo.

Como esperado, nos gráficos dos *scores* das primeiras componentes da Figura 5.1 (a) e (b), não se observa tendência dos dados, confirmando a ausência de correlação entre as componentes. Algumas amostras apresentam-se relativamente distantes da massa de dados, o que pode indicar a presença de pontos influentes ou *outliers* na matriz de Dados 1.

Apesar do uso dos gráficos dos *scores* ou das cartas de controle multivariado das componentes principais parecerem ser uma alternativa viável na identificação de *outliers*, não é possível analisar graficamente todas as componentes ao mesmo tempo. Além disso, a falta de uma fronteira nos gráficos dos *scores* não permite decidir com certeza que dados do conjunto de validação podem estar fora do domínio.

Para contornar esta dificuldade de interpretação, o parâmetro T^2 de Hotelling é utilizado aqui para identificar possíveis *outliers* no conjunto de validação. Nas Figuras 5.2 e 5.3 são apresentados os gráficos de T^2 considerando as quatro e cinco primeiras CPs, para os modelos estacionário e dinâmico, respectivamente. Nestas figuras verifica-se que aproximadamente 7% de amostras tanto do conjunto de modelagem, tanto como dos conjuntos de validação e teste, apresentam T^2 acima do limite de controle. Para o modelo estacionário, as amostras 78 e 144, do conjunto de validação, e 376, do conjunto de teste, apresentam T^2 superiores aos do conjunto de modelagem. Destas amostras, apenas a 144, do conjunto de validação, continua apresentando T^2 superior aos do conjunto de modelagem, para o modelo dinâmico.

Uma análise detalhada destes resultados foi realizada em conjunto com especialistas do processo. Estes não encontram explicação física que justificasse a exclusão das amostras que apresentam T^2 acima do limite de controle. Tais especialistas expuseram entretanto a relevância na construção de modelos que levassem em consideração tais amostras, que poderiam vir a serem utilizados para auxiliar na atuação no processo em situações críticas.

Desta forma, os modelos de predição da DBO de entrada e saída são construídos considerando todas amostras. Entretanto, uma análise criteriosa dos erros individuais obtidos para cada amostra que apresenta T^2 acima dos limites de controle é realizada.



Figura 5.2: Carta de T^2 das componentes principais para o modelo estacionário – Dados 1; (a) Conjunto de modelagem; (b) Conjuntos de validação e teste.



Figura 5.3: Carta de T^2 das componentes principais para o modelo dinâmico – Dados 1; (a) Conjunto de modelagem; (b) Conjuntos de validação e teste.

5.1.2 Dados 2

Para o conjunto Dados 2, composto por todas as 12 variáveis de entrada (preditoras) disponibilizadas pela empresa, apenas modelos estacionários foram construídos.

Na Tabela A3 do Anexo 2 são apresentados os autovalores e variâncias explicada e acumulada das componentes principais (CPs) obtidas.

Aplicando o critério de TODESCHINE (1997) para a matriz de dados DADOS 2, K = 0,3614 (que indica um grau baixo de correlação), KL = 8 (representando uma variância acumulada de 91,3 %) e KP = 5 (representando 74,8 % da variância acumulada).

Considerando então as componentes que apresentam autovalor maior que 1, as cinco primeiras CPs devem ser utilizadas para a construção dos modelos, concordando com o resultado obtido pelo método de Todeschine. Foi obtido portanto uma redução de dimensionalidade significativa, tendo em vista que o sistema de doze dimensões foi reduzido a cinco.

Na Tabela A6 do Anexo 2 são apresentados os *loadings* das CPs e na Figura 5.4 os gráficos dos *loadings* e *scores* das quatro primeiras CPs.



Figura 5.4: Gráfico dos loadings e scores correspondentes às quatro primeiras CPs (Dados 2).

Do gráfico dos *loadings* observa-se que a primeira componente, que carrega 28,1% da variação total dos dados, representa um contraste entre a condutividade *versus* a vazão, temperatura e produção de celulose. A segunda componente, com 15% da variação dos dados, traz informação da DQO e sólidos suspensos, parecendo representar um fator microbiológico. Para as terceira, quarta e quinta componentes, não ocorre concordância quanto à importância das variáveis. Nenhum significado físico destas componentes é claramente observado, ocasionando maior dificuldade de interpretação. Mais uma vez, uma interpretação mais detalhada do significado das componentes pode ser realizada com o auxílio de um ou mais especialistas do processo.

Como esperado, nos gráficos dos *scores* das primeiras componentes, não se observa tendência dos dados, confirmando a ausência de correlação entre as componentes. Algumas amostras apresentam-se relativamente distantes da massa de dados, o que pode indicar novamente a presença de amostras influentes ou *outliers* na matriz de Dados 2.

Na Figura 5.5 são então apresentados os gráficos da estatística de T^2 de Hotelling, considerando as cinco primeiras componentes principais, identificadas como mais importantes pelo método de Todeschine e por Mardia.



Figura 5.5: Carta de T² das componentes principais – Dados 2, (a) Conjunto de modelagem; (b) Conjuntos de validação (teste).

Na Figura 5.5, verifica-se que 7 amostras do conjunto de treinamento apresentam T^2 acima do limite de controle, já para o conjunto de validação, apenas 1 amostra encontra-se acima do limite. Mais uma vez, modelos de predição são construídos sem exclusão das amostras que se apresentam como possíveis outliers, entretanto, uma análise criteriosa dos seus erros de predição é realizada.

Na Sessão 5.2 são apresentados os resultados e discussões referentes aos modelos de predição da DBO de saída e DBO de entrada, respectivamente. Os conjuntos Dados 1 e Dados 2 são utilizados para a construção, validação e teste destes modelos.

5.2 Modelagem

5.2.1 Dados 1 - Predição da DBO de entrada

Modelagem via técnicas de regressão linear multivariada

Modelos de predição da DBO de entrada são inicialmente construídos utilizando o conjunto de Dados 1.

O método stepwise foi o primeiro a ser utilizado na construção dos modelos lineares. Utilizou-se $\alpha = 0,10$ (nível de 90% de confiança) e $\alpha = 0,15$ (nível de 85% de confiança), para os testes de inclusão e eliminação de variáveis. Considerando os resultados obtidos pelo método stepwise, a produção de papel e a condutividade não foram utilizadas, para a construção do modelo estacionário. Para o modelo dinâmico, os seguintes parâmetros foram incluídos no modelo: DQO, VAZ, pH, CEL, e os valores passados da COR e T. Os resultados de ambos modelos são apresentados na Tabela 5.1.

Como descrito anteriormente, a técnica PLS direciona as variáveis independentes de maneira a maximizar a variância dos dados levando-se em consideração a variável dependente. Nas Tabelas A1 e A2 do Anexo 2 são apresentadas as variâncias explicada e acumulada das componentes obtidas, para os modelos estacionário e dinâmico, respectivamente.

Analisando as variâncias apresentadas pelas CPs e VLs, como esperado, tanto para o modelo estacionário como para o dinâmico, observa-se que as CPs descrevem uma quantidade maior de variação que pode explicar o conjunto de variáveis dependentes (**X**) que as VLs. Por outro lado, as VLs descrevem mais variância da variável dependente (DBO) que as CPs.

Considerando então um limiar de pelo menos 2% na variância acumulada com a inclusão de uma VL no modelo, tem-se que as três primeiras VLs devem permanecer no modelo PLS estacionário e dinâmico. A inclusão da quarta VL ao modelo representa apenas um acréscimo de 1,0% e 1,57% na variância acumulada para o modelo estacionário e dinâmico, respectivamente. Obtém-se portanto uma redução de dimensionalidade superior à obtida através do método de Todeschine para as componentes principais.

Nas Tabelas A7 e A8 do Apêndice 2 são apresentados os *loadings* das VLs, dos modelos estacionário e dinâmico, respectivamente. Para os modelos estacionário e dinâmico tem-se que a primeira variável latente, que carrega 21,93% e 19,75% da variação total dos dados e representa

31,46% e 26,92% da variação da variável dependente, respectivamente, traz informação principalmente da DQO, seguida pela produção de celulose e temperatura. Apenas a variável COR não se apresenta significante para as três primeiras variáveis tanto para modelo estacionário, como para o modelo dinâmico. Na Tabela 5.1 são apresentados os resultados de modelagem. Nas Figuras 5.6 (a) a (c) são apresentados os gráficos de probabilidade dos resíduos para os modelos MLR e PLSR, respectivamente.

	^a Modelo	$R^2 a d j. (R^2)$	^b MSE	^c p-valor	^d F
	MLR - 8	41,8 (42,6)	11,64	0	dl1 01
Modelagem	MLR – 5	41,5 (42,0)	11,65	0	1,01
estacionária	PLSR-8	41,8 (42,6)	11,64	0	d21 10
	PLSR-3	41,8 (42,1)	11,74	0	-1,10
	MLR - 16	43,7 (45,1)	11,12	0,008	d3 t < 1
Modelagem	MLR – 6	43,1 (43,7)	11,34	0,001	-1,01
dinâmica	PLSR - 16	43,7 (45,1)	11,12	0,008	d40.02
	PLSR - 3	43,8 (44,0)	11,34	0,011	-0,93

Tabela 5.1: Resultados MLR e PLSR obtidos para o conjunto de modelagem - DBO de entrada (Dados 1).

Nota: Os valores de p-valor para o teste de significância na regressão dos oito modelos são menores que 0,05. ^a Modelo construído e número de variáveis independentes. ^b Os resultados MSE devem ser multiplicados por 10⁻⁴. ^c P-valor para o teste de normalidade dos resíduos. ^d Teste indicado por BARROS *et al.* (1996): ^{d1} $F_{0,05; 3; 617} = 2,60$, ^{d2} $F_{0,05; 5; 617} = 2,21$, ^{d3} $F_{0,05; 10; 609} = 1,83$ e ^{d4} $F_{0,05; 13; 609} \cong 1,70$

Analisando os resultados obtidos para o teste F e de R^2 *ajustado*, tem-se que tanto para os modelos estacionários, como para os dinâmicos, não existe ganho em se utilizar modelos que contém todas as variáveis originais (MLR – 8 e PLSR – 16) ou todas VLs (PLSR – 8 e PLSR – 16). Nenhum modelo passou no teste de normalidade dos resíduos, como pode ser verificado na Figura 5.6.



Figura 5.6: Teste de normalidade dos resíduos dos modelos para DBO de entrada; (a) MLR -5 estacionário; (b) MLR - 6 dinâmico; (c) PLSR - 3 estacionário; (d) PLS - 3 dinâmico.

Com a inclusão da quarta VL, os modelos estacionário e dinâmico passariam a apresentar um R^2 ajustado de 42,1% e 44,4%, respectivamente, permanecendo os mesmos valores do p-valor para o teste dos resíduos. Assim, concordando com o critério utilizado para a definição do número de VLs a serem utilizadas na modelagem, nenhuma melhora significativa na performance de predição PLS é obtida com a inclusão de mais VLs.

Os resultados obtidos para os conjuntos de validação e teste são apresentados na Tabela 5.2. Nesta, verifica-se novamente que não há diferenças significativas entres os maiores e menores modelos, nem mesmo entre os modelos estacionários e dinâmicos. Entretanto, verifica-se que informações importantes foram perdidas com a exclusão das variáveis originais (pelo método *stepwise*) ou das VLs. Por sua simplicidade e conseqüente maior facilidade de implementação, o modelo MLR estacionário com cinco variáveis preditoras pode ser considerado o melhor modelo linear de predição da DBO de entrada.

		Val	idação		Т	este		Média ± Desvio Padrão				
	^a Modelo	R^2	^b MSE	R^2	^b MSE	R^2	^b MSE	$\overline{R}^2 \pm \sigma$	$^{b}\overline{MSE}\pm\sigma$			
	MLR - 8	47,6	11,8	47,0	14,9	54,1	14,0	50,6 ± 3,9	13,5 ± 1,6			
Modelagem	MLR - 5	45,8	12,1	46,9	14,9	54,7	13,9	49,2±4,9	13,6±1,4			
estacionária	PLSR - 8	47,6	11,8	47,0	14,9	54,1	14,0	$50,6 \pm 3,9$	$13,5 \pm 1,6$			
	PLSR - 3	48,0	11,6	46,6	15,0	52,8	14,3	$49,2 \pm 3,3$	13,6±1,8			
	MLR - 16	47,1	11,9	47,4	14,8	55,0	13,8	$49,8 \pm 4,5$	$13,5 \pm 1,5$			
Modelagem	MLR - 6	47,0	11,9	47,0	14,9	53,4	14,3	$46,2 \pm 3,7$	$13,7 \pm 1,6$			
dinâmica	PLSR - 16	47,1	11,9	47,4	14,8	55,0	13,8	$49,8 \pm 4,5$	$13,5 \pm 1,5$			
	PLSR - 3	45,8	12,2	46,2	15,1	54,2	14,0	48,8±4,7	$13,8 \pm 1,5$			

Tabela 5.2: Resultados dos modelos MLR e PLSR obtidos para os conjuntos de validação e teste – DBO de entrada (Dados 1).

^a Modelo construído e número de variáveis independentes. ^bOs resultados MSE devem ser multiplicados por 10⁻⁴.

O gráfico da variável predita versus a medida mostra que, embora ainda longe do perfeito, o modelo MLR – 5 estacionário é capaz de fornecer resultados de predição da DBO de entrada dentro de um intervalo de predição de 95%, como pode ser observado na Figura 5.7. Isto também pode ser observado na Figura 5.8, onde é apresentada uma representação gráfica dos valores da DBO de entrada medidos e preditos para este modelo. Cerca de 95% das amostras apresentam desvios relativos entre a DBO média e a predita menor que 10%.



Figura 5.7: Relação entre a DBO de entrada predita *versus* medida (linha sólida) – Dados 1. As linhas tracejadas superior e inferior indicam o intervalo de predição de 95% para o modelo estacionário MLR –5.



Figura 5.8: Gráficos de séries temporais – Dados 1; (a) DBO de entrada medida e predita para o modelo estacionário MLR – 5; (b) Resíduos – Linhas superior e inferior indicam um intervalo de confiança de 95%.

Modelos PCR foram também construídos. Primeiramente o método *stepwise* foi utilizado para eliminação das componentes menos correlacionadas com a variável dependente. Os resultados indicaram que apenas a componente cinco deveria ser excluída do modelo estacionário, sendo obtido um modelo com R^2 ajustado praticamente igual ao obtido pelo método MLR, 41,8%. A performance dos modelos PCR são piores, R^2 ajustado igual a 23,8%, quando apenas quatro primeiras componentes principais são utilizadas na construção do modelo. Para o modelo dinâmico tem-se que nove componentes (CP6, CP8, CP9, CP11, CP12, CP13, CP14, CP15 e CP16) devem ser excluídas do modelo, sendo obtido um R^2 ajustado igual a 42,2%.

Quando apenas as cinco componentes identificadas como significantes pelo método de Todeschine são utilizadas para a construção do modelo dinâmico, um R^2 igual a 15,7% é obtido.

Modelagem via redes neurais artificiais

Modelos FLN foram inicialmente construídos. Para os modelos estacionários foram obtidos nove e quarenta e cinco monômios, para expansões polinomiais de grau um e dois, respectivamente. Já para os modelos dinâmicos, dezessete e cento e cinqüenta e três monômios foram obtidos, para expansões de ordem um e dois, respectivamente. Quando apenas as três primeiras VLs foram usadas como entrada das FLNs, quatro e dez monômios foram obtidos, para expansões de ordem um e dois, respectivamente obtidos, para expansões de ordem um e dois, respectivamente dez monômios foram obtidos, para expansões de ordem um e dois, respectivamente, tanto para os modelos estacionários como dinâmicos.

Nas Tabelas 5.3 e 5.4 são apresentados os resultados de modelagem quando utilizados graus de monômios iguais a um e dois, respectivamente. Os resultados de validação e teste são apresentados na Tabela 5.5 (modelos FLNs de primeira ordem) e Tabela 5.6 (modelos FLNs segunda ordem). Das três funções de transferência testadas, a função logsig apresentou melhores resultados considerando os valores de R^2 e *MSE*.

Tabela	5.3:	Resultados	FLN	(ordem	1)	obtidos	para	0	conjunto	de	modelagem	-	DBO	de	entrada
(Dados	1).														

	^a Modelo	R^2	^b MSE	° p-valor
	FLN – 8	42,2	11,75	0
Modelagem Estacionária	PLS-FLN – 8	42,2	11,75	0
Estacionaria	PLS-FLN – 3	41,3	14,15	0
	FLN – 16	44,8	11,23	0,007
Modelagem Dinâmica	PLS-FLN – 16	44,8	11,22	0,007
	PLS-FLN – 3	44,0	11,37	0,006

^a Modelo construído e número de variáveis independentes. ^b Os resultados MSE devem ser multiplicados por 10⁻⁴. ^c P-valor para o teste de normalidade dos resíduos.
	^a Modelo	R^2	^b MSE	°p-valor
	FLN-8	54,3	9,27	0,022
Modelagem	PLS-FLN-8	54,3	9,27	0,022
estacionaria	PLS-FLN-3	44,2	11,32	0,001
A7 - 3 - 3	FLN - 16	67,2	6,64	0
Modelagem	PLS-FLN-16	67,3	6,62	0
umannea	PLS-FLN-3	46,8	10,8	0,030

Tabela 5.4: Resultados FLN (ordem 2) obtidos para o conjunto de modelagem – DBO de entrada (Dados 1).

^a Modelo construído e número de variáveis independentes. ^b Os resultados MSE devem ser multiplicados por 10⁻⁴. ^c P-valor para o teste de normalidade dos resíduos.

Tabela 5.5: Resultados FLN (ordem 1) obtidos para os conjuntos de validação e teste - DBO de entrada (Dados 1).

		Vali	idação		Te	ste		Média ± De	svio Padrão
	^a Modelo	R^2	^b MSE	R^2	^b MSE	R^2	^b MSE	$\overline{R}^2 \pm \sigma$	$b \overline{MSE} \pm \sigma$
	FLN-8	46,1	25,8	44,8	20,5	54,8	18,2	48,6±5,4	$21,5 \pm 3,9$
Modelagem	PLS-FLN-8	46,4	22,1	45,7	23,0	54,5	13,6	$48,9\pm4,9$	19,6 ± 5,2
Colacional la	PLS-FLN – 3	46,7	22,4	45,2	23,4	53,2	13,9	48,4 ± 4,3	19,9 ± 5,2
	FLN - 16	44,7	24,6	45,0	19,9	55,4	15,5	48,4 ± 6,1	$20,0 \pm 4,6$
Modelagem	PLS-FLN-16	45,9	24,9	46,2	27,9	55,6	13,7	$49,2 \pm 5,5$	$22,2 \pm 7,5$
umamva	PLS-FLN-3	45,2	24,3	44,5	28,4	54,8	13,8	48,2 ± 5,8	$22,2 \pm 7,5$

^a Modelo construído e número de variáveis independentes. ^bOs resultados MSE devem ser multiplicados por 10⁻⁴.

Tabela 5.6: Resultados FLN (ordem 2) obtidos para os conjuntos de validação e teste – DBO de entrada (Dados 1).

	·	Vali	idação		Te	ste		Média ± Do	svio Padrão
	^a Modelo	R^2	^b MSE	R^2	^b MSE	R^2	^b MSE	$\overline{R}^2 \pm \sigma$	$b \overline{MSE} \pm \sigma$
	FLN – 8	34,6	33,5	41,2	36,0	46,3	23,0	40,7 ± 5,9	30,8 ± 6,9
Modelagem	PLS-FLN – 8	37,6	39,0	47,5	24,2	38,0	18,8	$41,0 \pm 5,6$	$27,3 \pm 10,5$
cstactonal la	PLS-FLN – 3	47,3	20,7	43,3	21,6	54,9	13,7	48,5 ± 5,9	18,7 ± 4,3
	FLN – 16	26,6	62,7	10,8	92,4	22,9	33,6	20,1 ± 8,3	62,9 ± 29,4
Modelagem	PLS-FLN - 16	27,5	54,2	10,8	83,3	24,1	33,3	$20,8\pm8,8$	56,9 ± 25,1
umannca	PLS-FLN-3	47,4	23,3	43,2	27,4	53,9	13,9	48.2 ± 5.4	21.5 ± 6.9

^a Modelo construído e número de variáveis independentes. ^bOs resultados MSE devem ser multiplicados por 10⁻⁴.

A perda de generalização das redes FLN com a inclusão de monômios de ordem dois pode ser notada claramente, tendo em vista que para o conjunto de modelagem tais modelos pareciam ser os melhores, discordando dos resultados obtidos para os conjuntos de validação e teste. Os resultados de validação e teste dos modelos FLN indicam que a redução da dimensionalidade das entradas da rede pela exclusão das últimas VLs realmente melhora a acurácia dos modelos de FLN de segunda ordem; o mesmo comportamento não é observado para o conjunto de modelagem. Para os modelos FLN de primeira ordem, nenhuma melhora significativa é observada para a mesma situação. De uma forma geral, os modelos estacionários apresentam performance superior aos modelos dinâmicos, com exceção do modelo dinâmico FLN (ordem 2) com três VLs que apresenta resultados de predição para os dados de validação e teste similares aos modelos estacionários.

Modelos de redes neurais MLP foram então construídos utilizando o algoritmo de treinamento DBD e a função de transferência sigmoidal. Diferentes topologias das redes foram testadas e os melhores resultados MLP obtidos são apresentados nas Tabelas 5.7 e 5.8, para os conjuntos de modelagem e validação (e teste), respectivamente.

Assim como ocorreu para o caso das redes FLN de primeira ordem, observa-se que a técnica PLS não auxilia o mapeamento não linear das redes MLP para o caso em estudo. Este fato pode ser explicado através dos seguintes fatores:

- a) O espaço de variáveis preditoras já ter sido previamente reduzido através da exclusão das quatro variáveis que continham muitas lacunas em suas medições, e
- b) A rede MLP ter sido capaz de encontrar relações entre as variáveis preditoras e a DBO de entrada, sem a necessidade de pré-processar suas variáveis originais.

Nota-se que mesmo utilizando uma quantidade menor de informação das variáveis preditoras, os modelos estacionários podem apresentar performances similares ou até mesmo superiores que os modelos dinâmicos. Considerando os resultados obtidos para o conjunto de validação e teste, assim como para o conjunto de modelagem, o modelo MLP estacionário com 10 neurônios intermediários fornece os melhores resultados de predição.

Como esperado, o gráfico de ajuste da DBO de entrada predita versus a medida mostra que o modelo MLP estacionário com dez neurônios é capaz de fornecer resultados de predição dentro de um intervalo de predição de 95%, como pode ser observado na Figura 5.9.

	^a Modelo	R^2	^b MSE	°p-valor
uyuun jan ka ja	MLP-8 (n=10)	64,3	7,3	0
Modelagem	PLS-MLP - 8 (n=4)	48,5	11,1	0,013
Estacionaria	PLS-MLP - 3 (n=10)	43,1	15,5	0
	MLP – 16 (n=8)	63,5	7,9	0,070
Modelagem	PLS-MLP – 16 (n=4)	51,0	10,7	0,042
umannea	PLS-MLP - 3 (n=10)	47,2	11,0	0,016

Tabela 5.7: Resultados MLP obtidos para o conjunto de modelagem - DBO de entrada (Dados 1).

^a Modelo construído e número de variáveis independentes. ^b Os resultados MSE devem ser multiplicados por 10^{-4} . ^c P-valor para o teste de normalidade dos resíduos. n = número de neurônios intermediários.

Tabela 5.8: Resultados MLP obtidos para os conjuntos de validação e teste - DBO de entrada (Dados 1).

		Vali	dação		Te	ste		Média ± De	svio Padrão
	^a Modelo	R^2	^b MSE	R^2	^b MSE	R^2	^b MSE	$\overline{R}^2 \pm \sigma$	$b \overline{MSE} \pm \sigma$
	MLP-8 (n=10)	60,1	9,0	59,8	11,1	65,3	10,3	61,7 ± 3,1	10,1 ± 1,1
Modelagem estacionária	PLS-MLP - 8 (n=4)	46,9	12,6	44,4	16,7	54,5	13,8	48,6±5,3	$14,4 \pm 2,1$
	PLS-MLP - 3 (n=10)	44,2	13,6	40,9	22,3	53,0	17,2	$46,0\pm6,3$	$17,7 \pm 4,4$
	MLP - 16 (n=8)	52,0	11,1	55,5	12,3	55,6	13,4	$54,4 \pm 2,1$	$12,3 \pm 1,2$
Modelagem	PLS-MLP – 16 (n=4)	51,0	13,4	47,2	15,5	42,1	17,8	46,8 ± 4,5	15,6±2,2
unamica	PLS-MLP - 3 (n=10)	46,6	12,4	37,1	18,2	56,0	13,2	$46,6 \pm 9,5$	$14,6 \pm 3,1$

^a Modelo construído e número de variáveis independentes. ^bOs resultados MSE devem ser multiplicados por 10⁻⁴. n = número de neurônios intermediários.



Figura 5.9: Relação entre a DBO de entrada predita versus medida (linha sólida) – Dados 1. As linhas tracejadas superior e inferior indicam o intervalo de predição de 95% para o modelo estacionário MLP com dez neurônios.

Na Figura 5.10 é apresentada uma representação gráfica dos valores da DBO de entrada medidos e preditos para este modelo. Mais de 95% das amostras apresentam desvios relativos entre a DBO medida e a predita menor que 10%.



Figura 5.10: Gráficos de séries temporais – Dados 1; (a) DBO de entrada medida e predita para o modelo estacionário MLP com dez neurônios; (b) Resíduos – Linhas superior e inferior indicam um intervalo de confiança de 95%.

Realizando uma análise geral dos resultados, é possível observar que, independente da técnica de modelagem utilizada, os melhores modelos de predição da DBO de entrada são os estacionários.

5.2.2 Dados 1 - Predição da DBO de saída

Modelagem via técnicas de regressão linear multivariada

Modelos de predição da DBO de saída são inicialmente construídos utilizando o conjunto de Dados 1.

O método stepwise foi o primeiro a ser utilizado na construção dos modelos lineares. Mais uma vez, utilizou-se $\alpha = 0,10$ (nível de 90% de confiança) e $\alpha = 0,15$ (nível de 85% de confiança), para os testes de inclusão e eliminação de variáveis.

Na predição da DBO de saída os resultados obtidos pelo método stepwise indicam que a DQO, a vazão, a condutividade e a temperatura devem ser utilizadas, para a construção do modelo estacionário. Já para o modelo dinâmico, além da DQO, vazão, condutividade e temperatura, os valores passados da DQO e vazão devem ser utilizados como variáveis preditoras. Os resultados de ambos modelos são apresentados na Tabela 5.9.

Modelos PLSR foram então construídos para a predição da DBO de saída. Na Tabela A1 e A2 do Anexo 2 são apresentadas as variâncias explicada e acumulada das componentes obtidas, para os modelos estacionário e dinâmico, respectivamente.

Comparando-se as variâncias apresentadas pelas CPs e VLs, tanto para o modelo estacionário como para o dinâmico, observa-se que as CPs descrevem uma quantidade maior de variação que pode explicar o conjunto de variáveis dependentes (**X**) que as VLs. Por outro lado, as VLs descrevem mais variância da variável dependente (DBO) que as CPs.

Considerando então um limiar de pelo menos 2% na variância acumulada com a inclusão de uma VL no modelo, tem-se que as três primeiras VLs devem permanecer no modelo PLS estacionário e dinâmico. A inclusão da quarta VL ao modelo representa apenas um acréscimo de 1,23% e 1,02% na variância acumulada para o modelo estacionário e dinâmico, respectivamente. Obtém-se portanto uma redução de dimensionalidade superior à obtida através do método de Todeschine para as componentes principais.

Nas Tabelas A10 e A11 do Anexo 2 são apresentados os *loadings* das VLs, dos modelos estacionário e dinâmico, respectivamente. Para os modelos estacionário e dinâmico tem-se que a primeira variável latente, que carrega 20,25% e 17,99% da variação total dos dados e representa 23,30 % e 24,75% da variação da variável dependente, respectivamente, traz informação

principalmente da DQO, seguida pela produção de celulose e temperatura. Apenas a variável COR não se apresenta significante para as três primeiras variáveis tanto para modelo estacionário, como para o modelo dinâmico. Na Tabela 5.9 são apresentados os resultados de modelagem.

	^a Modelo	$R^2 adj.(R^2)$	^b MSE	°p-valor	^d F
*********	MLR - 8	31,1 (32,0)	25,3	0,004	dla cc
Modelagem	MLR - 4	30,8 (31,3)	25,6	0,003	-1,66
estacionária	PLSR-8	31,1 (32,0)	25,33	0,004	d20.97
	PLSR - 3	31,2 (31,5)	25,51	0,002	-0,80
	MLR - 16	36,6 (38,2)	23,0	0,054	d31 04
Modelagem	MLR - 6	35,7 (36,3)	23,7	0,025	1,94
dinâmica	PLSR - 16	36,6 (38,2)	23,01	0,054	d40 52
	PLSR-3	37,2 (37,5)	3,27	0,031	0,55

Tabela 5.9: Resultados MLR e PLSR obtidos para o conjunto de modelagem - DBO de saída (Dados 1).

Nota: Os valores de p-valor para o teste de significância na regressão dos oito modelos é menor que 0,05.

^a Modelo construído e número de variáveis independentes. ^b Os resultados MSE devem ser multiplicados por 10⁻⁴. ^c P-valor para o teste de normalidade dos resíduos. ^d Teste indicado por BARROS *et al.* (1996): ^{d1} $F_{0.05; 4; 617} = 2,37$, ^{d2} $F_{0.05; 5; 617} = 2,21$, ^{d3} $F_{0.05; 10; 609} = 1,83$ e ^{d4} $F_{0.05; 13; 609} \cong 1,70$

Analisando os resultados obtidos para o teste F, tem-se que tanto para os modelos estacionários, como para os dinâmicos, não existe ganho em se utilizar modelos MLR – 8 ou MLR – 16, respectivamente, apesar dos resíduos deste último passarem no teste de normalidade (ver Figura 5.11). Analisando os gráficos de probabilidade dos resíduos dos modelos MLR – 4 e MLR – 6, Figuras 5.11 (a) e (b), observa-se que o modelo dinâmico apresenta-se melhor até mesmo para os valores extremos. Vale ressaltar que, quando um intervalo de confiança de 99% é considerado, os resíduos do modelo MLR – 6 passam no teste de normalidade, assim como, o valor do seu *F* calculado passa a ser menor que o *F* tabelado ($F_{0,01; 10; 609} = 2,32$).

Analisando os resultados obtidos para o teste F para os modelos PLS, tem-se que tanto para os modelos estacionários, como para os dinâmicos, não existe ganho em se utilizar modelos que contém todas VLs, ou seja, PLSR – 8 ou PLSR – 16, respectivamente. Analisando os gráficos de probabilidade dos resíduos dos modelos PLSR – 3, estacionário e dinâmico, Figuras 5.11 (c) e (d), respectivamente, observa-se mais uma vez que o modelo dinâmico apresenta-se melhor até mesmo para os valores extremos.



Figura 5.11: Teste de normalidade dos resíduos dos modelos para DBO de saída – Dados 1; (a) MLR – 4 estacionário; (b) MLR – 6 dinâmico; (c) PLS – 3 estacionário; (d) PLS – 3 dinâmico.

Com a inclusão da quarta VL, os modelos estacionário e dinâmico passariam a apresentar um R^2 ajustado de 31,5% e 37,5%, respectivamente, sendo seus respectivos p-valor iguais a 0,060 e 0,051. Assim, concordando com o critério utilizado para redução de dimensionalidade do sistema, nenhuma melhora significativa na performance de predição dos modelos PLS é obtida com a inclusão de mais VLs.

Os resultados obtidos para os conjuntos de validação e teste são apresentados na Tabela 5.10.

		Val	idação		T	este		Média ± De	esvio Padrão
	^a Modelo	R^2	^b MSE	R^2	^b MSE	R^2	^b MSE	$\overline{R}^2 \pm \sigma$	$b \overline{MSE} \pm \sigma$
	MLR – 8	37,8	20,4	28,4	26,2	38,7	25,8	34,9±5,7	$24,1 \pm 3,2$
Modelagem	MLR - 4	39,0	20,0	28,5	26,2	38,9	25,5	$35,5 \pm 6,0$	$23,9 \pm 3,4$
estacionária	PLSR-8	37,8	20,4	28,4	26,2	38,7	25,8	$34,9 \pm 5,7$	$24,1 \pm 3,2$
	PLSR - 3	36,6	20,8	26,8	26,8	37,7	26,4	$33,7 \pm 6,0$	$24,6 \pm 3,4$
	MLR – 16	39,6	20,0	30,8	26,7	46,4	22,8	39,0 ± 7,8	$23,2 \pm 3,4$
Modelagem	MLR-6	40,6	19,8	31,4	26,5	46,7	22,3	39,6±7,7	$22,9 \pm 3,4$
dinâmica	PLSR - 16	40,6	19,8	31,4	26,5	46,7	22,3	39,6 ± 7,7	$22,9 \pm 3,4$
	PLSR-3	39,9	19,8	28,4	28,1	44,4	23,7	37,6 ± 8,3	$23,9 \pm 4,2$

Tabela 5.10: Resultados dos modelos MLR e PLSR obtidos para os conjuntos de validação e teste – DBO de saída (Dados 1).

^a Modelo construído e número de variáveis independentes. ^bOs resultados MSE devem ser multiplicados por 10⁻⁴.

A análise da performance dos modelos através dos conjuntos de validação e teste é dificultada devido aos altos valores de desvios padrão de R^2 e *MSE* encontrados. Entretanto, é possível observar que, concordando com os resultados de modelagem, os modelos dinâmicos apresentam performance de predição superior ao estacionário para os dados de validação e teste.

O modelo MLR com seis variáveis preditoras apresenta resultados levemente melhores e é o modelo dinâmico mais simples obtido. Deve-se notar que, apesar do modelo PLS apresentar apenas três VLs como variáveis preditoras, a estimativa dos seus *scores* requer informações das oito variáveis originais, o que pode representar um modelo com maior custo de implementação.

Como esperado, através do gráfico do ajuste da variável predita versus a medida (Figura 5.12) verifica-se que é possível obter resultados de predição da DBO de saída dentro de um intervalo de predição de 95% para a grande maioria das amostras utilizando o modelo dinâmico MLR com seis variáveis preditoras. Isto também pode ser observado na Figura 5.13, onde é apresentada uma representação gráfica dos valores da DBO de saída medidos e preditos para este modelo. Mais de 85% das amostras apresentam desvios relativos entre a DBO medida e a predita menor que 10%.



Figura 5.12: Relação entre a DBO saída predita vs. medida (linha sólida) – Dados 1. As linhas tracejadas superior e inferior indicam o intervalo de predição de 95% para o modelo MLR – 6.

Modelos PCR foram também construídos. Primeiramente o método stepwise foi utilizado para eliminação das componentes menos correlacionadas com a variável dependente. Para $\alpha_{in} = 0,10$ e $\alpha_{out} = 0,15$ os resultados indicam que nenhuma componente dever ser excluída do modelo estacionário, e dez componentes devem ser utilizadas para o modelo dinâmico.

Quando níveis de confiança mais exigentes são utilizados ($\alpha_{in} = 0,15$ e e $\alpha_{out} = 0,15$), obtém-se que apenas a componente cinco deve ser excluída do modelo estacionário, sendo obtido um modelo com R^2 ajustado praticamente igual ao obtido pelo método MLR, 31,2%. A performance dos modelos PCR é ainda pior, R^2 ajustado igual a 17,2%, quando apenas as quatro primeiras componentes principais são utilizadas na construção do modelo. Para o modelo dinâmico, os resultados *stepwise* discordam com os obtidos segundo Todeschine, ou seja, tem-se que as seis primeiras componentes devem ser utilizadas, sendo obtido R^2 ajustado igual a 18,36%. Quando apenas as cinco primeiras CPs são utilizadas, obtém-se R^2 ajustado igual a 16,6%. Tais resultados já eram esperados tendo em vista que as informações das variáveis encontram-se espalhadas praticamente em todas as componentes, como pode ser observado através dos seus *loadings*.



(b)

Figura 5.13: Gráficos de séries temporais – Dados 1; (a) DBO de saída medida e predita para o modelo dinâmico MLR - 6; (b) Resíduos – Linhas superior e inferior indicam um intervalo de confiança de 95%.

Comparando-se os resultados obtidos por técnicas estatísticas multivariadas, ou seja, regressão linear múltipla, regressão por componentes principais e regressão por mínimos quadrados parciais, observa-se que o modelo MLR dinâmico, contendo apenas seis variáveis preditoras, é o que apresenta melhor performance de predição da DBO de saída.

Modelos FLN e PLS-FLN foram inicialmente construídos, sendo obtidos os mesmos números de monômios que os da DBO de entrada quando realizadas expansões polinomiais de grau um e dois.

Nas Tabelas 5.11 e 5.12 são apresentados os resultados de modelagem quando utilizados graus de monômios iguais a um e dois, respectivamente. Os resultados de validação e teste são apresentados na Tabela 5.13 (modelos FLNs de primeira ordem) e Tabela 5.14 (modelos FLNs segunda ordem). Das três funções de transferência testadas, a função logsig apresentou melhores resultados considerando os valores de R^2 e *MSE* obtidos.

Tabela 5.11: Resultados FLN (ordem 1) obtidos para o conjunto de modelagem – DBO de saída (Dados 1).

	^a Modelo	R^2	^b MSE	° p-valor
	FLN – 8	32,8	25,3	0,008
Modelagem Estacionária	PLS-FLN-8	32,8	25,3	0,008
PStacional la	PLS-FLN – 3	32,5	25,2	0,005
	FLN – 16	38,6	22,8	0,062
Modelagem Dinâmica	PLS-FLN - 16	38,9	22,8	0,062
**	PLS-FLN-3	38,5	22,9	0,041

^a Modelo construído e número de variáveis independentes. ^b Os resultados MSE devem ser multiplicados por 10⁻⁴. ^c P-valor para o teste de normalidade dos resíduos.

	* Modelo	R^2	^b MSE	°p-valor
	FLN – 8	43,9	20,9	0,074
Modelo	PLS-FLN-8	43,9	20,9	0,074
estacionario	PLS-FLN – 3	34,3	24,5	0,007
N (FLN – 16	60,0	14,9	0,109
Modelo	PLS-FLN – 16	60,0	14,9	0,109
	PLS-FLN-3	40,4	22,3	0,054

Tabela 5.12: Resultados FLN (ordem 2) obtidos para o conjunto de modelagem – DBO de saída (Dados 1).

^a Modelo construído e número de variáveis independentes. ^b Os resultados MSE devem ser multiplicados por 10⁻⁴.

° P-valor para o teste de normalidade dos resíduos.

		Vali	dação		Te	ste		Média ± De	esvio Padrão
	^a Modelo	R^2	^b MSE	R^2	^b MSE	R^2	^b MSE	$\overline{R}^2 \pm \sigma$	$b\overline{MSE}\pm\sigma$
	FLN-8	37,6	20,6	27,4	26,1	38,8	25,2	34,6 ± 6,3	$24,0 \pm 3,0$
Modelo	PLS-FLN-8	42,2	21,7	30,1	26,9	39,7	33,2	37,3 ± 6,4	27,3 ± 5,8
estacionario	PLS-FLN-3	41,3	23,1	27,9	28,0	38,7	34,8	36,0 ± 7,1	2 8, 6±5,9
	FLN - 16	39,2	20,6	29,8	25,2	44,8	25,0	37,9 ± 7,6	23,6 ± 2,6
Modelo dinâmico	PLS-FLN-16	41,2	21,0	34,9	23,1	47,8	39,2	41,3 ± 6,5	27 ,8 ± 10,0
unameo	PLS-FLN-3	41,7	21,6	30,6	22,6	41,6	21,3	3 8, 0 ± 6,4	21,8±0,7

Tabela 5.13: Resultados FLN (ordem 1) obtidos para os conjuntos de validação e teste – DBO de saída (Dados 1).

^a Modelo construído e número de variáveis independentes. ^bOs resultados MSE devem ser multiplicados por 10⁻⁴.

Tabela 5.14: Resultados FLN (ordem 2) obtidos para os conjuntos de validação e teste – DBO de saída (Dados 1).

		Vali	idação		Te	ste		Média ± De	svio Padrão
	^a Modelo	R^2	^b MSE	R^2	^b MSE	R^2	^b MSE	$\overline{R}^2 \pm \sigma$	$b \overline{MSE} \pm \sigma$
	FLN – 8	36,3	26,0	28,2	36,1	36,0	32,9	33,5 ± 4,6	31,7 ± 5,2
Modelo estacionário	PLS-FLN – 8	45,3	24,7	30,4	46,7	41,2	32,2	39,0 ± 7,7	34,5 ± 11,2
	PLS-FLN-3	47,1	22,1	34,3	34,2	40,2	40,4	$40,5 \pm 6,4$	32,2 ± 9,3
	FLN - 16	22,2	67,8	10,4	47,6	5,5	87,2	12,7 ± 8,6	67,5 ± 19,8
Modelo dinâmico	PLS-FLN-16	18,6	64,8	21,1	55,0	14,5	118,0	18,1 ± 3,3	79,3 ± 33,9
	PLS-FLN-3	47,2	21,4	34,8	21,2	47,2	21,2	43,1±7,2	21,3 ± 0,1

^a Modelo construído e número de variáveis independentes. ^bOs resultados MSE devem ser multiplicados por 10⁻⁴.

Assim como ocorreu para a DBO de entrada, a perda de generalização das redes FLN com a inclusão de monômios de ordem dois pode ser notada claramente, tendo em vista que para o conjunto de modelagem tais modelos pareciam ser os melhores, discordando dos resultados obtidos para os conjuntos de validação e teste. Os resultados de validação e teste dos modelos FLN indicam que a redução da dimensionalidade das entradas da rede pela exclusão das últimas VLs realmente melhora a acurácia dos modelos de FLN de segunda ordem; o mesmo comportamento não é observado para o conjunto de modelagem. Para os modelos FLN de primeira ordem, apenas uma melhora modesta é observada para a mesma situação. Mais uma vez, a identificação do melhor modelo FLN de predição da DBO é dificultada pelos valores altos de desvios encontrados. Modelos de redes neurais MLP foram então construídos utilizando o algoritmo de treinamento DBD e a função de transferência sigmoidal. Diferentes topologias das redes foram testadas e os melhores resultados MLP obtidos são apresentados nas Tabelas 5.15. Na Tabela 5.16 são apresentados os resultados para os conjuntos de validação e teste.

	* Modelo	R^2	^b MSE	°p-valor
	MLP - 8 (n=8)	40,4	22,3	0,017
Modelagem Estacionária	PLS-MLP - 8 (n=8)	35,9	25,6	0,237
Estacionaria	PLS-MLP - 3 (n=6)	32,4	47,2	0,001
	MLP – 16 (n=3)	46,8	20,1	0,070
Modelagem Dinâmica	PLS-MLP – 16 (n=6)	41,6	21,9	0,007
"** *****************	PLS-MLP - 3 (n=6)	42,5	43,4	0,008

Tabela 5.15: Resultados MLP obtidos para o conjunto de modelagem - DBO de saída (Dados 1).

^a Modelo construído e número de variáveis independentes. ^b Os resultados MSE devem ser multiplicados por 10⁴. ^c P-valor para o teste de normalidade dos resíduos. n = número de neurônios intermediários.

		Vali	dação		Te	ste		Média ± Desvio Padrão		
	^a Modelo	R^2	^b MSE	R^2	^b MSE	R^2	^b MSE	$\overline{R}^2 \pm \sigma$	$b \overline{MSE} \pm \sigma$	
Modelagem	MLP - 8 (n=8)	49,8	16,6	30,7	24,5	46,1	23,6	42,2 ± 10,1	21,6±4,3	
	PLS-MLP - 8 (n=8)	37,8	22,0	25,1	32,5	33,1	32,8	32,0 ± 6,4	$29,1 \pm 6,2$	
cstacionai ia	PLS-MLP - 3 (n=6)	39,7	43,3	29,8	59,1	35,6	57,3	35,0 ± 5,0	53,2 ± 8,6	
	MLP - 16 (n=3)	56,7	14,4	35,6	23,7	47,9	23,1	46,7 ± 10,6	20,4 ± 5,2	
Modelagem dinâmica	PLS-MLP - 16 (n=6)	49,4	17,5	32,7	27,4	31,0	31,0	37,7 ± 10,2	$25,3\pm7,0$	
	PLS-MLP - 3 (n=6)	43,4	44,1	35,2	35,2	49,2	33,1	$42,6\pm7,0$	37,5 ± 5,8	

Tabela 5.16: Resultados MLP obtidos para os conjuntos de validação e teste - DBO de saída (Dados 1).

^a Modelo construído e número de variáveis independentes. ^bOs resultados MSE devem ser multiplicados por 10⁻⁴. n = número de neurônios intermediários.

Assim como ocorreu para o caso das redes FLN de primeira ordem, observa-se que a técnica PLS não auxilia o mapeamento não linear das redes MLP para o caso em estudo. Os mesmos fatores apresentados para a DBO de entrada podem explicar a ocorrência deste fato para a DBO de saída.

Note que apesar do conjunto de validação "influenciar" na definição da topologia das redes MLP, diferenças significativas entre os conjuntos de validação e teste são observadas apenas nos modelos MLP para predição da DBO de saída. Considerando apenas os resultados obtidos para o conjunto de teste, verifica-se que os modelos dinâmicos MLP com três neurônios e PLS-FLN com três VLs, são os melhores modelos obtidos. Entretanto, os resultados de validação indicam que o primeiro apresenta melhor performance de predição; os resíduos deste modelo apresentam-se normalmente distribuídos.

Como esperado, o gráfico de ajuste da DBO de saída predita versus a medida (Figura 5.14) mostra que o modelo MLP dinâmico com três neurônios é realmente capaz de fornecer resultados de predição dentro de um intervalo de predição de 95%.



Figura 5.14: Relação entre a DBO de saída predita versus medida (linha sólida) – Dados 1. As linhas tracejadas superior e inferior indicam o intervalo de predição de 95% para o modelo dinâmico MLP com três neurônios.

Na Figura 5.14 é apresentada uma representação gráfica dos valores da DBO de saída medidos e preditos para este modelo. Mais de 90% das amostras apresentam desvios relativos entre a DBO de saída medida e a predita menor que 10%.



Figura 5.15: Gráficos de séries temporais – Dados 1; (a) DBO de saída medida e predita para o modelo dinâmico MLP com três neurônios; (b) Resíduos – Linhas superior e inferior indicam um intervalo de confiança de 95%.

Realizando uma análise geral dos resultados, é possível observar que, independente da técnica de modelagem utilizada, os melhores modelos de predição da DBO de saída são obtidos quando a dinâmica do processo é considerada.

5.2.3 DADOS 2 - Predição da DBO de entrada e saída

A análise dos resultados de modelagem obtidos utilizando o conjunto Dados 2 é realizada de forma sucinta, dando-se enfoque à comparação da performance destes modelos com a dos modelos construídos com o conjunto Dados 1. Entretanto, vale ressaltar, que os mesmos cuidados de desenvolvimento e validação dos modelos foram tomados, seja utilizando-se o conjunto Dados 1 como o conjunto Dados 2.

Modelos de predição MLR foram inicialmente construídos. O método stepwise foi então utilizado na construção dos modelos lineares, sendo considerados $\alpha = 0,10$ (nível de 90% de confiança) e $\alpha = 0,15$ (nível de 85% de confiança), para os testes de inclusão e eliminação de variáveis.

Na predição da DBO de entrada, os resultados obtidos pelo método stepwise indicam que a DQO e nitrogênio amoniacal devem ser utilizados para a construção do modelo MLR. Já para a DBO de saída, além da DQO, os sólidos suspensos, temperatura, condutividade, produção de papel e celulose devem ser utilizados como variáveis preditoras. Tais resultados diferem dos obtidos pelo conjunto Dados 1. Os resultados de ambos modelos são apresentados na Tabela 5.17.

Modelos PLS são então construídos. Na Tabela A3 do Anexo 2 são apresentadas as variância explicada e acumulada das componentes obtidas, para predição da DBO de entrada e saída. Comparando-se as variâncias apresentadas pelas CPs e VLs, tanto para da DBO de entrada como para a de saída, observa-se mais uma vez que as CPs descrevem uma quantidade maior de variação que pode explicar o conjunto de variáveis dependentes (X) que as VLs. Por outro lado, as VLs descrevem mais variância da variável dependente (DBO) que as CPs.

Considerando então um limiar de pelo menos 2% na variância acumulada com a inclusão de uma VL no modelo, tem-se que as quatro primeiras VLs devem permanecer no modelo PLS de predição da DBO de entrada e saída. A inclusão da quinta VL ao modelo representa apenas um acréscimo de 0,64% e 0,18% na variância acumulada para o modelo para a DBO de entrada e saída, respectivamente.

Nas Tabelas A9 e A12 do Anexo 2 são apresentados os *loadings* das VLs, para a DBO de entrada e saída, respectivamente. Para os modelos de predição da DBO de entrada e saída tem-se que a primeira variável latente, que carrega 55,8% e 48,8% da variação total dos dados,

respectivamente, e representa 14,1 % e 12,8% da variação da variável dependente, respectivamente, traz informação principalmente da DQO. Na Tabela 5.17 e 5.18 são apresentados os resultados de modelagem e validação, respectivamente.

Tabela 5.17: Resultados MLR e PLSR obtidos para o conjunto de modelagem – DBO de entrada e saída (Dados 2).

	DBO de	entrada		DBO de saída								
^a Modelo	R^2 adj. (R^2)	^b MSE	p-valor	^{d}F	^a Modelo	$R^2 adj.(R^2)$	^b MSE	p-valor	^d F			
MLR - 12	62,1 (66,4)	8,4	0,593	dl1 20	MLR - 12	54,9 (63,5)	12,7	0,885	din or			
MLR-2	62,1 (63,3)	10,6	0,082	1,30	MLR-6	55,6 (59,9)	13,9	0,490	- 0,80			
PLSR - 12	62,1 (66,4)	8,4	0,593	d21 05	PLSR - 12	54,9 (63,5)	12,7	0,885	d2 0 70			
PLSR-4	63,9 (66,1)	8,5	0,596	1,25	PLSR-4	60,2 (62,8)	9,5	0,790	0,70			

^a Modelo construído e número de variáveis independentes. ^b Os resultados MSE devem ser multiplicados por 10^4 . ^c FLN de primeira ordem e ^c FLN de segunda ordem. ^d Teste indicado por Barros et al. (1996): ^{d1}F_{0,05; 10; 51} \cong 1,95, ^{d2}F_{0,05; 8; 51} \cong 2,10 e ^{d3}F_{0,05; 6; 51} \cong 2,30.

Tabela 5.18: Resultados MLR e PLSR obtidos para o conjunto de validação – DBO de entrada e saída (Dados 2).

		DBO de	e entrada		DBO	de saída	
		* Modelo	R^2	^b MSE	^a Modelo	R^2	^b MSE
		MLR – 12	57,4	13,8	MLR – 12	67,6	11,8
	Dadas 7	MLR - 2	71,8	8,6	MLR – 6	61,8	13,5
	Dados 2	PLSR - 12	57,4	13,8	PLSR - 12	67,6	11,8
Modelagem estacionária		PLSR-4	64,8	8,0	PLSR-4	65,4	10,9
	Dados 1	MLR – 8	70,3	10,8	MLR – 8	73,9	5,7
		MLR-5	66,7	11,7	MLR - 4	67,4	7,4
		PLSR – 8	70,3	10,8	PLSR – 8	73,9	5,7
		PLSR – 3	69,1	10,5	PLSR – 3	77,1	4,7
		MLR – 16	66,2	13,4	MLR – 16	61,0	9,6
Modelagem dinâmica	Dados 1	MLR-6	71,3	10,0	MLR-6	51,9	12,4
		PLSR – 16	66,2	13,4	PLSR - 16	61,0	9,6
		PLSR-3	63,9	11,7	PLSR – 3	66,3	7,7

^a Modelo construído e número de variáveis independentes. ^b Os resultados MSE devem ser multiplicados por 10⁻⁴. Os melhores modelos obtidos através do conjunto Dados 1 são apresentados em negrito.

O teste $F e R^2$ ajustado indicam que não existe melhora significativa em trabalhar-se com os modelos construídos com as doze variáveis latentes nem com 12 VLs, tanto para a DBO de entrada como para a de saída. Tais resultados indicam que os modelos PLSR com quatro variáveis latentes são os melhores em ambos os casos. Vale lembrar que quando utilizado o conjunto Dados 1, o modelos MLR apresentaram performance de predição similar, ou até mesmo superior, aos modelos PLSR.

Os resultados de R^2 e *MSE* para o conjunto de validação são contraditórios. Enquanto os valores de MSE indicam que o modelo PLSR – 4 apresenta o melhor resultado de predição da DBO de entrada e saída, o R^2 indica o modelo MLR – 2 e o MLR – 12 são os melhores para a DBO de entrada e saída, respectivamente.

Comparando-se as performances dos modelos construídos com os conjuntos Dados 1 e 2, observa-se que apesar dos modelo estacionário MLR com cinco variáveis preditoras e o modelo dinâmico MLR com seis preditoras, para a DBO de entrada e saída, respectivamente, não apresentarem os melhores resultados de predição para o conjunto de validação com 15 amostras, os demais modelos construídos com 626 amostras (Dados 1) podem fornecer resultados de predição superiores aos obtidos utilizando o conjunto Dados 2. Com base num conjunto de validação tão reduzido, é difícil realizar uma análise conclusiva.

Modelos FLN foram então construídos. Utilizando expansão polinomial de ordem dois foram obtidos vinte e seis, vinte e oito e quatorze monômios para os modelo FLN, PLS-FLN (com 13 VLs) e PLS-FLN (com quatro VLs), respectivamente. Já para a DBO de saída foram obtidos trinta monômios para os modelos FLN e PLS-FLN com dezesseis *inputs* (variáveis originais e VLs, respectivamente) e quatorze monômios para o modelo PLS-FLN com três VLs.

Modelos de redes neurais MLP foram construídos utilizando o algoritmo de treinamento DBD e a função de transferência sigmoidal. Diferentes topologias das redes foram testadas.

Os resultados das melhores redes FLN e MLP são apresentados nas Tabelas 5.19 e 5.20, para os conjuntos de modelagem e validação, respectivamente.

DBO de	e entrad	a	**************************************	DBO de saída							
^a Modelo	R^2	^b MSE	p-valor	^a Modelo	R^2	^b MSE	p-valor				
°FLN 12	59,9	8,9	0,650	°FLN – 12	54,7	13,6	0,758				
°PLS-FLN-12	59,9	8,8	0,648	°PLS-FLN – 12	54,7	12,2	0,868				
°PLS-FLN-4	60,0	8,9	0,552	°PLS-FLN – 4	54,5	13,8	0,796				
^d FLN - 12	92,1	1,7	0,016	^d FLN – 12	90,0	4,6	0,173				
^d PLS-FLN - 12	92,3	1,6	0,007	^d PLS-FLN – 12	89,0	2,81	0,023				
^d PLS-FLN-4	69,8	6,4	0,075	^d PLS-FLN-4	59,5	12,1	0,667				
MLP - 12 (n=3)	73,7	6,7	0,574	MLP – 12 (n=3)	71,3	13,2	0,801				
PLS-MLP - 12 (n=3)	73,6	6,6	0,696	PLS-MLP - 12 (n=3)	42,8	19,8	0,487				
PLS-MLP -4 (n=2)	61,0	10,1	0,727	PLS-MLP -4 (n=6)	51,3	19,3	0,363				

Tabela 5.19: Resultados FLN e MLP obtidos para o conjunto de modelagem – DBO de entrada e saída (Dados 2).

^a Modelo construído e número de variáveis independentes (n, número de neurônios intermediários). ^b Os resultados MSE devem ser multiplicados por 10^4 . ^c FLN de primeira ordem e ^d FLN de segunda ordem. n = número de neurônios intermediários.

	DBO de er	ntrada	· · · · · · · · · · · · · · · · · · ·	DBO de saída					
	* Modelo	R^2	^b MSE	^a Modelo	R ²	^b MSE			
	°FLN – 12	48,7	15,2	°FLN – 12	57,7	18,3			
	°PLS-FLN-12	48,7	15,2	°PLS-FLN – 12	68,4	19,2			
	°PLS-FLN-4	48,2	14,6	°PLS-FLN-4	71,3	21,2			
Modelagem	^d FLN - 12	0,2	105,0	^d FLN – 12	7,2	26,9			
estacionária	^d PLS-FLN-12	16,2	62,8	^d PLS-FLN-12	9,9	87,2			
(Dados 2)	^d PLS-FLN-4	38,3	33,8	^d PLS-FLN-4	63,7	23,6			
	MLP – 12 (n=3)	66,0	10,5	MLP – 12 (n=3)	66,0	4,1			
	PLS-MLP $- 12$ (n=3)	62,5	11,6	PLS-MLP $- 12$ (n=3)	74,4	6,3			
	PLS-MLP -4 (n=2)	67,5	9,7	PLS-MLP -4 (n=6)	74,8	5,3			
Melhor modelo obtido a partir de Dados 1	MLP - 8 (n=10)	81,3	6,9	MLP - 16 (n=3)	69,7	6,8			

Tabela 5.20: Resultados FLN e MLP obtidos para o conjunto de validação para a DBO de entrada e saída (Dados 2).

^a Modelo construído e número de variáveis independentes (n, número de neurônios intermediários). ^b Os resultados MSE devem ser multiplicados por 10^4 . ^c FLN de primeira ordem e ^d FLN de segunda ordem. n = número de neurônios intermediários.

Para a DBO de saída, a performance dos modelos FLN de primeira ordem e MLP é substancialmente melhorada utilizando a técnica PLS para reduzir o espaço das variáveis preditoras. Por outro lado, os modelos FLN e MLP apresentaram uma melhora modesta na predição da DBO de entrada na mesma situação. Para os modelos FLN de segunda ordem, verifica-se uma notável melhora dos resultados com a redução das variáveis de entrada pela técnica PLS, tanto para a DBO de entrada como para a de saída.

Os melhores resultados para o conjunto de validação são obtidos usando o modelo de estrutura PLS-MLP com dois e seis neurônios intermediários na predição da DBO de entrada e saída, respectivamente. Entretanto, para a DBO de entrada, tais resultados de predição são superados quando utilizado o modelo construído com o conjunto Dados 1. O fato do conjunto Dados 2 apresentar um maior número de variáveis preditoras e um número reduzido de amostras, justifica a necessidade em utilizar-se a técnica PLS no pré-processamento de dados. Vale lembrar que o mesmo não aconteceu para os modelos construídos utilizando oito variáveis preditoras e 626 amostras.

Tais resultados devem ser comparados com cautela tendo em vista que apenas dados históricos foram disponibilizados pela empresa para a construção dos modelos.

5.3 Considerações finais sobre o capítulo

O objetivo principal deste capítulo foi apresentar os resultados de predição da DBO de entrada e saída da lagoa aerada da IPB, utilizando-se técnicas estatísticas multivariadas de regressão e de redes neurais artificiais para a construção dos modelos. Os melhores modelos obtidos para a predição da DBO de entrada e saída são apresentados nas Tabelas 5.21 e 5.22.

Tabela 5.21: Resultados dos melhores modelos obtidos para os conjuntos de validação e teste – DBO de entrada (Dados 1).

		Validação			Teste			Média ± Desvio Padrão		
	Melhores Modelos MLR - 5 FLN - 8 MLP - 8 (n=10)	R^2	^b MSE	R^2	^b MSE	R^2	^b MSE	$\overline{R}^2 \pm \sigma$	$b \overline{MSE} \pm \sigma$	
Modelagem estacionária	MLR - 5	45,8	12,1	46,9	14,9	54,7	13,9	$49,2 \pm 4,9$	13,6 ± 1,4	
	FLN - 8	46,1	25,8	44,8	20,5	54,8	18,2	48,6 ± 5,4	21,5 ± 3,9	
	MLP-8 (n=10)	60,1	9,0	59,8	11,1	65,3	10,3	61,7 ± 3,1	$10,1\pm1,1$	

Tabela 5.22: Resultados dos melhores modelos obtidos para os conjuntos de validação e teste – DBO de saída (Dados 1).

		Vali	dação		Te	ste		Média ± Desvio Padrão		
	Melhores Modelos	R^2	^b MSE	R^2	^b MSE	R^2	^b MSE	$\overline{R}^2 \pm \sigma$	$b \overline{MSE} \pm \sigma$	
Modelagem	MLR – 6	40,6	19,8	31,4	26,5	46,7	22,3	39,6 ± 7,7	$22,9 \pm 3,4$	
	PLS-FLN – 3	47,2	21,4	34,8	21,2	47,2	21,2	43.1 ± 7.2	$21,3 \pm 0,1$	
umannea	MLP – 16 (3n)	56,7	14,4	35,6	23,7	47,9	23,1	46,7±10,6	$20,4 \pm 5,2$	

Com base nos resultados obtidos, os seguintes pontos devem ser evidenciados:

- De um modo geral, os resultados estatísticos indicam que os modelos estacionários são os que melhor se ajustam à predição da DBO de entrada, enquanto os modelos dinâmicos são os melhores para a DBO de saída.

- Os modelos MLP estacionário e dinâmico são os melhores obtidos para a predição da DBO de entrada e saída, respectivamente. No caso da DBO de entrada, além destes modelos apresentarem maiores R^2 , seus valores de *MSE* e de desvio são os menores obtidos.

- Invariavelmente, os resultados de predição da DBO de entrada são melhores que os de predição da DBO de saída. Isso já era esperado, tendo em vista que variáveis, tais como a taxa de crescimento microbiano e as rotas preferenciais na lagoa, apesar de influenciarem na variação dos valores da DBO de saída, não foram consideradas na construção dos seus modelos de predição.

- Para a DBO de entrada, a performance de predição do modelo MLR é consideravelmente melhorada quando utilizada a técnica *stepwise* para seleção de variáveis; este modelo apresenta resultados melhores que os modelos PLSR. Para a DBO de saída, nenhuma diferença significativa é observada entre os modelos MLR e PLSR.

- Os modelos lineares clássicos parecem ser mais robustos que os modelos FLN de primeira e segunda ordem, pois apresentam *MSE* com menor variabilidade.

 - É possível obter uma redução de dimensionalidade maior utilizando-se a técnica PLS do que a PCA, sendo que os modelos PCR apresentaram a pior performance de predição de todos os modelos construídos.

- A técnica PLS somente parece auxiliar na estimativa dos parâmetros das redes FLN de segunda ordem. Isto pode ser justificado pelo fato destas redes apresentarem um número considerável de monômios gerados.

- A redução de dimensionalidade pela técnica PLS mostrou ser eficaz para os modelos MLP e FLN, melhorando suas performances de predição quando apenas um número reduzido de amostras é disponibilizada para o desenvolvimento dos modelos (conjunto Dados 2). Comparando os resultados de regressão linear multivariada com os de redes neurais para esse conjunto de dados, pode-se observar que os modelos MLP e PLS-MLP apresentam performance de predição superior aos modelos MLR e PLS, respectivamente. Os modelos PLS-MLR fornecem os melhores resultados de predição da DBO de entrada e de saída. Os modelos FLN e PLS-FLN foram os piores obtidos para este caso.

- Os gráficos de controle da estatística T^2 de Hotelling se mostraram uma ferramenta de fácil interpretação e uso. Entretanto, as amostras indicadas como *outliers* pelos gráficos de controle T^2 não aparecem destoantes nos gráficos de ajuste e dos resíduos dos melhores modelos obtidos. Tais resultados indicam que os modelos conseguiram predizer tais *outliers* no mesmo intervalo de confiança que as demais amostras.

As estatísticas R^2 , R^2 ajustado, *MSE*, p-valor e teste F se mostraram parâmetros consideravelmente eficazes na análise comparativa entre os modelos MLR, PLS e PCR desenvolvidos. Já para os modelos de redes neurais, apenas as estatísticas R^2 , MSE e p-valor puderam ser utilizadas. Infelizmente, não existe nenhum critério universal que considere o número de parâmetros ajustados durante o treinamento para fins de comparação entre modelos com diferentes topologias.

110

CAPÍTULO 6 CONSIDERAÇÕES FINAIS

6.1 Conclusões

Ao escolher esse tema para desenvolver a tese de doutorado pretendia-se mostrar se a utilização das técnicas de redes neurais seria uma ferramenta viável, alternativa às técnicas clássicas de regressão linear e aos modelos teóricos (com base fenomenológica), na modelagem de um bioprocesso de complexidade reconhecida.

Foram escolhidos três técnicas estatísticas clássicas e duas de redes neurais para construção dos modelos para predição de um dos parâmetros de qualidade mais importantes, e de medição em laboratório mais demorada, da lagoa aerada, ou seja, a demanda bioquímica de oxigênio (DBO). Tais técnicas foram: a) Regressão linear multivariada – MLR; b) Regressão por mínimos quadrados parciais – PLSR; c) Regressão dos componentes principais – PCR; d) Redes neurais perceptron de múltiplas camadas – MLP; e) Redes neurais *functional-link* – FLN.

As principais conclusões sobre os temas abordados nesta tese foram apresentadas detalhadamente ao final de cada capítulo. De uma forma geral, apesar das redes MLP apresentarem uma etapa tediosa e demorada de definição da sua topologia, seus modelos apresentam resultados de predição superiores aos obtidos através das demais técnicas e são bastante satisfatórios tanto para predição da DBO de entrada como para a de saída. Surpreendentemente, os modelos MLR desenvolvidos com o auxílio do método *stepwise* para seleção de variáveis, apresentaram resultados similares ou até mesmo levemente superiores aos das FLN, mesmo sendo utilizada as variáveis latentes como *inputs*, ou funções de transferência reversíveis para auxiliar o seu mapeamento.

Com base nos resultados obtidos, é possível concluir-se que o número de dados e o seu grau de representabilidade do processo são fatores essenciais para se obter um modelo satisfatório com redes neurais. Quando um número reduzido de dados foi utilizado, o uso da técnica PLS para minimizar as entradas das redes se mostrou essencialmente necessário, principalmente para o caso mais complexo, de predição da DBO de saída. Modelos MLR e PLSR se mostraram mais eficientes na predição da DBO de entrada e saída, respectivamente, para este caso.

111

Mesmo não havendo exclusão de amostras indicadas como possíveis *outliers*, a importância na análise destas amostras, como etapa prévia à construção dos modelos, foi apresentada de forma clara, podendo ser utilizada em outros trabalhos. Neste trabalho, foi possível observar que os modelos construídos conseguiram realizar predições dos potenciais *outliers*, dentro do intervalo de confiança considerado.

A apresentação de uma revisão unificada, detalhada e comparativa das técnicas estatísticas multivariadas e de redes neurais artificias (área sombreada da Figura 6.1), aplicadas a um caso real, pode ser considerada o principal aporte desta tese. Considera-se também de grande relevância a apresentação da importância dos modelos de estrutura PLS-FLN e PLS-MLP (áreas mais escuras da Figura 6.1) na modelagem do sistema em estudo, principalmente quando apenas um número reduzido de amostras é utilizado para a construção dos modelos.



Figura 6.1: Desenho esquemático das ferramentas utilizadas na construção dos modelos preditivos.

Espera-se que este trabalho possa beneficiar estudantes, professores e pesquisadores das mais diferentes áreas, contribuindo para esclarecer as potencialidades e os problemas de utilizar técnicas clássicas de regressão e das redes neurais. Vale ressaltar que um ponto de grande preocupação sempre presente neste trabalho foi de alertar para que o uso das técnicas estatísticas clássicas e de redes neurais não seja feita de forma indiscriminada e descuidada, sendo apresentado diversos parâmetros estatísticos para auxiliar na escolha e na comparação entre modelos de diferentes dimensões e estruturas.

6.2 Perspectivas futuras

Como sugestões para prosseguimento deste trabalho sugere-se:

- a) Estudo e aplicação de modelos estatísticos empíricos não-lineares ao caso estudado.
- b) Construção de modelos de redes neurais onde são excluídas as variáveis de *input* consideradas como menos importantes por algum método de análise se sensitividade, tais como os propostos por GARSON (1991) ou SUNG (1998).
- c) Uso de algum algoritmo de eliminação de conexões (pesos) não importantes para os modelos neurais, tal como o *optimal brain surgeon* (HASSIB e STORK, 1993).
- d) Estudo e uso de uma estrutura de rede neural definida automaticamente por algum algoritmo de aprendizado construtivo ao caso estudado.
- e) Integração entre modelos de predição da DBO de entrada e de saída, ou seja, utilizar os valores preditos da DBO de entrada, como *input* da rede para predição da DBO de saída, tendo em vista a evidente correlação existente entre estes dois parâmetros.
- f) Construção de modelos de predição utilizando-se a DBO e DQO como variáveis de carga, ou seja, em unidades de concentração por tempo (mg/dia, por exemplo).

Para o processo estudado neste trabalho seria também interessante reestruturar a periodicidade das análises de forma a garantir um maior aproveitamento das mesmas e principalmente o uso das variáveis controladas do processo (carga de nitrogênio e fósforo) como variáveis preditoras.

CAPÍTULO 7

REFERÊNCIAS BIBLIOGRÁGICAS

- ADAMS, M.J. (1995). Chemometrics in analytical spectroscopy, Cambridge, UK.
- ALKULAIBI, A., SORAGHAN, J. J. (1997). Hybrid higher-order cepstrum and functional link networkbased blind equaliser (HOCFLN), *Signal Processing*, v. 62, 101-109.
- BAFFI, G., MARTIN, E. B. e MORRIS, A. J. (1999). Non-linear projection to latent structures revisited (the neural network PLS algorithm), Comp. Chem. Engng. 23, 1293-1307.
- BARROS, B. N., SCARMINIO, I. S. e BRUNS, R. E. (1996). Planejamento e otimização de experimentos, *Editora da UNICAMP*, Campinas, São Paulo, Brasil.
- BAKSI, B. R. e CATTERJEE, R. (1998). Unification of neural and statistical methods as applied to materials structure-property mapping, Journal of Alloys and Compounds 279, 39-46.
- BAKSI, B. R. e UTOJO, U. (1999). A common framework for the unification of neural, chemometric and statistical modeling methods, Analitica Chimica Acta. 384, 227-247.
- BAUGHMAN, D. R. e LIU, Y. A. (1995). Neural networks in bioprocessing and chemical engineering, Academic Press (1995).
- BEEBE, K. R., PELL, R. J. e SEASHOLTZ, M. B. (1998). Chemometrics A Pratical Guide, Wiley Press.
- BILLINGS, S. A., CHEN, S. e KORENBERG, M. J. (1989). Identification of MIMO non-linear systems using a forward-regression orthogonal estimator, *Int. J. Control*, v. 49, n. 6, 2157-2189.
- BISHOP, C.M. (1994). Neural networks and their applications, Rev. Sci. Instrum., v. 65, n. 6, 1803-1831 (1994).
- BEEBE K. R., PELL, R. J. e SEASHOLTZ, M. B. (1998). Chemometrics A Pratical Guide", Wiley Press.
- BLANK, T. B. e BROWN, S.D. (1993a). Data preprocessing using neural networks, Analytica. Chimica Acta, v. 277, 273-287.
- BLANK, T.B. e BROWN, S.D. (1993b). Nonlinear multivariate mapping of chemical data using feedforward neural networks", *Anal. Chem.*, v. 65, 3081-3089.
- BROWN e HARRIS (1994). Neurofuzzy Adaptative Modelling and Control, Prentice Hall, New York.
- CANCILLA, D.A e FAND, X. (1996). Evaluation and quality control of environment analytical data from the Niagara River using multiple chemometric method", *Journal Great Lakes Res.*, v. 22, 241-253.

- CHEN, S. e BILLINGS, S. A. (1992). Neural networks for nonlinear dynamic system modelling and identification, *Int. J. Control*, v. 56, n. 2, 319-346.
- CHEN G., MCAVOY, T. J. e PIOVOSO, M. J. (1998). A multivariate statistical controller for on-line quality improvement", Journal of Process Control, v. 8, n. 2, 139-149.
- COLACIOPPO, R. (2001). Controle estatístico multivariado de processos para observações individuais, Tese de Mestrado, Instituto de Matemática, Estatística e Ciência da Computação/UNICAMP, Campinas – SP.
- COSTA, A. C., DECHECHI, E. C., SILVA, F. L. H., MAUGERI FILHO, F. e MACIEL FILHO, R. (2000). Simulated Dynamics and Control of an Extractive Alcoholic Fermentation, *Applied Biochemistry and Biotechnology*, v. 84-6, SPR 577-593.
- COSTA, A. C., HENRIQUES, A. S. W., ALVES, T. L. M., MACIEL FILHO, R. e LIMA, E. L. (1999). A hybrid neural model for optimization of fed-batch fermentations, *Brazilian Journal of Chemical Engineering*, v. 16, n. 1, 53-63.
- COSTA, A. C., ALVES, T. L. M., HENRIQUES, A. W. S., MACIEL FILHO, R. e LIMA, E. L. (1998). An adaptative optimal control scheme based on hybrid neural modelling, *Computers chem. Engng.*, v. 22, suppl., S859-S862.
- COTE, M., GRANDIJEAN, B. P. A., LESSARD, P. e YHIBAULT, J. (1995). Dynamic modeling of the activated sludge process: improving prediction using neural networks, Wat. Res. 29, 995-1004.
- DRAPER, N. R. e SMITH, H. (1981). Applied Regression Analysis, Wiley Press, New York,.
- DESPAGNE, F. e MASSART, D. L. (1998). Neural networks in multivariate calibration, *Analyst*, 123, 157-178.
- DUTTA, D., MOHANTY, A. K., CHOUDHURY, R.K. e CHAND, P. (1998). Pattern recognition of particles tracks using principal component analysis and artificial neural network, *Nuclear Instruments & Methods in Physics Research*, v. 404, 445-454.
- FELDMAN, J.A e BALLARD, D. H. (1982). Connectionist models and their properties, *Cognitive Sci.*, v. 6, 205-254.
- GARSON, G.D. (1991). Interpreting Neural-Network Connection Weights. AI Expert. Apr, 47-51.
- GELADI, P. e KOWALSKI, B. R. (1986). Partial least-squares regression: A tutorial, Analytica Chimica Acta, v. 185, 1-17.
- GONTARSKI C.A., RODRIGUES, P. R., MORI, M. e PRENEM, L.F. (2000a). Simulation of an industrial wastewater treatment plant using artificial neural networks, *Computers and Chemical Engineering*, v. 24, 1719-1723.
- GONTARSKI C.A. (2000b). Avaliação da utilização de redes neuronais aplicadas a processos químicos, Tese – Doutorado, Faculdade de Engenharia Química/UNICAMP, Campinas - SP.

- GONZÁLEZ, A.G. e GONZÁLEZ-ARJONA, D. (1995). Statistical assessment of a new criterion for selecting the number of factors in factor analysis, *Analytica. Chimica Acta*, v. 314, 251-252.
- HÄCK, M. e KÖHNE, M. (1996). Estimation of wastewater process parameters using neural networks, Water Sci. Technol., v. 33, n. 1, 101-107.
- HAMODA, M. F., AL-GHUSAIN, I. A. e HASSAN, A. H. (1999). Integrated wastewater treatment plant performance evaluation using artificial neural networks, *Water. Sci. Tech.*, 40 (7), 55.
- HARADA, L. H. P. (2001). Modelagem Híbrido Neuronal Aplicada a Processos Fermentativos, Tese de Mestrado, Faculdade de Engenharia Química/UNICAMP, Campinas SP.
- HARADA, L. H. P., COSTA, A. C. e MACIEL FILHO, R. (2002), aceito para publicação na Appl. Biochem. Biotechnol.
- HARREMOËS, P., CAPODAGLIO, A. G., HELLSTROM, B. G., HENZE, M., JENSEN, K. N., LYNGGAARD-JENSEN, A., OTTERPOHL, R. e SOEBORG, H. (1993). Wastewater treatment plants under transient loading-performance, modeling and control, Wat. Sci. Tech. 27, 71-115.
- HAYKIN, S. (1994). Neural Networks, Macmillan College Publishing Company, New York.
- HENRIQUE, H. M. (1999). Tese de Doutorado, Programa de Engenharia Química/COPPE/UFRJ, Rio de Janeiro RJ.
- HENRIQUE, H. M. e LIMA, E. L. (1996). Model Structure Determination in Neural Network Models, submetido.
- HENRIQUES, A. W. S., COSTA, A. C., ALVES, T. L. M. e LIMA, E. L. (1999). Optimization of fedbatch processes: challenges and solutions, *Brazilian Journal of Chemical Engineering*, v. 16, n. 2, 171-177.
- HOLCOM, T.R. e MORARI, M. (1992). PLS/ Neural Networks, Computers Chem. Eng., v. 16, n. 4, 393-411.
- HONG-GUANG, N. e JI-ZONG, W. (2000). Prediction of compreensive strength of concrete by neural networks, *Cement and Concrete Research*, v. 30, 1245-1250.
- HOLPFIELD, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities, *in Proc. Nat. Academy Sci.*, v. 79, 2554-2558.
- JACKSON, J. E. (1980). Principal components and factor analysis: part I principal components, *Journal* of *Quality Technology*, 12 (4), 201-213.
- JACKSON, J. E. (1985). Multivariate Quality Control. Communications in Statistics Theory and Methods, v. 14, n. 11, 2657-2688.
- JACKSON, J. E. e MUDHOLKAR, G. S. (1979). Control procedures for residuals associated with principal component analysis, *Technometrics*, v. 21, n. 3, 341-349.
- JACKSON, J.E. (1991). A user guide to principal components, Wiley Press, New York.

- JACOBS, R.A. (1998).Increased Rates of Convergence Through Learning Rate Adaptation, Neural Networks, v.1, p. 295-307.
- JOHNSON, M.E. (1987). Multivariate Statistical Distribution, Wiley Press, New York.
- JOHNSON, R. A. e WICHERN, D. W. (1998). Applied multivariate statistical analysis, Prentice Hall, Upper Saddle River, New Jersey.
- KANJILAL, P.P. (1995). On the application of orthogonal transformation for design and analysis of feedforward networks, *IEEE Trans. Neural Networks*, v. 6, 1061-1070.
- KANO, M., NAGAO, K., HASEBE, S., HASHIMOTO, I., OHNO, H., STRAUSS, R. e BAKSHI, B. (2000). Comparison of statistical process monitoring methods:application to the Eastman challenge problem, *Computers and Chem. Eng.*, v. 24, 175-181.
- KENNARD, R. W. e STONE, L. A. (1969). Computer aided design of experiments, Technometrics, v. 11, n. 1, 137.
- KOHONEN, T. (1984). Self-organization and associative memory, Springer-Velarg, Berlin.
- KOMPANY-ZARED, M., MASSOUMI, A. e PEZESHK-ZADEH, S.H. (1999). Simultaneous spectrophotometric determination of Fe and Ni with xylenol orange using principal component analysis and artificial neural networks in some industrial samples", *Talanta*, 48, 283-292.
- KOURTI, T. e MACGREGOR, J. F. (1996). Multivariate SPC methods for process and product monitoring, *Journal of Quality Technology*, v. 28, n. 4, 409-428.
- KOVÁCS, Z.L. (1996). Redes Neurais Artificiais: Fundamentos e Aplicações, São Paulo: Edição Acadêmica.
- LEE, D. S. E PARK, J. M. (1999). Neural network modeling for on-line estimation of nutrient dynamics in a sequentially-operated batch reactor, *Journal of Biotechnology*, 75, 229.
- LOZANO, J., NOVIC, M., RIUS, F. X. e ZUPAN, J. (1995). Modeling metabolic energy by neural networks, *Chemom. Intell. Lab. Syst.*, v. 28, n. 1, 67-72.
- MACLAIN, R. B. e HENSON, M. A. (2000). Principal component analysis for nonlinear model reference adaptative control, *Computers and Chem. Eng.*, v. 24, 99-110.
- MAIER, H. R. e DANDY, G. C. (2000). Neural networks for prediction and forecasting of water resources variables: a review of modelling issues and applications, *Environmental Modelling & Software*, v. 15, 101-124.
- MAJCEN, N., RAJER-KANDUC, K., NOVIC, M. e ZUPAN, J. (1995). Modeling of property prediction from multicomponent analytical data using different neural networks", *Anal. Chem.*, v. 67, 2154-2161.
- MALKAVAARA, P, HARJULA, P e KNUUTINEN, J. (2000). Chemometric investigation on structural changes in pine kraft lignin during pulping, *Chem. Intel. Lab. Systems*, v. 52, 117-122.
- MARDIA, K. V., KENT, J. T. e BIBBY, J. M. (1979). Multivariate Analysis, Academic Press.

MARTENS, H. e NAES, T. (1989). Multivariate Calibration, Wiley Press.

MATLAB PLS_Toolbox 2.0, Eigenvector Research Inc., Manson, WA (1998).

- MELEIRO, L. A. C. (2002). Projeto e aplicação de controladores baseados em modelos lineares, neurais e nebulosos. Tese de Doutorado, Faculdade de Engenharia Química, UNICAMP, Campinas SP Brasil
- METCALF & EDDY Inc. (1991). Wastewater engineering: treatment, disposal and reuse, 3rd edition, New York: McGraw-Hill.
- MILLER, G. F., TODD, P. M., HEDGE, S. U. e SHAFFER, J. D. (1989). Designing Neural Networks using Genetic Algorithms, *Proceedings of the Third International Conference on Genetic Algorithms*, Morgan Kaufmann Publishers, 379-384.
- MILTON, J. S. e ARNOLD, J. C. (1990). Introduction to Probability and Statistics, McGraw Hill. New York.
- MINSKY, M. e PAPERTS, S. (1969). Perceptrons: an introduction to computational geometry, MIT Press, Cambridge, Mass..
- MONTGOMERY, D. C. (2002). Applied Statistics and Probability for Engineers, John Wiley & Sons Press, New York.
- MONTGOMERY, D. C. e PECK, E. A (1992). Introduction to linear regression analysis, *Wiley Press*, New York.
- MONTGOMERY, D. C., RUNGER, G. C., HUBELE, N. F. (1998). Engineering Statistics, *Wiley Press*, New York.
- MORALES, M. M., MARTÍ, P., CAMPOS, L. e SAGRADO, S. (1999). An environmental study by factor analysis of surface seawaters in the Gulf of Valencia. *Anal. Chim. Acta*, v. 394, 109-117.
- MORRIS, A.J., MONTAGUE, G. A. e WILLS, M. J. (1994). Artificial neural networks: studies in process modeling and control", *Transactions of IChemE*, 72, A, 3-19.
- NASCIMENTO, C. A. O., GIUDICI, R. e GUARDANI, R. (2000). Neural network based approach for optimization for industrial chemical process, *Computers and chemical Engineering*, v. 24, 2303-2314.
- OLIVEIRA, K.P.S. (Fev/2000). Aplicação da técnica de Redes Neurais e de Análise de Componentes Principais na Modelagem de uma Lagoa Aerada da RIPASA S/A. Campinas, Faculdade de Engenharia Química/UNICAMP, Dissertação de Mestrado.
- OLIVEIRA-ESQUERRE, K. P., COSTA, A C., BRUNS, R. E. e MORI, M. (2003a). Simulation of an aerated lagoon using artificial neural networks and multivariate regression techniques, aceito para publicação em uma edição especial da *Applied Biochemistry and Biotechnology*.
- OLIVEIRA-ESQUERRE, K. P., MORI, M. e BRUNS, R. E. (2002). Simulation of an industrial wastewater treatment plant using artificial neural networks and principal components analysis, *Brazilian Journal of Chemical Engineering*, v. 19, n. 4, 365-370.

- OLIVEIRA-ESQUERRE, K. P., SEBORG, D. E., BRUNS, R. E. e MORI, M. (Abril/2003b). Application of Steady-State and Dynamic Modeling for the Prediction of BOD for an Aerated Lagoon at a Pulp and Paper Mill. Part I: Linear Approaches, submetido para publicação na *Biochemical Engineering Journal*.
- OLIVEIRA-ESQUERRE, K. P., SEBORG, D. E., BRUNS, R. E. e MORI, M. (Abril/2003c). Application of Steady-State and Dynamic Modeling for the Prediction of BOD for an Aerated Lagoon at a Pulp and Paper Mill. Part II: Nonlinear Approaches, submetido para publicação *na Biochemical Engineering Journal.*
- ORTIZ-ESTARELLES, O, MARTÍN-BIOSCA, Y, MEDINA-HERNÁNDEZ, M. J., SAGRADO, S. e BONET-DOMINGO, E. (2001). On the internal multivariate quality control of analytical laboratories. A case study: the quality of drinking water, *Chem. Intel. Lab. Systems*, v. 56, 93-103.
- PAO, Y.H. (1989). Adaptative Pattern Recognition and Neural Networks. *Reading. Mass: Addison-Wesley*.
- PIELOU, E. C. (1984). The interpretation of ecological data, Wiley Press, New York, USA.
- POGGIO, T. e GIROSI, F. (1990). Networks for approximation and learning, Proc. IEEE, 78 (9) 1481-1497.
- PONTON, J.W. e KLEMES, J. (1993). Alternatives to neural networks for inferential measurement, *Computers Chem. Eng.*, v. 17, n. 10, 991-1000 (1993).
- POPPI R. J. (Julho/1993). Construção de um espectrofotômetro com transformada de Hadamard e sua aplicação na análise por injeção em fluxo. Campinas, Instituto de Química/UNICAMP, Tese Doutorado.
- PU, H. e HUNG, Y. (1995). Use of artificial neural networks: Predicting trickling filter performance in a municipal wastewater treatment plant", *Envir. Manag. Health*, v. 6, n. 2, p. 16-27.
- QIN, S. J. e MCAVOY, T. J. (1992). Nonlinear PLS modeling using neural networks, *Computers and Chemical Engineering*, 16, 379-391.
- ROSENBLATT, R. (1962). Principles of Neurodynamics, Spartan Books, New York.
- RUMELHART, D. E. e McCLELLAND, J.L. (1986). Parallel distributed processing: explorations in the microstructure of cognition, v. 1, Foundations Mit Press.
- SIMOGLOU, A., MARTIN, E. B. e MORRIS, A. J. (2000). Multivariate statistical process control of an industrial fluidized-bed reactor", *Control Engineering Practice*, v. 8, 893-909.
- SUNG, A. H. (1998). Ranking importance of input parameters of neural networks", *Expert Systems with Applications*, v. 15, 405-411.

- STEYER, J. P., ROLLAND, D., BOUVIER, J. C. e MOLETTA, R. (1997). Hybrid fuzzy neural network for diagnosis – Application to the anaerobic treatment of wine distillery wastewater in a fluidized bed reactor, Wat. Sci. Tech. 36, 209-217.
- YANG, G.B.O. (1996). Managing secondary treatment systems, TAPPI Journal, v. 52-54.
- TODESCHINE, R. (1997). Data correlation, number of significant principal components and shape of molecules, The K correlation index, *Anal. Chim. Acta*, v. 348, 419-430.
- UTOJO, U. e BAKSHI, B. R. (1995). Neural Networks in Bioprocessing and Chemical Engineering, Academic Press Inc., Chapter: Connections Between Neural Networks and Multivariate Statistical Methods: An Overview.
- WALCZAK, B. e MASSART, D.L. (1996). The Radial Basis Functions Partial Least Squares approach as a flexible non-linear regression technique, *Anal. Chim.*, v. 331, 177-185.
- WILCOX, S. J., HAWKES, D. L., HAWKES, F. R. e GUWY, A. J. (1995). A neural network, based on bicarbonate monitoring, to control anaerobic digestion. *Wat. Res.*, v. 29, n. 6, 1465.
- ZHAO, H., HAO, O. I., FELLOW, A. S. C. E., MCAVOY, T. J. e CHANG, C. H. (1997). Modeling nutrient dynamics in sequencing batch reactor. *Journal of Environ. Eng.*, v. 123, n. 4, 863.

ZUPAN, Z. e GASTEIGER, J. (1993). Neural Networks for Chemists, New York.

CAPÍTULO 8

PUBLICAÇÕES ASSOCIADAS À TESE

8.1 Trabalhos publicados

- OLIVEIRA-ESQUERRE, K. P., SEBORG, D. E., BRUNS, R. E. e MORI, M. (Abril, 2003).
 Application of Steady-State and Dynamic Modeling for the Prediction of the BOD of an Aerated
 Lagoon at a Pulp and Paper Mill. Part I: Linear Approaches, submetido para publicação no
 Biochemical Engineering Journal.
- OLIVEIRA-ESQUERRE, K. P., SEBORG, D. E., MORI, M. e BRUNS, R. E. (Abril, 2003).
 Application of Steady-State and Dynamic Modeling for the Prediction of the BOD of an Aerated
 Lagoon at a Pulp and Paper Mill. Part II: Nonlinear Approaches, submetido para publicação no
 Biochemical Engineering Journal.
- OLIVEIRA-ESQUERRE, K. P., COSTA, A. C., BRUNS, R. E. e MORI, M. (2003). Simulation of an aerated lagoon using artificial neural networks and multivariate regression techniques, *Applied Biochemistry and Biotechnology*, v. 105-108, 437-449.
- OLIVEIRA-ESQUERRE, K. P., BRUNS, R. E. e MORI, M. (2002). Simulação de uma lagoa aerada utilizando técnicas estatísticas multivariadas. Anais do XIV Congresso Brasileiro de Engenharia Química, Natal – RN - Brasil.
- OLIVEIRA-ESQUERRE, K. P., BRUNS, R. E. e MORI, M. (Outubro, 2003). Multivariate Linear Modeling of Bioprocess Treatment for Industrial Wastewater: Improving Prediction Using the PLS Approach, submetido para publicação no *Brazilian Journal of Chemical Engineering*.
- OLIVEIRA-ESQUERRE, K. P.; COSTA, A. C.; BRUNS, R. E.; MORI, M. (2002). Simulation of an aerated lagoon using artificial neural networks and multivariate regression techniques, livro de resumos do 24th Symposium on Biotechnology for Fuels and Chemicals, Gatlinburg Tenessee EUA.
- ¹OLIVEIRA-ESQUERRE, K. P., MORI, M. e BRUNS, R. E. (2002). Simulation of an industrial wastewater treatment plant using artificial neural networks and principal components analysis, *Brazilian Journal of Chemical Engineering*, v. 19, n. 4, 365-370.

¹ Fruto do trabalho de OLIVEIRA (2000). Parte das análises estatísticas apresentadas neste artigo foi realizada durante o período de desenvolvimento deste trabalho de pesquisa de doutorado.

Vale ressaltar que, os trabalhos submetidos para publicação do *Biochemical Engeneering* Journal, em abril de 2003, são frutos da pesquisa realizada no Process Control Laboratory na University of California, Santa Barbara (UCSB), sob supervisão do professor Dale E. Seborg.

8.2 Trabalhos apresentados

Evento: American Institute of Chemical Engineering Annual Meeting (AIChE). Trabalho: Prediction of BOD in an industrial wastewater treatment plant using linear and nonlinar dynamic models. Autores: Karla P. Oliveira-Esquerre, ² Dale E. Seborg, Milton Mori e Roy E. Bruns. Período: 03 a 08 de novembro de 2002, Indianopolis – Indiana – EUA.

Evento: Congresso Brasileiro de Engenharia Química (XIV COBEQ). Trabalho: Simulação de uma lagoa aerada usando técnicas estatísticas multivariadas. Autores: ² Karla P. Oliveira-Esquerre, Roy E. Bruns e Milton Mori. Período: 24 a 27 de agosto de 2002, Natal – RN – Brasil.

Evento: 24th Symposium on Biotechnology for Fuels and Chemicals. Trabalho: Simulation of an aerated lagoon using artificial neural networks and multivariate regression techniques.

Autores: ² Karla P. Oliveira-Esquerre, Aline C. Costa, Roy E. Bruns e Milton Mori. Período: 28 de abril a 01 de maio de 2002, Gatlinburg – Tennessee – EUA.

8.3 Trabalho premiado

Prêmio: OUTSTANDING POSTER PRESENTATION (1° dentre 64 trabalhos).

Trabalho: Simulation of an Aerated Lagoon Using Artificial Neural Networks and Multivariate Regression Techniques.

Autores: ² Karla P. Oliveira-Esquerre, Aline C. Costa, Roy E. Bruns e Milton Mori.

Evento: 24th Symposium on Biotechnology for Fuels and Chemicals.

Período: 28 de abril a 01 de maio de 2002, Gatlinburg - Tennessee - EUA).

² Autor que apresentou o trabalho.

8.4 Considerações finais

Frente os trabalhos publicados, submetidos e apresentados torna-se evidente a relevância do trabalho de pesquisa desenvolvido no decorrer de três anos e dois meses. O trabalho premiado é fruto do reconhecimento da excelência de um dos trabalhos desenvolvidos.

Com relação às principais atividades extras realizadas paralelamente ao desenvolvimento desta pesquisa de doutorado, tem-se:

- a) Participação do Programa de Estágio Docente PED, na Faculdade de Engenharia Química da UNICAMP, relacionadas às seguintes disciplinas:
- EQ481 Introdução à Engenharia Química (agosto a dezembro de 2000);
- EQ052 Tópicos em Processos Químicos (fevereiro a junho de 2001);
- EQ921 Projetos Químicos (fevereiro a junho de 2003), e
- EQ041 Tópicos em Projeto Químico (fevereiro a junho de 2003).
- b) Aluna do curso de Especialização em Engenharia Ambiental, na Faculdade de Engenharia Química da UNICAMP (outubro de 1999 a dezembro de 2000).
- c) Aluna do curso de Melhoria de Processos Formação Black Belt, no Instituto de Matemática, Estatística e Computação Científica da UNICAMP (setembro de 2000 a março de 2001).



Anexo 1: Gráficos de séries temporais e matriz de correlação.

Figura 1A: Gráficos de séries temporais das variáveis preditoras e preditas da lagoa aerada da IPB (Set/1996 a Jul/2000), onde cada amostra representa o valor individual ou médio das análises realizadas em um dia.

	BODin	BODout	DQO	CHUVA	COND	COR	NAM	NN	pН	SST	Т	VAZ	CEL	PAP
n on	1	0,51	0,62	0,01	0,23	0,12	0,15	0,09	0,09	0,24	0,29	0,07	0,31	0,09
BOD ⁱⁿ	0	0	0	0,75	0	0	0	0,15	0	0	0	0,01	0	0
DOD	0,51	1	0,47	-0,05	0,2	0,15	0,07	0,18	0,07	0,09	0,19	0,2	0,24	0,09
BOD _{out}	0	0	0	0,14	0	0	0,1	0	0,01	0,03	0	0	0	0
DOO	0,627	0,47	1,00	-0,06	0,42	0,20	0,29	0,24	0,03	0,32	0,47	-0,22	0,36	0,19
DQU	0	0	0	0,07	0	0	0	0	0,27	0	0	0	0	0
CHIIVA	0,01	-0,05	-0,06	l	0,04	0,07	0,11	0,32	0,02	0,13	0,01	0,26	0,04	0,03
СПОУА	0,75	0,14	0,07	0	0,14	0,02	0,01	0	0,48	0	0,7	0	0,2	0,31
COND	0,23	0,2	0,42	0,04	1	0,15	0,17	0,31	0,47	0,33	-0	-0,33	-0,1	0,03
COND	0	0	0	0,14	0	0	0	0	0	0	0,29	0	0,08	0,36
COP	0,12	0,15	0,20	0,07	0,15	1	0	0,17	0,12	-0,1	0,19	0,16	0,1	0,18
CUR	0	0	0	0,02	0	0	0,98	0	0	0	0	0	0	0
NAM	0,15	0,07	0,29	0,11	0,17	0	1	0,47	0,07	0,24	0.02	-0,2	0.03	0,09
	0	0,1	0	0,01	0	0,98	0	0	0,06	0	0,58	0	0,45	0,03
NINI	0,09	0,18	0,24	0,32	0,31	0,17	0,47	1	0,21	0,24	0,09	-0,12	-0,1	0,21
1414	0,15	0	0	0	0	0	0	0	0	0,01	0,13	0,04	0,2	0
	0,09	0,07	0,03	0,02	0,47	0,12	0,07	0,21	1	0,2	-0,1	-0,09	-0,2	-0,2
pri	0	0,01	0,27	0,48	0	0	0,06	0	0	0	0	0	0	0
eer	0,24	0,09	0,32	0,13	0,33	-0,12	0,24	0,24	0,2	1	0,04	-0,19	0,04	-0
	0	0,03	0	0	0	0	0	0,01	0	0	0,38	0	0,37	0,89
т	0,29	0,19	0,47	0,01	-0,03	0,19	0,02	0,09	-0,1	0,04	1	0,3	0,51	0,36
1	0	0	0	0,7	0,29	0	0,58	0,13	0	0,38	0	0	0	0
VA7	0,07	0,2	-0,22	0,26	-0,33	0,16	-0,2	-0,12	-0,1	-0,2	0,3	1	0,27	0,06
VAL	0,01	0	0	0	0	0	0	0,04	0	0	0	0	0	0,02
CEI	0,31	0,24	0,36	0,04	-0,05	0,1	0,03	-0,08	-0,2	0,04	0,51	0,27	1	0,42
CEL	0	0	0	0,2	0,08	0	0,45	0,2	0	0,37	0	0	0	0
DAD	0,09	0,09	0,19	0,03	0,03	0,18	0,09	0,21	-0,2	-0	0.36	0,06	0.42	1
rAr	0	0	0	0,31	0,36	0	0,03	0	0	0,89	0	0,02	0	0
Nota: Na	área som	breada s	ão apre	esentados o	os valor	es do í	ndice d	e corre	lação l	inear e	ntre ca	da par	de var	iáveis,

Tabela 1A: Matriz de correlação.

sendo os demais valores referentes à estimativa do p-valor para o teste de significância da correlação. Neste caso, a hipótese nula consiste na ausência de correlação entre duas variáveis, logo, para um nível de confiança de 95%, p-valor maior que 0,05 indica uma correlação não significativa.
ANEXO 2: Resultados do pré-processamento de dados via PCA e PLS.

						Predição d	a BOD de	saída				Predição da	DBO de er	itrada	
CPs	Autovalores	² VE (%)	² VA (%)	VLs	¹ VE (%)	¹ VA (%)	² VE (%)	² VA (%)	Inc. (%)	VLs	¹ VE (%)	¹ VA (%)	² VE (%)	² VA (%)	Inc.(%)
1	2.288	28.60	28.60	1	23.30	23.30	20.25	20.25		1	31.43	31.43	21.93	21.93	
2	1.759	22.00	50.60	2	6.05	29.34	17.58	37.82	25.95	2	8.45	39.89	14.55	36.48	26.90
3	1.206	15.10	65.70	3	2.18	31.52	15.17	52.99	7.44	3	2.19	42.08	19.36	55.84	5.49
4	0.864	10.80	75.60	4	0.39	31.91	16.63	69.63	1.23	4	0.42	42.50	12.42	68.26	1.00
5	0.737	9.20	85.70	5	0.07	31.98	9.45	79.07	0.22	5	0.07	42.57	9.98	78.25	0.17
6	0.474	5.90	91.60	6	0.01	32.00	7.27	86.35	0.04	6	0.02	42.59	8.06	86.30	0.04
7	0.362	4.50	96.10	7	0.00	32.00	8.11	94.46	0.00	7	0.00	42.59	7.03	93.33	0.00
8	0.311	3.90	100.00	8	0.00	32.00	5.54	100.00	0.00	8	0.00	42.59	6.67	100.00	0.00

Tabela 2.1A: Variância explicada e acumulada das componentes principais e variáveis latentes (Dados 1 - Modelagem estacionária).

Variância explicada (VE) e variância acumulada (VA), espaço das variáveis preditas (1) e espaço das variáveis preditoras (2). Aumento na VA (Inc.).

Tabela 2.2A: Variância explicada e acumulada das componentes principais e va	ariáveis latentes (Dados	 Modelagem dinâmica)
--	--------------------------	---

						Predição d	a BOD de s	saída		Predição da DBO de entrada							
CPs	Autovalores	² VE (%)	² VA (%)	VLs	⁺ VE (%)	¹ VA (%)	² VE (%)	² VA (%)	Inc. (%)	VLs	¹ VE (%)	¹ VA (%)	² VE (%)	² VA (%)	Inc.(%)		
1	3.811	23.80	23.80	1	24.75	24.75	17.99	17.99	=	1	26.92	26.92	19.75	19.75	~		
2	3.168	19.80	43.60	2	9.95	34.70	13.50	31.49	40.20	2	15.12	42.03	9.69	29.44	56.16		
3	1.843	11.50	55.10	3	2.85	37.54	9.63	41.11	8.21	3	2.00	44.03	17.30	46.74	4.75		
4	1.382	8.60	63.80	4	0.38	37.93	15.84	56.95	1.02	4	0.69	44.72	9.82	56.56	1.57		
5	1.150	7.20	71.00	5	0.20	38.13	8.31	65.26	0.53	5	0.25	44.97	6.39	62.95	0.55		
6	0.999	6.20	77.20	6	0.07	38.20	5.56	70.82	0.20	6	0.09	45.06	7.40	70.35	0.20		
7	0.740	4.60	81.80	7	0.03	38.23	4.44	75.26	0.07	7	0.05	45.11	3.47	73.82	0.11		
8	0.638	4.0	85.80	8	0.01	38.24	4.07	79.33	0.03	8	0.01	45.12	5.58	79.40	0.03		
9	0.526	3.30	89.10	9	0.00	38.24	4.69	84.02	0.00	9	0.00	45.12	2.97	82.37	0.01		
10	0.462	2.90	92.00	10	0.00	38.24	3.92	87.94	0.00	10	0.00	45.12	4.26	86.63	0.00		
11	0.368	2.30	94.30	11	0.00	38.24	2.93	90.87	0.00	11	0.00	45.12	3.48	90.11	0.00		
12	0.272	1.70	96.00	12	0.00	38.24	2.36	93.23	0.00	12	0.00	45.12	2.70	92.81	0.00		
13	0.190	1.20	97.20	13	0.00	38.24	2.46	95.70	0.00	13	0.00	45.12	2.70	95.51	0.00		
14	0.174	1.10	98.40	14	0.00	38.85	2.14	97.90	0.00	14	0.00	45.12	2.47	97.98	0.00		
15	0.138	0.90	99.20	15	0.00	38.85	1.07	98.9 7	0.00	15	0.00	45.12	1.03	99.01	0.00		
16	0.124	0.80	100.00	16	0.00	38.85	1.03	100.00	0.00	16	0.00	45.12	0.99	100.00	0.00		

Variância explicada (VE) e variância acumulada (VA), espaço das variáveis preditas (1) e espaço das variáveis preditoras (2). Aumento na VA (Inc.).

						Predição d	a BOD de s	saída			********	Predição da	DBO de er	itrada	
CPs	Autovalores	² VE (%)	² VA (%)	VLs	¹ VE (%)	¹ VA (%)	² VE (%)	² VA (%)	Inc. (%)	VLs	¹ VE (%)	¹ VA (%)	² VE (%)	² VA (%)	Inc.(%)
1	3.375	28.10	28.10	1	48.81	48.81	12.76	12.76	-	1	55.85	55.85	14.12	14.12	-
2	1.800	15.00	43.10	2	10.13	58.94	13.41	26.17	20.75	2	7.61	63.47	11.13	25.25	13.63
3	1.535	12.80	55.90	3	2.00	60.94	22.35	48.52	3.40	3	1.22	64.68	23.76	49.01	1.92
4	1.230	10.20	66.20	4	1.83	62.77	5.61	54.14	3.01	4	1.46	66.15	4.12	53.13	2.26
5	1.036	8.60	74.80	5	0.40	63.18	8.20	62.34	0.64	5	0.12	66.27	11.00	64,13	0.18
6	0.820	6.80	81.60	6	0.21	63.39	5.29	67.63	0.33	6	0.05	66.32	5.50	69.63	0.08
7	0.748	6.20	87.90	7	0.06	63.45	9.05	76.67	0.10	7	0.02	66.34	7.06	76.69	0.03
8	0.411	3.40	91.30	8	0.03	63.48	5.59	82.26	0.05	8	0.01	66.35	4.39	81.08	0.02
9	0.3780	3.10	94.40	9	0.01	63.49	3.67	85.93	0.02	9	0.00	66.35	6.58	87.66	0.00
10	0.295	2.50	96.90	10	0.00	63.49	7.92	93.85	0.00	10	0.00	66.35	6.11	93.77	0.00
11	0.2521	2.10	99.00	11	0.00	63.49	3.68	97.53	0.00	11	0.00	66.35	3.12	96.89	0.00
12	0.1203	1.00	100.00	12	0.00	63.49	2.47	100.00	0.00	12	0.00	66.35	3.11	100.00	0.00

Tabela 2.3A: Variância explicada e acumulada das componentes principais e variáveis latentes (Dados 2).

Variância explicada (VE) e variância acumulada (VA), espaço das variáveis preditas (1) e espaço das variáveis preditoras (2). Aumento na VA (Inc.).

130

Tabela 2	2.4A: LOA	aings das	componen	tes princip	ais (Dados	1 - Mode	elagem esta	icionaria)
VARIÁVIEIS	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8
DQO	0,19	0,62	0,01	0,08	0,24	0,49	0,18	-0,49
CEL	0,51	0,16	-0,02	0,29	-0,40	0,35	-0,39	0,45
COND	-0,33	0,53	-0,14	-0,07	-0,26	-0,10	0,51	0,49
COR	0,16	0,03	-0,69	-0,51	0,38	0,10	-0,18	0,21
PAP	0,41	0,09	0,12	-0,62	-0,53	-0,23	0,10	-0,30
pН	-0,40	0,18	-0,50	0,21	-0,42	-0,14	-0,41	-0,39
Т	0,44	0,26	-0,13	0,37	0,26	-0,72	0,04	-0,02
VAZ	0,24	-0,45	-0,47	0,29	-0,22	0,17	0,59	-0,14

Tabela 2.4A: Loadings das componentes principais (Dados 1 – Modelagem estacionária).

VARIÁVIEIS	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11	CP12	CP13	CP14	CP15	CP16
$\frac{1}{DOO(k)}$	0.21	0.39	-0.07	-0.07	-0.08	-0.06	0.49	0.11	-0.07	0,45	0.16	0,14	-0.29	-0.06	0.09	-0.44
DQO (k-1)	0,23	0,39	-0,07	-0,12	-0,18	-0,03	0,22	-0,34	-0,04	0,19	-0,48	-0,12	0,31	0,00	-0,05	0,45
CEL (k)	0,37	0,01	0,13	-0,17	0,05	0,31	0,33	0,32	0,07	-0,37	0,20	-0,50	0,26	-0,08	-0,06	-0,05
CEL (k-1)	0,37	0,06	0,06	-0,14	0,30	0,13	0,07	-0,34	0,51	-0,29	0,06	0,46	-0,20	0,10	-0,03	0,06
COND (k)	-0,14	0,37	-0,23	0,03	0,42	-0,05	0,08	0,00	-0,42	-0,23	0,22	0,09	0,17	0,53	0,10	0,06
COND (k-1)	-0,15	0,38	-0,21	-0,02	-0,05	0,35	-0,25	-0,39	-0,19	-0,21	0,20	-0,18	-0,24	-0,48	-0,09	-0,07
COR (k)	0,15	-0,11	-0,56	0,28	-0,21	-0,10	0,14	0,19	0,04	-0,09	0,30	0,34	0,26	-0,33	0,08	0,28
COR (k-1)	0,19	-0,12	-0,54	0,24	-0,24	-0,10	-0,03	-0,16	0,17	-0,18	-0,20	-0,34	-0,22	0,39	-0,10	-0,30
pH (k)	-0,20	0,07	-0,37	-0,18	0,54	-0,08	0,06	0,37	0,18	0,00	-0,43	-0,08	-0,17	-0,29	-0,10	0,04
pH (k-1)	-0,21	0,07	-0,27	-0,32	-0,18	0,58	-0,21	0,17	0,30	0,33	0,09	0,08	0,17	0,30	0,05	0,02
PAP (k)	0,25	0,01	0,10	0,43	-0,01	0,53	-0,07	0,27	-0,32	-0,06	-0,41	0,27	-0,17	0,04	0,01	0,01
PAP (k-1)	0,21	0,06	0,03	0,52	0,45	0,06	-0,21	-0,12	0,23	0,46	0,19	-0,28	0,20	-0,05	0,04	0,01
T (k)	0,40	0,09	-0,03	-0,22	-0,04	-0,20	-0,39	0,25	-0, 17	0,15	0,20	0,01	-0,21	0,10	-0,58	0,23
T (k-1)	0,38	0,12	-0,06	-0,26	-0,01	-0,19	-0,49	0,09	-0,08	-0,08	-0,14	0,02	0,16	-0,11	0,60	-0,23
VAZ (k)	0,15	-0,43	-0,15	-0,20	0,14	0,15	0,15	-0,16	-0,26	0,20	0,12	-0,21	-0,41	0,04	0,38	0,39
VAZ (k-1)	0,14	-0,41	-0,18	-0,23	0,21	0,13	0,05	-0,30	-0,32	0,14	-0,10	0,18	0,38	-0,11	-0,31	-0,40

Tabela 2.5A: Loadings das componentes principais (Dados 1 – Modelagem dinâmica).

Tabela 2.6A: Loadings das componentes principais (Dados 2).

					0	1	I	- I - · · ·				
	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11	CP12
DQO	0,11	0,58	0,14	0,06	0,00	-0,44	0,34	-0,20	0,13	0,20	-0,22	-0,42
CEL	0,43	0,29	-0,21	0,00	-0,24	-0,08	0,14	0,03	0,01	0,16	-0,07	0,76
COND	-0,40	0,29	0,13	-0,23	0,15	0,09	0,15	0,05	0,14	-0,62	-0,39	0,27
COR	0,22	0,00	0,35	-0,58	-0,15	-0,15	-0,28	0,22	-0,49	0,07	-0,27	-0,07
CHUVA	0,17	0,17	0,01	0,16	0,82	0,11	-0,35	-0,04	-0,07	0,22	-0,22	0,12
NAM	-0,11	0,09	-0,58	-0,07	-0,14	-0,39	-0,55	-0,33	-0,04	-0,22	-0,09	-0,07
NN	-0,32	0,11	-0,25	-0,38	0,35	-0,26	0,31	0,07	-0,32	0,13	0,51	0,12
PAP	0,12	0,15	-0,40	-0,52	-0,03	0,59	0,09	-0,12	0,23	0,20	-0,10	-0,24
pН	-0,33	0,13	0,42	-0,23	-0,11	-0,04	-0,37	-0,22	0,46	0,36	0,25	0,23
SST	-0,20	0,46	0,11	0,27	-0,23	0,43	-0,11	-0,27	-0,56	0,00	0,17	-0,01
Т	0,39	0,38	0,07	-0,04	0,05	0,04	-0,24	0,36	0,21	-0,42	0,52	-0,16
VAZ	0,38	-0,22	0,24	-0,19	0,15	0,00	0,16	-0,73	-0,05	-0,31	0,18	0,06

		5 11 A	T TT - 2			*** <	X 17 CT	111.0
		VL2	VL3	VL4	VL5	VL6	VL7	<u>VL8</u>
DQO (k)	0.725	0.155	0.038	0.394	0.030	0.097	-0.186	0.274
CEL (k)	0.488	-0.309	0.454	-0.431	-0.487	-0.180	0.409	-0.123
COND (k)	0.300	0.311	-0.763	0.203	-0.059	-0.198	-0.143	-0.649
COR (k)	0.153	-0.189	-0.012	-0.247	0.814	-0.953	-0.160	0.332
PAP (k)	0.193	-0.735	0.328	0.357	0.310	-0.121	0.588	-0.538
pH (k)	0.051	0.628	-0.619	-0.301	0.384	0.089	0.657	-0.028
T (k)	0.462	-0.637	0.030	-0.473	0.171	0.521	-0.334	0.141
VAZ (k)	-0.100	0.229	0.641	-0.668	0.492	-0.008	-0.103	-0.454

Tabela 2.7A: Loadings das variáveis latentes - DBO de entrada (Dados 1 - Modelagem estacionária).

Tabela 2.8A: Loadings das variáveis latentes - DBO de entrada (Dados 1 – Modelagem dinâmica).

	VL1	VL2	VL3	VL4	VL5	VL6	VL7	VL8	VL9	VL10	VLII	VL12	VL13	VL14	VL15	VL16
DQO (k)	0.556	0.215	-0.032	0.338	0.256	0.060	0.025	0.213	0.033	0.013	0.141	0.050	-0.037	0.056	-0.066	-0.032
DQO (k-1)	0.512	-0.066	-0.163	0.091	-0.077	-0.196	-0.522	0.137	-0.139	0.205	0.057	-0.640	0.227	0.181	-0.054	0.047
CEL (k)	0.385	-0,100	0.331	-0.235	-0.518	-0.004	-0.011	0.065	0.046	0.333	-0.276	0.446	-0.419	-0.294	0.042	0.099
CEL (k)	0.375	-0.278	0.188	-0.329	-0.337	0.206	0.679	-0.183	-0.618	-0.040	0.270	-0.079	0.056	0.299	0.035	-0.047
COND(k)	0.198	0.283	-0.471	0.096	0.004	-0.228	0.353	-0.512	0.283	-0.318	-0.002	0.200	0.402	-0.379	0.371	-0.417
COND (k-1)	0.125	0.008	-0.586	0.085	0.176	-0.077	0.363	-0.048	-0.345	0.523	-0.221	-0.166	0.186	-0.514	-0.214	0.375
COR (k)	0.071	-0.088	0.181	-0.121	0.671	-0.935	0.123	0.065	-0.274	-0.285	0.029	0.421	0.014	-0.013	-0.366	0.170
COR (k-1)	0.057	-0.282	0.146	-0.262	0.594	-0.861	0.381	0.287	0.082	0.141	0.172	-0.310	-0.133	0.019	0.425	-0.189
PAP (k)	0.156	-0.441	0.258	0.283	-0.091	-0.178	0.387	-0.248	0.142	0.464	-0.993	0.218	0.235	0.211	-0.021	-0.031
PAP (k-1)	0.144	-0.444	0.153	0.300	0.101	-0.091	0.418	-0.965	0.112	0.103	0.390	-0.133	-0.326	0.017	-0.048	0.009
pH (k)	-0.016	0.484	-0.298	-0.324	0.173	-0.193	0.486	-0.465	0.639	-0.386	-0.136	-0.060	-0.336	0.400	-0.195	0.261
pH (k-1)	-0.073	0.313	-0.332	-0.234	0.435	0.046	-0.263	-0.051	-0.535	0.799	-0.428	0.390	-0.350	0.342	0.139	-0.228
T (k)	0.387	-0.477	0.131	-0.250	0.354	0.313	-0.410	0.043	0.147	-0.402	-0.148	0.160	-0.101	-0.109	0.460	0.356
T (k-1)	0.374	-0.552	0.013	-0.389	0.194	0.241	-0.287	0.113	0.387	-0.144	0.066	0.095	0.070	0.035	-0.469	-0.371
VAZ (k)	-0.098	0.157	0.528	-0.486	0.318	0.069	0.068	-0.194	-0.139	0.071	-0.256	-0.342	0.018	-0.370	-0.106	-0.339
VAZ (k-1)	-0.102	0.112	0.448	-0.570	0.256	0.026	0.018	-0.289	0.127	0.337	0.135	-0.080	0.568	-0.051	0.055	0.332

L.

-	VL1	VL2	VL3	VL4	VL5	VL6	VL7	VL8	VL9	VL10	VL11	VL12
DQO (k)	0.735	0.107	0.147	0.023	0.040	-0.298	-0.076	0.003	-0.243	-0.235	0.132	0.189
CEL (k)	0.360	-0.370	0.530	0.055	-0.163	0.487	-0.078	0.318	0.295	-0.170	0.275	0.037
CHUVA (k)	0.141	-0.348	0.162	0.390	-0.019	-0.294	0.368	0.293	-0.765	0.839	-0.247	-0.001
COND (k)	0.080	0.356	-0.375	0.104	-0.643	-0.383	0.609	0.119	-0.071	0.011	0.411	-0.272
COR (k)	0.126	0.245	-0.466	-0.041	0.378	-0.358	0.299	-0.424	0.581	0.290	0.026	0.588
NAM (k)	0.392	-0.635	0.306	-0.759	-0.042	0.137	0.201	-0.229	-0.016	0.149	-0.251	0.063
NN (k)	0.220	0.150	-0.589	0.368	0.039	-0.085	0.629	0.004	-0.067	-0.310	-0.483	-0.038
PAP (k)	-0.031	-0.640	0.023	0.721	-0.551	-0.094	0.181	-0.488	0.325	-0.311	0.171	0.277
pH (k)	0.363	-0.305	-0.450	0.254	0.250	-0.100	-0.654	0.360	0.243	0.050	0.046	-0.462
SST (k)	0.077	-0.074	0.309	-0.069	0.264	-0.303	0.689	-1.135	0.632	0.046	0.049	-0.501
T (k)	0.099	0.295	-0.047	0.295	-0.681	0.501	-0.456	-0.259	0.439	0.394	-0.603	-0.001
VAZ (k)	-0.144	-0.189	0.539	-0.195	0.121	-0.918	0.253	0.399	0.259	-0.123	-0.360	0.043

Tabela 2.9A: Loadings das variáveis latentes - DBO de entrada (Dados 2 - Modelagem dinâmica).

Tabela 2.10A: Loadings das variáveis latentes - DBO de saída (Dados 1 - Modelagem estacionária).

	VL1	VL2	VL3	VL4	VL5	VL6	VL7	VL8
DQO (k)	0.613	-0.114	0.058	0.641	-0.056	0.020	-0.163	0.454
CEL (k)	0.535	-0.420	0.356	-0.282	-0.677	-0.180	0.134	-0.056
COND (k)	0.226	0.335	-0.669	0.545	-0.173	-0.035	-0.011	-0.752
COR (k)	0.367	-0.047	-0.309	-0.322	0.770	-0.777	-0.074	0.131
PAP(k)	0.206	-0.606	0.334	0.089	0.287	-0.120	0.844	-0.455
pH (k)	0.144	0.656	-0.744	-0.092	-0.123	0.261	0.504	0.187
T(k)	0.490	-0.621	-0.119	-0.198	0.088	0.659	-0.371	0.011
VAZ (k)	0.222	0.359	0.606	-0.700	0.313	0.186	-0.032	-0.293

	VL1	VL2	VL3	VL4	VL5	VL6	VL7	VL8	VL9	VL10	VL11	VL12	VL13	VL14	VL15	VL16
DQO (k)	0.442	-0.023	0.093	0.493	-0.117	0.005	-0.223	0.175	-0.196	-0.102	0.281	-0.531	-0.285	0.126	-0.098	-0.063
DQO (k-1)	0.438	-0.101	0.000	0.455	-0.113	0.111	-0.300	0.212	-0.161	0.073	-0.031	0.750	-0.006	-0.245	0.226	0.000
CEL (k)	0.398	-0.276	0.342	-0.171	-0.447	-0.081	0.291	0.179	-0.344	-0.132	-0.047	-0.174	0.587	0.147	-0.303	-0.077
CEL (k)	0.406	-0.324	0.142	-0.177	-0.484	-0.116	0.183	0.042	0.395	-0.047	-0.012	0.325	-0.663	0.065	0.113	0.095
COND (k)	0.121	0.296	-0.395	0.411	-0.250	-0.119	0.033	-0.384	0.636	-0.187	-0.090	-0.280	0.443	0.032	0.023	0.529
COND(k-1)	0.074	0.227	-0.476	0.418	-0.228	0.083	0.302	-0.149	0.059	0.482	-0.786	0.167	-0.052	0.135	-0.145	-0.484
COR (k)	0.279	0.014	-0.329	-0.183	0.572	-0.755	-0.073	0.103	-0.059	0.226	0.316	-0.185	0.150	0.235	0.043	-0.331
COR (k-1)	0.297	-0.055	-0.319	-0.201	0.673	-0.453	0.210	0.339	-0.184	-0.142	-0.363	0.205	-0.153	-0.169	-0.116	0.386
PAP (k)	0.162	-0.379	0.264	0.067	0.072	-0.349	0.744	-0.566	-0.234	0.537	-0.033	-0.286	0.223	-0.586	0.353	0.036
PAP (k-1)	0.131	-0.375	0.103	0.147	0.238	-0.194	0.668	-0.828	0.301	-0.487	0.174	0.303	-0.253	0.464	-0.245	-0.047
pH (k)	0.051	0.478	-0.457	-0.051	-0.094	0.176	0.393	-0.166	0.291	-0.707	0.564	-0.131	0.085	-0.403	0.172	-0.289
pH (k-1)	0.004	0.440	-0.412	-0.065	-0.218	0.263	0.231	-0.304	-0.763	0.580	0.135	0.115	-0.250	0.249	-0.078	0.296
T(k)	0.407	-0.492	-0.091	-0.067	0.091	0.574	-0.156	0.110	0.152	0.233	0.255	-0.264	0.192	0.063	-0.377	0.095
T(k-1)	0,401	-0.496	-0.231	-0.105	-0.070	0.439	-0.447	-0.188	-0.049	-0.153	-0.193	-0.099	0.078	-0.100	0.411	-0.106
VAZ (k)	0.178	0.193	0.415	-0.560	0.373	0.188	0.150	-0.045	0.139	0.068	-0.193	-0.076	0.063	0.290	0.448	0.042
VAZ (k-1)	0.183	0.201	0.326	-0.575	0.275	0.061	-0.228	-0.446	0.232	0.086	-0.206	0.163	-0.052	-0.257	-0.408	-0.112

Tabela 2.11A: Loadings das variáveis latentes - DBO de saída (Dados 1 - Modelagem dinâmica).

Tabela 2.12A: Loadings das variáveis latentes - DBO de saída (Dados 2).

	VL1	VL2	VL3	VL4	VL5	VL6	VL7	VL8	VL9	VL10	VL11	VL12
DQO (k)	0.794	0.182	-0.221	0.187	-0.139	-0.233	0.053	-0.256	-0.136	-0.004	-0.056	-0.205
CEL (k)	-0.186	-0.262	-0.202	0.997	-0.702	0.028	-0.346	0.401	-0.324	-0.132	0.167	-0.222
CHUVA (k)	0.023	-0.108	-0.342	0.234	0.286	0.070	-0.044	-0.011	-0.280	0.852	-0.511	-0.205
COND (k)	0.356	-0.522	0.434	0.371	-0.100	-0.015	0.420	-0.044	-0.292	0.117	0.114	0.587
COR (k)	-0.174	0.055	-0.308	0.223	-0.424	0.199	0.853	-0.948	0.146	-0.225	-0.246	-0.035
NAM (k)	0.068	-0.133	0.138	0.173	-0.229	0.938	-1.034	-0.396	0.606	-0.026	-0.251	0.243
NN (k)	0.200	-0.240	0.415	-0.177	-0.884	0.328	0.074	0.331	-0.002	0.321	-0.294	-0.084
PAP(k)	0.226	0.358	-0.585	0.403	-0.118	0.575	-0.128	0.367	0.183	-0.235	0.139	-0.162
pH (k)	0.152	-0.478	0.340	0.171	0.103	-0.495	0.480	-0.618	0.844	0.022	0.219	-0.382
SST(k)	0.390	-0.685	-0.063	-0.127	0.354	-0.329	-0.033	0.360	0.055	-0.376	-0.417	0.074
T(k)	0.176	-0.019	-0.715	0.049	-0.167	-0.063	0.024	-0.173	0.239	0.184	0.381	0.387
VAZ (k)	-0.253	0.587	-0.301	0.453	-0.115	-0.727	0.342	0.038	0.356	0.078	-0.435	0.358