

Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica e de Computação

Identificação de Modelos para Controle Preditivo: Aplicação a uma Planta de Lodos Ativados

Autor: José María Vargas Lara

Orientador: Prof. Dr. Basílio Ernesto de Almeida Milani

Tese de Doutorado apresentada à Faculdade de Engenharia Elétrica e de Computação como parte dos requisitos para obtenção do título de Doutor em Engenharia Elétrica. Área de concentração: **Automação**.

Banca Examinadora

Basílio E. A. Milani DT/FEEC/Unicamp
Cláudio Garcia PTC/LAC/Poli-USP
Gilmar Barreto DMCSI/FEEC/Unicamp
Jussara Farias Fardin DEL/UFES
Wagner Caradori do Amaral DCA/FEEC/Unicamp

Campinas, SP

Fevereiro/2005

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

V426i Vargas Lara, José María
Identificação de modelos para controle preditivo:
aplicação a uma planta de lodos ativados /
José María Vargas Lara. – Campinas, SP: [s.n.], 2005.

Orientador: Basílio Ernesto de Almeida Milani.
Tese (doutorado) - Universidade Estadual de Campinas,
Faculdade de Engenharia Elétrica e de Computação.

1. Controle preditivo. 2. Identificação de sistemas. 3.
Águas residuais - Purificação - Processo de lodo ativado.
4. Águas residuais - Purificação - remoção de nitrogênio. I.
Milani, Basílio Ernesto de Almeida. II. Universidade Estadual
de Campinas. Faculdade de Engenharia Elétrica e de
Computação. III. Título

Titulo em Inglês: Identification of models for predictive control: application to
activated sludge plant

Palavras-chave em Inglês: Predictive control, System identification, Sewage -
Purification - Activated Sludge process e Sewage -
Purification - Nitrogen removal

Área de concentração: Automação

Titulação: Doutor em Engenharia Elétrica

Banca examinadora: Cláudio Garcia, Gilmar Barreto, Jussara Farias Fardin e Wagner
Caradori do Amaral.

Data de defesa: 14/02/2005

Vargas Lara, J. M. (2005). *Identificação de Modelos para Controle Preditivo: Aplicação a uma Planta de Lodos Ativados*, Tese de Doutorado, DT/FEEC - Universidade Estadual de Campinas, Campinas, Brasil.

Resumo

Este trabalho trata da síntese de um controlador preditivo adaptativo, baseado em modelos reduzidos obtidos mediante identificação orientada para controle preditivo. A metodologia proposta envolve duas áreas importantes para o sucesso do experimento de identificação. Na primeira, projeto de sinais de excitação, é mostrado como sinais de excitação pseudo-aleatórios multi-níveis podem ser projetados para atender com eficácia requisitos que sinais de excitação relevantes para controle devem satisfazer. Na segunda, modelagem de preditores, é abordada a modelagem por identificação de preditores de horizonte estendido. O objetivo central da metodologia proposta é o controle da remoção de matéria nitrogenada de uma planta de tratamento de esgotos por lodos ativados com pré-desnitrificação. A metodologia é aplicada para controlar a concentração de amônia mediante o controle do *set-point* de oxigênio dissolvido. A metodologia também é aplicada para o controle da concentração de nitrato no reator desnitrificante. Os controladores são avaliados em um estudo de simulação, mostrando a eficácia da metodologia proposta para o controle das concentrações de amônia e nitrato.

Palavras-chave: Controle preditivo, identificação de sistemas, sinais de excitação, preditores k -passos à frente, lodos ativados, remoção de nitrogênio.

Abstract

This work deals with the synthesis of an adaptive predictive controller, based on reduced models obtained via identification oriented to predictive control. The methodology proposed involves two important areas for the success of the identification experiment. In the first one, excitation signal design, it is shown how pseudo-random multilevel signals can be designed to effectively satisfy the requirements that excitation signals relevant for control should satisfy. In the second, predictor modelling, the identification of long-range horizon predictors is approached. The main objective of the proposed methodology is the control of the nitrogen removal in an predenitrifying activated sludge wastewater treatment plant. This methodology is applied to control the ammonium concentration in the denitrifying reactor. The controllers are evaluated in a simulation study, showing the effectiveness of the proposed methodology to control ammonium and nitrate concentrations.

Keywords: Predictive control, system identification, excitation signals, k -step ahead predictors, activated sludge, nitrogen removal.

*A Deus pela força e coragem nesta caminhada,
a minha amada família a quem dedico esta conquista,
ao meu amado país o Peru e
a esta querida terra que levo no coração, o Brasil.*

Agradecimentos

Ao meu orientador o Prof. Basílio E. A. Milani, pelo constante apoio e disposição ao desenvolvimento de minha pesquisa, sou grato pela orientação. Ao Prof. Wagner Caradori de Amaral pelas sugestões e críticas a parte deste trabalho.

Ao Dr. Oscar A. Z. Sotomayor por disponibilizar o simulador *bechmark* ASWWT-USP, assim como também informação valiosa respeito do processo. Ao Prof. Anthony Barker da Universidade de Wales Swansea, UK, por enviar tão gentilmente o programa Galois®.

Aos colegas do Laboratório de Telemática que tornaram a permanência neste belo país mais grata ainda. Aos colegas conterrâneos com os que dividi gratos momentos, Yusef, Luis, Miguel, Alex, entre tantos outros.

Aos amigos que mostraram que a frase o Brasil é bonito tem sentido não somente pelas suas belezas naturais senão também pela sincera amizade que sua gente sabe demonstrar.

À Coordenação de Aperfeiçoamento do Pessoal de Nível Superior, CAPES pelo apoio financeiro.

E finalmente agradeço aos meus pais Celestino e Margarita pela paciência, pelo incentivo e especialmente por entenderem a minha ausência. A minhas irmãs Sofia, Liliana e Milagros pelo constante apoio e principalmente o carinho. Aos meus sobrinhos José Raul, Andrea e Edson por estarem sempre prestes a me esperar. Eles nunca estiveram longe.

Sumário

Lista de Figuras	xiii
Lista de Tabelas	xvii
Glossário	xix
Lista de Símbolos	xxiii
Trabalhos Publicados Pelo Autor	xxvii
1 Introdução	1
1.1 Introdução ao Tratamento de Esgotos Domésticos	1
1.1.1 Importância do Tratamento de Esgotos	2
1.1.2 Panorama atual	2
1.1.3 O papel da automação e controle em estações de tratamento de esgotos	2
1.2 Estações de Tratamento de Esgotos	3
1.3 Processo de Lodos Ativados	5
1.3.1 Modelo Matemático do PLA	6
1.3.2 <i>Benchmark</i> do PLA	6
1.4 Dificuldades e Problemas no Controle de PLA	7
1.4.1 Problemas de Observabilidade e Controlabilidade	7
1.4.2 Distúrbios	8
1.4.3 Incertezas	8
1.5 Modelagem para controle do PLA	8
1.5.1 Sinais de Excitação Relevantes para Controle	9
1.5.2 Identificação de Modelos para Controle Preditivo	10
1.6 O MPC Aplicado à Remoção Biológica de Nitrogênio	10
1.7 Objetivos da tese	12
1.7.1 Restrições	12
1.8 Estrutura da Tese	13
2 Modelo do Processo de Lodos Ativados e Modelo de Simulação	15
2.1 Mecanismos Biológicos	15
2.1.1 Microrganismos e suas atividades	16
2.1.2 Hidrólise	17

2.1.3	Crescimento da biomassa	18
2.1.4	Decaimento da biomassa	18
2.1.5	Fatores ambientais que afetam o crescimento de biomassa	19
2.2	Processo Aeróbio de Remoção de Matéria Carbonácea	20
2.2.1	Divisão da matéria carbonácea	20
2.3	Processo de Remoção de Matéria Nitrogenada	23
2.3.1	Divisão da matéria nitrogenada	23
2.3.2	Processo de Nitrificação	24
2.3.3	Processo de Desnitrificação	26
2.4	O Modelo IAWQ No. 1	27
2.5	Modelo de Simulação	29
2.6	Conclusões	33
3	Projeto de Sinais de Identificação para Controle	35
3.1	Introdução	35
3.2	Tipos de sinais de excitação	37
3.2.1	Sinais de excitação aleatórios	37
3.2.2	Sinais de excitação pseudo-aleatórios	38
3.3	Geração de sinais pseudo-aleatórios multi-níveis	44
3.3.1	Campos de Galois	44
3.3.2	Seqüências m em $GF(q)$	45
3.3.3	Sinal de seqüência m multi-nível	46
3.4	Análise de harmônicos	47
3.4.1	Conteúdo de harmônicos de sinais pseudo-aleatórios multi-níveis	47
3.4.2	Supressão de harmônicos por mapeamento	48
3.5	Medida de qualidade para um sinal de excitação	51
3.6	Critérios e parâmetros para o projeto de sinais MLPRS	52
3.6.1	Faixa de freqüência relevante para excitação do sistema	52
3.6.2	Faixa de freqüência relevante para controle	52
3.6.3	Parâmetros para o projeto do sinal	53
3.7	Exemplo de aplicação da metodologia	55
3.7.1	Condições experimentais	56
3.7.2	Projeto dos sinais de excitação	56
3.7.3	Algoritmo de identificação	58
3.7.4	Avaliação de desempenho	58
3.8	Identificação do processo de remoção de matéria carbonácea	64
3.8.1	Identificação do Processo de Oxigênio Dissolvido	64
3.8.2	Projeto dos sinais de excitação da planta	65
3.8.3	Identificação e validação dos modelos	67
3.9	Conclusão	73

4	Identificação de Modelos para Controle Preditivo	75
4.1	Identificação e controle	75
4.2	Identificação de modelos para controle	76
4.2.1	Especificação de modelos para controle usando pré-filtragem	77
4.3	Identificação para Controle Preditivo	79
4.3.1	Conceitos Gerais e Formulação do MPC	80
4.3.2	Limitações do preditor um-passo-à frente	86
4.4	Modelagem de Preditores k -passos-à frente	88
4.4.1	Modelo preditor por iteração sucessiva padrão	88
4.4.2	Modelo preditor de identificação preditiva de horizonte estendido	90
4.4.3	Modelo preditor de múltiplos-passos-à frente	94
4.4.4	Erros de Polarização	98
4.4.5	Relação erro de predição	100
4.4.6	Comparação dos diferentes tipos de modelos preditores: estudo de caso . . .	101
4.5	Conclusão	106
5	Controle Preditivo Baseado em Modelos de Predição Múltiplos	109
5.1	Controle Preditivo Generalizado (GPC)	109
5.1.1	Formulação padrão do controlador GPC	110
5.1.2	Formulação Modificada: Controle Preditivo Generalizado usando Modelos de Predição Múltiplos	117
5.1.3	GPC auto-ajustável	120
5.2	Desempenho do controlador preditivo usando os diferentes modelos de predição . . .	122
5.3	Conclusão	127
6	O MPC Aplicado à Remoção Biológica de Nutrientes	129
6.1	A Remoção Biológica de Nitrogênio	130
6.1.1	O aspecto do acoplamento	130
6.2	Controle da nitrificação	131
6.2.1	Identificação do modelo reduzido para a concentração de amônia	132
6.2.2	Resultados da simulação em malha fechada	135
6.3	Controle da desnitrificação	141
6.3.1	Identificação do modelo de ordem reduzida da concentração de nitrato	142
6.3.2	Controle da concentração de nitrato	145
6.4	Conclusões	150
7	Conclusões e Tópicos para Futuras Pesquisas	153
7.1	Introdução	153
7.2	Contribuições do trabalho	153
7.2.1	Geração de sinais	153
7.2.2	Identificação	154
7.2.3	Modelagem para controle preditivo	154
7.2.4	Controle GPC baseado em modelos preditores	154
7.2.5	Controle do processo de remoção de nitrogênio	155

7.3	Conclusões	155
7.3.1	Geração de sinais	155
7.3.2	Identificação	155
7.3.3	Modelagem para controle preditivo	156
7.3.4	Controle GPC baseado em modelos preditores	156
7.3.5	Controle do processo de remoção de nitrogênio	157
7.4	Tópicos para futuras pesquisas	158
7.4.1	Geração de sinais de excitação	158
7.4.2	Identificação para controle preditivo	158
7.4.3	Controle preditivo baseado em multimodelos	159
7.4.4	Controle da remoção de nutrientes	159
Referências bibliográficas		161
A Indicadores para Caracterização de Afluentes em ETE's		171
A.1	Matéria Orgânica	171
A.2	Nutrientes	172
A.2.1	Nitrogênio	172
A.2.2	Fósforo	172
A.3	Outros Parâmetros	172
B Definições importantes para o experimento de identificação		175
B.1	Estabilidade de uma função de transferência	175
B.2	Funções de autocorrelação e de potência de um sinal	175
B.3	Experimentos informativos	176
B.3.1	Persistência de excitação	176
C Algoritmo de Identificação Não Linear Multivariável		179
C.1	Representação do processo	179
C.1.1	Aproximação polinomial	179
C.1.2	Conjunto de termos candidatos	181
C.1.3	Redução do modelo	181
C.2	Estimação de parâmetros	182
C.2.1	Método de Gram-Schmidt Modificado	182
C.3	Deteccão da estrutura	184
C.3.1	Taxa de redução do erro	184
C.3.2	Critério de parada	184
C.4	Algoritmo combinado de estimação de parâmetros e deteccão de estrutura	185

Lista de Figuras

1.1	Diagrama típico de uma Estação de Tratamento de Esgotos.	4
1.2	Sistema de lodos ativados básico.	5
1.3	Sistema de lodos ativados com pré-desnitrificação.	6
2.1	Mecanismos biológicos do processo de lodos ativados (Carstensen, 1994).	17
2.2	Divisão da matéria orgânica carbonácea segundo sua biodegradabilidade.	20
2.3	Crescimento heterotrófico segundo a hipótese de morte-regeneração.	21
2.4	Principais reações na remoção de carbono orgânico.	22
2.5	Divisão da matéria nitrogenada segundo a medição do NTK.	23
2.6	Principais reações no processo de nitrificação e desnitrificação na remoção de matéria nitrogenada.	26
2.7	Processos biológicos de oxidação de carbono, nitrificação e desnitrificação considerados pelo modelo IAWQ No. 1: (ae) denota o processo aeróbio e (an) denota o processo anóxico.	28
2.8	Esquema do modelo <i>benchmark</i> de simulação ASWWTP-USP.	29
2.9	Modelo do decantador segundo o princípio de múltiplas camadas.	32
3.1	Sinais de excitação estocásticos: a) Ruído branco binário; b) Ruído branco gaussiano.	38
3.2	Sinais de espectro ótimo.	39
3.3	Esquema do gerador PRBS de 4-estágios (comprimento máximo de 15 dígitos).	40
3.4	Exemplo de sinal pseudo-aleatório binário.	41
3.5	Exemplo de sinal pseudo-aleatório multi-nível (com supressão de harmônicos múltiplos de 2).	43
3.6	Gerador de sinais MLPRS em Simulink.	44
3.7	Esquema do gerador do sinal de seqüência m multi-nível (Braun et al., 1999).	46
3.8	Espectro de potência após o ZOH.	48
3.9	Modelo de Hammerstein.	50
3.10	Esquema do reator de neutralização de pH	55
3.11	Saída da planta e sinais de excitação para $q = 7$, $n_r = 3$ e $m = 3$	59
3.12	Espectro de potência do sinal MLPRS para $q = 7$, $n_r = 3$ e $m = 3$	59
3.13	Saída do modelo identificado (pontos) e saída da planta (traço contínuo).	61
3.14	Conjunto independente de dados para c_{q1} e q_3 (grafo superior). Validação entre a saída do planta (-) com a saída do modelo identificado (-.) (gráfico inferior).	61
3.15	Vazão de injeção de ar e concentração de oxigênio dissolvido na zona 3.	68

3.16	Espectro de potência do sinal com $q = 7$, $n_r = 2$.	69
3.17	Concentração do oxigênio dissolvido (traço contínuo) e saída do modelo identificado NARX linear (pontos) com o sinal s_1 .	71
3.18	Validação com os modelos identificados linear e não linear.	71
3.19	Validação dos modelos da dinâmica de OD.	73
4.1	Estratégia do controlador preditivo.	84
4.2	Princípio do preditor de identificação preditiva de horizonte estendido (Haber et al., 2003).	92
4.3	Princípio do preditor de múltiplos-passos-à frente (Haber et al., 2003).	95
4.4	Resposta do modelo real (—) e dos preditores LS (...), LRPI (- -) e MSPI (-.-) 10-passos-à frente.	103
4.5	Resposta do modelo real (—) e dos preditores LS (...) e MSPI (- -) 10-passos-à frente e do preditor LRPI (-.-) 40-passos-à frente.	104
4.6	Diagramas de Bode dos preditores LS, LRPI e MSPI 10-passos-à frente.	105
4.7	Relação de erro de predição entre os preditores LRPI e MSPI 20-passos-à frente.	105
4.8	Relação de erro de predição entre os preditores LRPI e MSPI 40-passos-à frente.	106
5.1	Sistema de controle em malha fechada com os polinômios R e S .	116
5.2	Sistema de controle em malha fechada do controlador GPC-MSPI com os polinômios \mathcal{R} , \mathcal{S} e \mathcal{H} .	120
5.3	Esquema do controlador GPC auto-ajustável baseado nos modelos preditores.	121
5.4	Desempenhos do controlador GPC com modelo MSPI e modelo LS, sem ruído aditivo na saída.	123
5.5	Desempenhos do controlador GPC com modelo MSPI e com modelo LRPI, sem ruído aditivo na saída.	124
5.6	Desempenho do controlador GPC com modelo LS, com ruído aditivo na saída.	124
5.7	Desempenho do controlador GPC com modelo LRPI, com ruído aditivo na saída.	125
5.8	Desempenho do controlador GPC com modelo MSPI, com ruído aditivo na saída.	125
6.1	Esquema de controle do set-point de OD para remoção de amônio.	132
6.2	Sinal de excitação do <i>set-point</i> de OD para gerar os dados de identificação da dinâmica entre o <i>set-point</i> de OD e a concentração de amônio. O gráfico superior corresponde à concentração de amônio, o intermediário é o <i>set-point</i> de OD e o inferior é o OD com ruído e atraso no sensor.	134
6.3	Característica de repetição inversa do sinal de excitação para o <i>set-point</i> de OD e a região de frequência relevante para controle deste sinal.	134
6.4	Identificação do modelo da concentração de amônio: saída da planta (—) e saída do modelo identificado (··). <i>Set-point</i> da concentração de amônio em 4 mg N/l.	136
6.5	Validação entre a saída da planta (—) e a predição do modelo (- -) frente a distúrbios no afluente, nas concentrações de substrato e amônio.	136
6.6	Rastreamento das mudanças do <i>set-point</i> da concentração de amônio sem ruído aditivo.	139
6.7	Rastreamento de mudanças do <i>set-point</i> da concentração de amônio com ruído aditivo no sensor.	140

6.8	Controle do <i>set-point</i> de OD. A linha contínua corresponde ao <i>set-point</i> controlado enquanto a linha tracejada corresponde ao <i>set-point</i> fixo de 2 mg O ₂ /l. Os distúrbios nas concentrações de amônio (- -) e de substrato (-) também são mostradas.	140
6.9	Seqüência decrescente de degraus para determinação do espaço entrada-saída.	142
6.10	Sinal de saída (concentração de nitrato) e sinal de entrada (vazão de recirculação interna) para o experimento de identificação.	144
6.11	Validação do modelo identificado e potência espectral indicando a região relevante para controle depois do ZOH.	144
6.12	Validação do modelo identificado com o conjunto independente de dados após o processo de remoção de tendências.	145
6.13	Desempenho dos controladores GPC-LRPI e PI para rejeição de distúrbios na vazão de afluente do esgoto bruto.	148
6.14	Desempenho dos controladores GPC-LRPI e PI para rejeição de distúrbios nas concentrações de substrato e de amônio no afluente.	148
6.15	Desempenho dos controladores GPC-LRPI e PI para rejeição de distúrbios na vazão da fonte externa de carbono.	149
6.16	Rastreamento do <i>set-point</i> da concentração de nitrato dos controladores GPC-LRPI e PI na última zona anóxica.	149
6.17	Rastreamento do <i>set-point</i> da concentração de nitrato dos controladores GPC-LRPI e PI com ruído na saída.	150

Lista de Tabelas

1.1	Escalas de tempo das dinâmicas em uma planta de lodos ativados (Olsson e Newell, 1999).	11
3.1	Condições de operação nominal do sistema (Henson e Seborg, 1997).	56
3.2	Análise do limite de baixa frequência para n_r e q	57
3.3	Número de harmônicos não suprimidos presentes dentro da largura de faixa de interesse com $\gamma = 1/2$	57
3.4	Parâmetros do modelo NARX não linear.	60
3.5	Índices de erro usando o sinal PRBS sem considerar a dinâmica do atuador.	62
3.6	Índices de erro usando sinais PRBS considerando a dinâmica do atuador.	62
3.7	Índices de erro entre o sinal pseudo-aleatório e os sinais aleatórios sem considerar a dinâmica do atuador.	63
3.8	Índices de erro entre o sinal pseudo-aleatório e os sinais aleatórios considerando a dinâmica do atuador.	63
3.9	Análise do limite de baixa frequência para n_r e q	66
3.10	Número de harmônicos não suprimidos para $\gamma = 1/2$	67
3.11	Número de harmônicos não suprimidos para $\gamma = 1/3$	67
3.12	Índices de erro para grau de não linearidade $\ell = 2$	68
3.13	Parâmetros do modelo NARX não linear com o sinal s_1	69
3.14	Índices de erro para grau de não linearidade $\ell = 1$	70
3.15	Parâmetros do modelo NARX linear com o sinal s_1	70
3.16	Parâmetros do modelo NARX não linear (Jerônimo et al., 2000).	72
3.17	Índices de erro para um conjunto independente de dados.	72
5.1	Índices de erro do desempenho em malha fechada dos controladores GPC sem ruído na saída.	126
5.2	Índices de erro do desempenho em malha fechada dos controladores GPC com ruído na saída.	126
6.1	Parâmetros do controlador GPC-LRPI.	138
6.2	Parâmetros do controlador GPC-LRPI.	146

Glossário

AIC - Critério de informação de Akaike

ARX - *Autoregressive and exogeneous input*

ASM1 - Modelo de lodos ativados IAWQ No. 1 (*activated sludge model No. 1*)

ASWWTP-USP - *Activated sludge wastewater treatment plant - University of São Paulo*

BIBO - *Bounded-input, bounded-output*

BIC - Critério de informação de Bayes

CARIMA - *Controller auto-regressive integrated moving-average*

CONAMA - Conselho nacional do meio ambiente

COT - Carbono orgânico total

DBO - Demanda bioquímica de oxigênio

DIB - Sinal binário de intervalo discreto (*discrete interval binary*)

DMC - Controle de Matriz Dinâmica (*Dynamic Matriz Control*)

DQO - Demanda química de oxigênio

EM - Erro máximo

ETE's - Estações de tratamento de esgotos

FAC - Função de autocorrelação

FDIDENT - *Frequency domain identification toolbox*

GMC - Controlador por modelo genérico (*generic model controller*)

GPC - Controlador preditivo generalizado (*generalized predictive control*)

GPC-LRPI - GPC baseado no preditor de identificação preditiva de horizonte estendido

GPC-LS - GPC baseado no preditor de iteração sucessiva padrão

- GPC-MSPI - GPC baseado em identificação do preditor múltiplos-passos-à frente
- GPC-STR - GPC *self-tuning regulator*
- HAB - Sinal binário de Hall (*hall binary*)
- ICA - Tecnologia de instrumentação, controle e automação
- IPRBS - Sinal pseudo-aleatório binário com repetição inversa (*inverse PRBS*)
- IWA - *International water association*
- IWSA - *International water services association*
- LMPC - Controle preditivo linear
- LRPI - Identificação preditiva de horizonte estendido (*long-range predictive identification*)
- LS - Método de mínimos quadrados (*least square method*)
- MGS - Algoritmo ortogonal de Gram-Schmidt modificado
- MLMH - Sinal com multi-harmônicos multi-nível (*multi-level multi-harmonics*)
- MLPRS - Sinal pseudo-aleatório multi-nível (*multi-level pseudo-random signal*)
- MPC - Controlador preditivo baseado em modelo (*model based predictive control*)
- MSPI - Identificação de preditores múltiplos-passos-à frente (*multi-step-ahead prediction identification*)
- MUSMAR - *Multistep multivariable adaptive regulator*
- NARMAX - *Non linear auto-regressive with moving average and exogeneous input*
- NARX - *Non linear autoregressive and exogeneous input*
- NLMPC - Controle preditivo não linear
- NTK - Nitrogênio total Kjeldahl
- PEM - *Prediction error methods*
- PID - Ação proporcional, integral e derivativa
- PLA - Processo de lodos ativados
- PRBS - Sinal pseudo-aleatório binário (*pseudo-random binary signal*)
- QRB - Sinal binário de resíduo quadrático (*quadratic-residue binary*)
- RBB - Ruído branco binário

RBG - Ruído branco gaussiano

RHC - Controle de horizonte retrocedente (*receding horizon control*)

RMS - Somatório do erro quadrático

SOH - Sinal multi-seno (*sum of harmonics*)

TPB - Sinal binário primo twin (*twin-prime binary*)

ZOH - Segurador de ordem zero (*zero-order hold*)

Lista de Símbolos

A_c	- Polinômio contendo os pólos desejados em malha fechada
$\mathcal{A}_k, \mathcal{G}_k$	- Polinômios da resposta livre predita da entrada e saída do GPC-MSPI
a_r	- Coeficiente do polinomial primitivo de grau q igual a 2
b	- Taxa de decaimento da biomassa (h^{-1})
\mathbf{b}	- Vetor com os termos de polarização
b_A	- Taxa de decaimento da biomassa autotrófica (h^{-1})
b_H	- Taxa de decaimento da biomassa heterotrófica (h^{-1})
c_r	- Coeficiente do polinomial primitivo de grau q maior que 2
c_{q1}	- Concentração do fluxo de ácido
d	- Tempo morto ou atraso do vetor de entrada do modelo NARX
\mathbf{d}	- Vetor com os termos para o distúrbio medido
$\mathbf{D}(t)$	- Matriz diagonal no instante t
$\mathcal{D}_{\mathcal{M}}$	- Conjunto de valores admissíveis de θ
D_s	- Deslocamento do sinal para múltiplas entradas
$e(t)$	- Ruído branco com propriedades de média nula e variância σ^2
\mathbf{e}	- Vetor com os termos do erro de modelagem
f	- Índice de desempenho <i>friendliness</i> de um sinal de excitação
\mathbf{f}	- Vetor das predições da resposta livre
$\tilde{\mathbf{f}}$	- Vetor das predições da resposta livre do GPC-MSPI
f_P	- Fração de biomassa produzindo produtos particulados (adimensional)
\mathbf{G}	- Matriz da resposta ao degrau do modelo CARIMA
$G(q^{-1})$	- Função de transferência desde u a y
$G(q^{-1}, \theta)$	- Função de transferência na estrutura do modelo, correspondente ao valor θ
$G_0(q^{-1})$	- Função de transferência "verdadeira" desde u a y
g_i	- Parâmetros da resposta ao degrau do modelo CARIMA
\bar{g}^T	- Vetor fila da matriz ganho GPC
K_h	- Constante de tempo do processo de hidrólise (h)
K_{NH}	- Constante de média-saturação de amônio para biomassa autotrófica ($\text{g NH}_4^+ - \text{N/m}^3$)
K_{NO}	- Constante de média-saturação de nitrato para biomassa heterotrófica desnitrificante ($\text{g NO}_3\text{-N/m}^3$)
$K_{O,A}$	- Constante de média-saturação do oxigênio para biomassa autotrófica ($\text{g O}_2/\text{l}$)
$K_{O,H}$	- Constante de média-saturação de oxigênio para biomassa heterotrófica ($\text{g O}_2/\text{l}$)
$K_{S,I}$	- Constante de média-inibição (g DQO/l)

$H(q^{-1})$	- Função de transferência desde e a y
$H(q^{-1}, \theta)$	- Função de transferência do modelo de ruído correspondente ao valor θ
$H_0(q^{-1})$	- Função de transferência "verdadeira" do modelo de ruído
k_a	- Taxa de amonificação ($\text{m}^3 (\text{g DQO h})^{-1}$)
K_{gpc}	- Matriz ganho do controlador GPC equivalente a $(\mathbf{G}^T \mathbf{G} + \lambda_u \mathbf{I})^{-1} \mathbf{G}^T$
k_h	- Taxa específica máxima de hidrólise ($\text{g biod. DQO} (\text{g DQO h})^{-1}$)
K_{Ia}	- Função de transferência de oxigênio
K_S	- Constante de média-saturação do substrato ($\text{g DQO}/\text{m}^3$)
J_{LRPI}	- Função objetivo da identificação preditiva de horizonte estendido
J_p	- Índice de erro quadrático de rastreamento da referência
J_u	- Índice da magnitude absoluta do incremento da ação de controle
$L(q^{-1})$	- Pré-filtro para o erro de predição
$\mathbf{L}(t)$	- Ganho do estimador
L_{lrpi}	- Pré-filtro de estimador LRPI
m	- Comprimento máximo
m	- Número de níveis de um sinal de seqüência m
$\mathcal{M}(\theta)$	- Conjunto de modelos parametrizados
N_1	- Horizonte de predição mínimo
N_2	- Horizonte de predição máximo
N_c	- Horizonte de controle
N_d	- Atraso máximo para o vetor de distúrbio medido
N_p	- Horizonte de predição, $N_p = N_2 - N_1$
n_r	- Número de estágios do registro de deslocamento
N_s	- Comprimento de um sinal de seqüência de comprimento máximo
n_T	- Número de transições do sinal
N_u	- Atraso máximo para o vetor de entrada
N_y	- Atraso máximo para o vetor de saída
$\#harm$	- Número de harmônicos não suprimidos
$\mathbf{P}(t)$	- Matriz de covariância no instante t
q_1	- Fluxo de de ácido no reator de neutralização
q_2	- Fluxo de buffer no reator de neutralização
q_3	- Fluxo de base no reator de neutralização
Q	- Penalização sobre o erro de rastreamento de predição
$Q_{ar,i}$	- Vazão de injeção de ar na zona aeróbia i (l/h)
Q_e	- Vazão do efluente (m^3/h)
Q_{ext}	- Vazão externa de carbono (l/h)
Q_f	- Vazão de alimentação do sedimentador (m^3/h)
Q_{in}	- Vazão do afluente (m^3/h)
Q_{int}	- Vazão de recirculação interna (m^3/h)
Q_{sl}	- Vazão de reciclo externo (m^3/h)
Q_w	- Vazão de descarte de lodo (m^3/h)
R	- Penalização sobre o desvio da variável de entrada
$r(t)$	- Referência ou <i>set-point</i> do controlador
\mathbf{r}	- Vetor do sinal de referência

$R(q^{-1}), S(q^{-1})$	- Polinômios de projeto do controlador GPC padrão
$\mathcal{R}(q^{-1}), \mathcal{S}(q^{-1}), \mathcal{H}(q^{-1})$	- Polinômios de projeto do controlador GPC-MSPI
R_k	- Relação erro de predição
r_{so}	- Taxa de consumo de oxigênio na zona aeróbia
\mathcal{S}	- "Sistema verdadeiro"
S_{ALK}	- Concentração de alcalinidade solúvel (mmol HCO_3^-/l)
S_h	- Concentração de substrato lentamente biodegradável (mg DQO/l)
S_I	- Concentração de matéria orgânica solúvel não biodegradável (mg DQO/l)
S_{ND}	- Concentração de nitrogênio orgânico solúvel biodegradável (mg N/l)
S_{NH}	- Concentração de amônio (e amônia) solúvel (mg N/l)
S_{NI}	- Concentração de nitrogênio orgânico solúvel não biodegradável (mg N/l)
S_{NO}	- Concentração de nitrato (e nitrito) solúvel (mg N/l)
S_O	- Concentração de oxigênio dissolvido (mg O_2/l)
S_S	- Concentração do substrato rapidamente biodegradável (mg DQO/l)
$S_{S,I}$	- Concentração de matéria inibitória (g DQO/l)
$T(q^{-1})$	- Filtro introduzido como parâmetro do projeto na estratégia LRPI
T_{cyc}	- Tempo de duração de um período do sinal
T_{sw}	- Tempo de comutação do sinal de seqüência m multi-nível
$u(t)$	- Variável de entrada (manipulada)
\mathbf{u}	- Vetor com os termos para a entrada
$\tilde{\mathbf{u}}$	- Vetor de futuros incrementos de controle
$\bar{\mathbf{u}}$	- Vetor restrito de futuros incrementos de controle
$\mathbf{U}(t)$	- Matriz triangular superior com diagonal unitária no instante t
w_s	- Largura de faixa do sinal de excitação
w_{inf}	- Frequência inferior do sinal de excitação
w_{sup}	- Frequência superior do sinal de excitação
$x(t)$	- Média móvel integrada do distúrbio aleatório no instante t
X_B	- Concentração de biomassa ativa (mg DQO/l)
$X_{B,A}$	- Concentração de biomassa autotrófica (mg DQO/l)
$X_{B,H}$	- Concentração de biomassa heterotrófica (mg DQO/l)
X_{NB}	- Concentração de biomassa ativa (mg N/l)
X_{ND}	- Concentração de nitrogênio orgânico particulado biodegradável (mg N/l)
X_{NP}	- Concentração de nitrogênio orgânico particulado não biodegradável ou inerte (mg N/l)
X_I	- Concentração de matéria particulada não biodegradável (mg DQO/l)
X_P	- Concentração de produtos particulados inertes (mg DQO/l)
X_S	- Concentração do substrato lentamente biodegradável (mg DQO/l)
$y(t)$	- Variável de saída
\mathbf{y}	- Vetor com os termos da saída
$\hat{\mathbf{y}}$	- Vetor de saídas controladas previstas
$\hat{y}(t/\theta)$	- modelo preditor o equação do preditor
$\hat{y}_1(t/\hat{\theta}_N)$	- Preditor um-passo-à frente

$\hat{y}_k(t/\hat{\theta}_N)$	-	Preditor k -passos-à frente
\hat{y}_{rl}	-	Predição da resposta livre
$Y_{A,NH}$	-	Coefficiente de produção de biomassa autotrófica observada para remoção de amônio (g cel. formado DQO/g DQO oxidado)
$Y_{A,NO}$	-	Coefficiente de produção de biomassa autotrófica observada para produção de nitrato (g cel. formado DQO/g DQO oxidado)
Y_H	-	Coefficiente de produção de biomassa heterotrófica observada (g cel. formado DQO/g DQO oxidado)
$Y_{H,NO}$	-	Coefficiente de produção de biomassa heterotrófica observada para remoção de nitrato (g cel. formado DQO/g DQO oxidado)
Y_{obs}	-	Coefficiente de produção de biomassa observado
Z^N	-	Conjunto finito de N medições da(s) entrada(s) e saída(s)
Z^{t-1}	-	Conjunto finito de medições da(s) entrada(s) e saída(s) até o instante $t - 1$

Letras gregas

α	-	Vetor dos coeficientes da primeira fila da matriz ganho GPC
$\alpha^{(k)}, \beta^{(k)}, \gamma^{(k)}$	-	Coefficientes do modelo linearizado para o k -ésimo passo de predição
α_s	-	Fator relativo à velocidade de resposta em malha fechada com a resposta em malha aberta
β	-	Matriz da resposta ao degrau para o GPC-MSPI
β_s	-	Fator relativo ao tempo de acomodação do processo
$\bar{\beta}^T$	-	Vetor fila da matriz ganho GPC-MSPI
$\varepsilon_F(t, \theta)$	-	Filtragem do erro de predição
$\varepsilon(t, \theta)$	-	Erro de predição
Δ	-	Operador deslocamento em atraso
$\Delta u(t)$	-	Forma variacional da variável de entrada
ϕ_u	-	Espectro de potência para a entrada
ϕ_v	-	Espectro de potência para o distúrbio
γ	-	Fator relativo ao número de harmônicos suprimidos
Γ_k, F_k	-	Polinômios da resposta livre predita da entrada e saída respectivamente
λ_u	-	Constante positiva que pondera a ação de controle
ℓ	-	Grau de não linearidade do modelo
η_g	-	Fator de correção para crescimento anóxico de heterótrofos (adimensional)
η_h	-	Fator de correção para hidrólise anóxica (adimensional)
θ	-	vetor de parâmetros dos modelos
θ^*	-	vetor de parâmetros do modelo limite
$\varphi(t)$	-	vetor de regressores no instante t
τ_{dom}^c	-	Constante de tempo dominante mais curta
τ_{dom}^l	-	Constante de tempo dominante mais longa
μ_{max}	-	Taxa específica máxima de crescimento da biomassa (h^{-1})
$\mu_{max,H}$	-	Taxa específica máxima de crescimento da biomassa heterotrófica (h^{-1})
$\mu_{max,A}$	-	Taxa específica máxima de crescimento da biomassa autotrófica (h^{-1})
$\xi(t)$	-	Distúrbio aleatório no instante t

Trabalhos Publicados Pelo Autor

1. Lara, J. M. V., Milani, B. E. A. Control of nitrate flow in pre-denitrification systems using identification for long-range predictive control. *The 7th IASTED International Conference on Control and Application, CA 2005*, Cancun, Mexico (article accepted).
2. Lara, J. M. V., Milani, B. E. A. Adaptive predictive control of ammonium level based on dissolved oxygen variable set-point control in an activated sludge process. *The 4th Mercosur Congress on Process Systems Engineering, EMPROMER 2005*, Rio de Janeiro, Brazil (abstract accepted).
3. Lara, J. M. V., Milani, B. E. A. Identificação de modelos para Controle Preditivo: Estudo de Caso (Em Espanhol), *XI Congresso Internacional de Engenharia Eletrônica, Elétrica e de Sistemas, INTERCON 2004*, Lambayeque, Peru.
4. Lara, J. M. V., Milani, B. E. A. Identificação do processo de remoção de nitrogênio em uma planta de tratamento de esgotos usando modelos NARX (Em Espanhol), *XI Congresso Internacional de Engenharia Eletrônica, Elétrica e de Sistemas, INTERCON 2004*, Lambayeque, Peru.
5. Lara, J. M. V., Milani, B. E. A. Identification of Neutralization Process Using Multi-level Pseudo-Random Signals. *Proceedings of the 2003 American Control Conference, ACC 2003*, Colorado, EUA, 2003.
6. Lara, J. M. V., Milani, B. E. A. Projeto de sinais relevantes para controle na identificação de sistemas não lineares (Em Espanhol), *IX Congresso Internacional de Engenharia Eletrônica, Elétrica e de Sistemas, INTERCON 2002*, Lima, Peru.

Capítulo 1

Introdução

1.1 Introdução ao Tratamento de Esgotos Domésticos

O homem é um ser inerentemente social e desde suas origens sempre procurou viver em grupos organizados por varias razões, seja por defesa, busca de alimentos, conquista da natureza, etc. Desde épocas antigas os conglomerados humanos se situaram nas proximidades de fontes de água que permitissem o desenvolvimento de suas atividades, recurso estratégico, que se pensava inesgotável.

No entanto, a compreensão de que ela é vital à sobrevivência existiu entre algumas civilizações antigas, que se preocuparam com a disponibilidade de água tratada e a utilização de redes de esgoto. Na antiga Roma, usava-se a Cloaca Maxima, um conjunto de aquedutos que utilizava água para transportar o esgoto proveniente de latrinas. É surpreendente que a utilização de redes de esgoto fossem esquecidas até o século 19, onde a falta de instalações adequadas para esgotos provocaram grandes epidemias, especialmente naquelas cidades que viviam o auge do comércio e da industrialização.

O desenvolvimento de redes de esgotos municipais marcaram o início dos modernos sistemas de distribuição. Devido à abundância do recurso hídrico e a existência do fenômeno de autodepuração¹, os despejos eram lançados diretamente nos corpos de água (rios e correntes, lagoas, estuários e águas costeiras), convertendo o problema da ameaça à saúde humana em um outro problema até então pouco valorizado: a poluição das águas.

O uso intensivo da água e a poluição resultante não conseguiram acompanhar o ciclo natural de autodepuração. Quando estes despejos são descarregados nos corpos de água, estes servem de alimento a bactérias que decompõem as substâncias complexas do despejo em simples compostos químicos, enquanto consomem oxigênio dissolvido. Quanto mais despejo é agregado, as bactérias se reproduzem rapidamente e a população cresce, consumindo mais oxigênio. Se as descargas de despejo forem muito altas, o crescimento das bactérias é tão grande que o oxigênio dissolvido se esgota dentro dos corpos de água receptores. Um dos principais objetivos dos processos de tratamento de esgotos é manter este material consumidor de oxigênio dentro de níveis adequados para os corpos de água receptores, com mínimo risco para a saúde humana.

¹Restabelecimento do equilíbrio no meio aquático, por mecanismos essencialmente naturais, onde os compostos orgânicos são convertidos em compostos inertes e não prejudiciais do ponto de vista ecológico

1.1.1 Importância do Tratamento de Esgotos

Descargas de esgoto provenientes de redes de distribuição municipais e industriais são uma fonte significativa de problemas na qualidade da água em vários países. Relatórios advertem que descargas municipais são a segunda fonte de deterioração da qualidade da água dos diferentes países. Poluentes associados com descargas municipais incluem nutrientes, bactérias e outros agentes patogênicos, assim como metais e produtos químicos tóxicos, provenientes da indústria.

A falta de tratamento de esgotos e condições de saneamento podem contribuir para a proliferação de inúmeras doenças parasitárias e infecciosas, como a tifoide, cólera, disenterias, hepatite infecciosa e vários casos de verminoses. Outra importante razão para tratar os esgotos é a preservação do meio ambiente. Conforme citado anteriormente, a matéria orgânica presente nos esgotos pode causar a diminuição da concentração de oxigênio dissolvido, provocando a morte de peixes e outros organismos aquáticos, escurecimento da água e exalação de odores desagradáveis.

1.1.2 Panorama atual

A degradação ambiental causada pelo processo de crescimento e desenvolvimento urbano-industrial no Brasil, que se manifesta na contaminação de rios e córregos, em muitos casos acaba impedindo o uso das águas para consumo da população, da própria indústria e da agricultura. Segundo indicadores atuais (IBGE, 2002), apenas 35% dos esgotos urbanos passam por tratamento. O restante é simplesmente lançado em rios e córregos ou em fossas sépticas, com elevada probabilidade de contaminação dos lençóis freáticos. Na região de Campinas o índice de tratamento de esgoto era de 10% até 2004, o que aumentou a carga de poluição orgânica nas três principais microbacias hidrográficas da região: as dos rios Capivari, Atibaia e Quilombo.

Existe também uma nova tendência em tratamento de esgotos, concebida na Conferência Internacional sobre Meio Ambiente, realizada em 1992 no Rio de Janeiro (RIO-92), e materializada em recomendações na Cúpula Mundial sobre Desenvolvimento Sustentável, realizada em 2002, em Johannesburgo (Rio+10). É a *sustentabilidade* do processo. Para conseguir um tratamento de esgotos sustentável, a atenção é direcionada não somente à remoção de matéria orgânica mas também à recuperação de nutrientes e energia, reutilização da água e utilização do lodo tratado. Com as demandas de tecnologias sustentáveis e padrões de lançamento cada vez mais estritas, a complexidade no tratamento de esgotos também tende a aumentar. Como consequência, a tecnologia de controle de processos e a simulação de modelos serão essenciais para a implementação bem sucedida destas novas tecnologias (Spanjers, 2000).

1.1.3 O papel da automação e controle em estações de tratamento de esgotos

Em termos de massa tratada e energia consumida, o processo de tratamento de esgotos domésticos pode ser considerado como a maior indústria de processos. Na Região do Estado de São Paulo, por exemplo, um volume diário médio de $3,0 \times 10^6$ m³ de esgotos são tratados, só nos municípios sob a jurisdição da Companhia de Saneamento Básico do Estado de São Paulo (SABESP, 2002).

São várias as razões que motivam a aplicação de tecnologia de instrumentação, controle e automação (ICA) em estações de tratamento de esgotos (Olsson, 1993):

- **Qualidade do efluente:** há uma crescente preocupação pública pela luta contra a poluição do Meio Ambiente, como a poluição dos cursos de água. Resoluções provenientes de instituições nacionais de administração para a qualidade da água, motivam o desenvolvimento de melhores estações para o tratamento de esgotos. Os padrões para o lançamento de efluentes e o adequado desempenho das estações de tratamento de esgotos devem ser garantidos.
- **Economia:** quiçá o principal incentivo seja a redução de custos de operação, tais como custos de aeração, consumo de energia, adição de produtos químicos e mão de obra. Pode ser demonstrado que a utilização da ICA pode prover menores custos operacionais com maior confiabilidade e segurança, levando ao limite a capacidade operacional da instalação atual.
- **Operação eficiente da planta:** plantas de tratamento biológico de esgotos são realmente complexas, complexidade que tenderá a aumentar à medida que o processo de remoção de nutrientes se torne cada vez mais comum. Assim, a necessidade de uma operação consistente em plantas de tratamento de esgotos resulta ser um requisito indispensável de operação.
- **Disponibilidade de tecnologia:** as freqüentes melhorias nos sensores é por si mesmo um motivo para aplicações de estratégias de controle. No caso do controle avançado, é importante obter modelos com a apropriada complexidade, modelos que poderiam ser atualizados com as medições *on line* do processo para poderem adaptar-se ao comportamento da planta.

O potencial para melhorar o desempenho de estações de tratamento de esgotos através do desenvolvimento e utilização de estratégias de automação e controle ainda não é bem compreendido. Apesar de ter sido abordado há um bom tempo atrás por Andrews (1975), a utilização da ICA é ainda sub-utilizada (Steffens, 1997). Muitas plantas de tratamento de esgotos utilizam atualmente tecnologias de controle muito simples, ou não utilizam todo o potencial do controle automático. Usualmente a automação e controle em plantas de tratamento de esgotos utilizam controle manual com assistência de instrumentação, controle seqüencial ou simplesmente controle proporcional. Eles são aplicados basicamente para monitoramento, medição, registro e controle de equipamentos elétricos, com o objetivo de reduzir o consumo de energia elétrica. Existem poucas aplicações para o controle do processo em si (Freitas et al., 2002), como consequência das dificuldades particulares encontradas em estações de tratamento.

1.2 Estações de Tratamento de Esgotos

As Estações de Tratamento de Esgotos (ETE's) são as instalações onde se desenvolvem os processos biológicos, químicos e as diversas operações físicas, de forma a adequar o lançamento a uma qualidade desejada ou ao padrão de qualidade vigente. Do ponto de vista operacional, o propósito das ETE's é a remoção biológica da matéria orgânica mediante a oxidação da matéria carbonácea. No entanto, devido às condições de clima tropical do Brasil, ocorre também a eventual oxidação em condições aeróbias da matéria nitrogenada, fenômeno denominado nitrificação². Em algumas configurações, é também possível a remoção biológica de nutrientes como o nitrogênio e o fósforo. No presente estudo a remoção biológica de fósforo não é considerada.

²Processo de oxidação da matéria nitrogenada onde a amônia é transformada em nitritos e estes em nitratos.

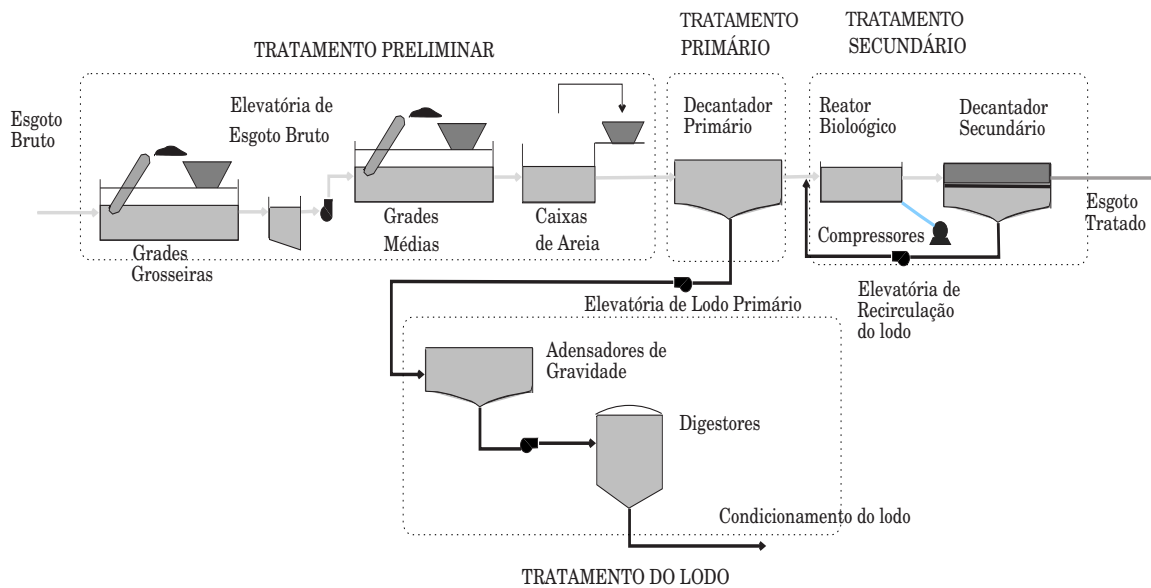


Fig. 1.1: Diagrama típico de uma Estação de Tratamento de Esgotos.

Nas ETE's, o esgoto bruto que chega às estações passa por diversas etapas de tratamento, como as mostradas no diagrama da Figura 1.1. O lodo em excesso do decantador secundário pode ser também conduzido mediante uma elevatória até os Adensadores de Gravidade junto com o lodo primário, linha de fluxo do processo que não está incluída no diagrama. Geralmente, as ETE's são divididas segundo seus níveis de tratamento em:

Tratamento Preliminar : destina-se à remoção de sólidos grosseiros em suspensão, onde materiais de maiores dimensões são retidos em grades e os sólidos decantáveis, como areia e gordura, são sedimentados por gravidade em desarenadores ou caixas de areia. São utilizados apenas mecanismos físicos e tem por finalidade proteger as unidades de tratamento subseqüentes, os dispositivos de transporte dos esgotos (como bombas e tubulações).

Tratamento Primário : composto geralmente de decantadores primários, visando a remoção de sólidos sedimentáveis que são adensados gradualmente no poço de lodo do decantador. Parte significativa destes sólidos em suspensão é composta pela matéria orgânica em suspensão. Predominam também os mecanismos físicos como forma de tratamento.

Tratamento Secundário : geralmente constituído pelo reator biológico, decantador secundário e sistema de recirculação. O principal objetivo do tratamento secundário é a remoção da matéria orgânica mediante a inclusão de uma etapa biológica, onde ocorrem reações bioquímicas, realizadas por microrganismos. Acelera o mecanismo de autodepuração em condições controladas.

Tratamento Terciário : nem sempre presente, é geralmente constituído de unidade de tratamento físico-químico, que tem por finalidade a remoção complementar da matéria orgânica, dos nutrientes, de poluentes específicos e da desinfecção dos esgotos tratados.

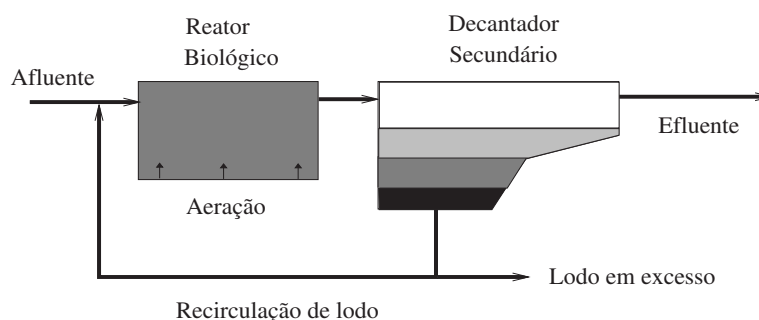


Fig. 1.2: Sistema de lodos ativados básico.

Tratamento de Lodo : são tratados os subprodutos sólidos gerados nas diversas unidades, com a finalidade de aumentar o teor de sólidos do lodo e de estabilizar total ou parcialmente as substâncias instáveis e matéria orgânica presentes no lodo fresco. É composto geralmente por adensadores e digestores anaeróbios.

1.3 Processo de Lodos Ativados

Dentre os processos de tratamento secundários, o Processo de Lodos Ativados (PLA) é o processo mais utilizado para o tratamento de esgotos domésticos e industriais. O PLA é utilizado em situações onde é necessária uma elevada qualidade de efluente com requisitos de área reduzidos. No entanto, requer um maior controle operacional devido ao emprego de uma quantidade maior de equipamentos com consumo também maior de energia elétrica, se comparado com outros tipos de tratamento secundários.

No sistema básico da Figura 1.2, as principais variáveis a controlar são: aeração, lodo em excesso e recirculação de lodo. É recomendável o controle destas variáveis para o tratamento eficiente do processo, garantindo a quantidade suficiente de lodo no sistema.

Os principais componentes do processo de lodos ativados são:

- **Reator biológico:** remove grande parte da matéria orgânica, mediante reações bioquímicas, realizadas por microrganismos aeróbios, podendo eventualmente remover nutrientes. Reproduzem e aceleram os fenômenos naturais de estabilização da matéria orgânica que ocorreria num corpo receptor.
- **Decantador secundário:** é responsável pela separação dos sólidos em suspensão do lodo que vem do reator biológico e pela sedimentação gradual dos sólidos em suspensão no fundo do decantador. O lodo acumulado no fundo é formado, em sua maior parte, por bactérias ainda ativas com capacidade de assimilar matéria orgânica. Este processo de separação permite a saída de um efluente clarificado.
- **Sistema de recirculação:** permite o retorno de uma parte do lodo sedimentado, rico em microrganismos e de maior concentração (*lodo ativado*) ao reator biológico. O retorno do lodo é necessário para suprir o reator biológico com uma quantidade suficiente de microrganismos, para decompor com maior eficiência o material orgânico.

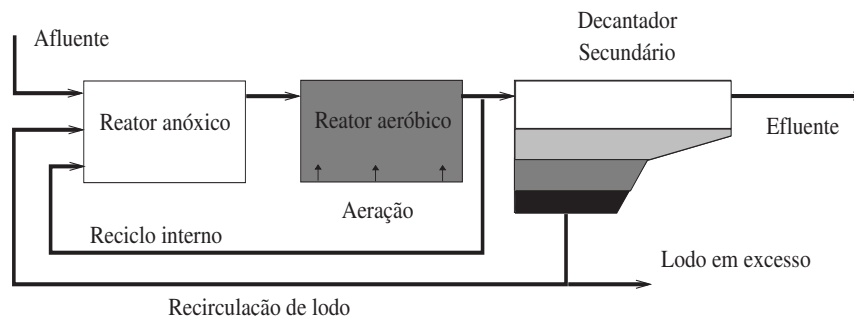


Fig. 1.3: Sistema de lodos ativados com pré-desnitrificação.

O processo de lodos ativados pode ser usado também para a remoção de matéria nitrogenada. Este processo é mais complicado, pois requer condições tanto aeróbias quanto anóxicas, onde, em lugar do oxigênio dissolvido, o nitrato é usado como agente oxidante. Uma das configurações de planta mais conhecidas, para a remoção de matéria nitrogenada, é o *sistema com pré-desnitrificação*, visto na Figura 1.3. Para satisfazer a demanda de matéria orgânica biodegradável no processo de desnitrificação, é adicionado um reator biológico desnitrificante, que utiliza a matéria orgânica do afluente. Para garantir a presença de nitrato neste reator, o lodo proveniente do reator biológico nitrificante é reciclado através do sistema de reciclo interno. O reator desnitrificante é mantido anóxico, enquanto no reator nitrificante é adicionado ar. Nesta configuração, os principais elementos a controlar são a aeração, o reciclo interno, o lodo em excesso e a recirculação de lodo.

1.3.1 Modelo Matemático do PLA

Existem diversos modelos que descrevem os processos biológicos em estações de tratamento por lodos ativados. De todos eles, o mais empregado é o Modelo de Lodos Ativados IAWQ³ No. 1, denominado também pelas siglas ASM1, descreve os processos biológicos de remoção de carbono, nitrificação e desnitrificação (Henze et al., 1987a; Henze et al., 1987b). Apesar deste modelo ter sido estendido, considerando a remoção de fósforo (Henze et al., 1995), a atividade desnitrificante dos organismos acumuladores de fósforo (Henze et al., 1999), o consumo de oxigênio e a produção de lodo (Gujer et al., 1999), o modelo original resulta sendo o mais conhecido e utilizado pela comunidade científica e industrial (Benchiheub, 1997).

Portanto, o ASM1 é o modelo adotado no presente trabalho, sendo que este é também adotado pelo simulador *benchmark* da planta de tratamento de lodos ativados, como explicado na seção seguinte.

1.3.2 Benchmark do PLA

Existe na atualidade um grande número de simuladores para processos de tratamento de esgotos, especializados para diferentes áreas de aplicação e plataformas, tanto no meio industrial, como

³International Association on Water Quality, que junto com a IWSA (International Water Services Association) são as predecessoras da atual IWA (International Water Association)

no acadêmico, com diversos preços e formas. Um estudo dos diversos simuladores para processos biológicos de tratamento de esgotos pode ser encontrado no manual de Copp (2000).

O simulador *benchmark* utilizado neste trabalho é o ASWWTP-USP (Sotomayor et al., 2001), implementado em MATLAB (MATrix LABORatory) usando o ambiente de simulação Simulink⁴, que representa o processo de lodos ativados numa configuração de fluxo contínuo com pré-desnitrificação para remoção de matéria orgânica e nitrogênio de esgotos domésticos. Este simulador *benchmark* foi escolhido por ser baseado no modelo ASM1, amplamente aceito e por complementar este modelo com o modelo do decantador secundário de múltiplas camadas, proposto por Takács et al. (1991), comumente utilizado em simuladores para plantas de lodos ativados.

1.4 Dificuldades e Problemas no Controle de PLA

Algumas características que tornam o processo de lodos ativados difícil de controlar são listadas a seguir (Lindberg, 1997):

- Variante no tempo: é um processo biológico onde temperatura, composição do afluente, população de biomassa e fluxos variam com o tempo.
- Não linear: os mecanismos biológicos que descrevem o crescimento, geração e morte dos microrganismos e a multiplicação de estados possuem características não lineares.
- Dinâmicas do tipo *stiff*: as constantes de tempo pertencem a uma ampla gama de respostas temporais, com intervalos de tempo de segundos a meses.
- Carência de sensores *on line* para algumas variáveis de processo, sendo que nem todas as variáveis envolvidas podem ser medidas diretamente.
- Múltiplas entradas e múltiplas saídas.
- Ações de controle são limitadas pelas restrições físicas de bombas, válvulas e outros atuadores.
- Dificuldade de incorporar modelos de processos complexos em algoritmos de controle avançado e pouca efetividade de estratégias de controle baseadas em modelos de processo muito simplificados.

1.4.1 Problemas de Observabilidade e Controlabilidade

O problema de observabilidade em plantas de tratamento de lodos ativados dificulta o desenvolvimento de projetos de controle, porque muitas variáveis de estado e parâmetros do processo não são diretamente mensuráveis ou observáveis. O número de variáveis de saída (medições) é muito menor que o número de variáveis de estado, indicando que o sistema pode ser não observável (Olsson e Newell, 1999).

⁴MATLAB® e Simulink® são produtos registrados de The MathWorks, Inc., 3 Apple Hill Drive, Natick, MA 01760-2098, URL: www.mathworks.com.

Também o número de variáveis manipuladas (agente de controle) é menor que o número de saídas (variável a ser controlada), que é menor ainda que o número de estados (Olsson e Newell, 1999), indicando que o sistema pode ser não controlável. Existe também a falta de atuadores suficientemente flexíveis e acionáveis que ajudem a rejeitar os distúrbios (Olsson, 1993; Steffens, 1997; Jeppsson et al., 2002).

1.4.2 Distúrbios

Plantas de tratamento estão sujeitas a um grande número de distúrbios externos e internos. Estes distúrbios podem ser classificados em três classes segundo a frequência com que se apresentam (Ingildsen, 2002)

- Distúrbios sazonais, presentes na carga, no fluxo, na temperatura da água, na taxa de nitrificação, etc. Os requisitos de medição podem variar entre um dia a uma semana.
- Distúrbios diurnos na concentração de poluentes e no fluxo. Em alguns casos, não é possível rejeitar completamente distúrbios diurnos, pois distúrbios no afluente apresentam variações de amplitude grandes demais para que o atuador possa manipular. As medições devem ser feitas no intervalo de minutos a horas.
- Distúrbios eventuais, tais como chuvas, eventos tóxicos e picos na carga do afluente. Requerem medições rápidas e mesmo assim os distúrbios são de tal magnitude que ações de controle normais não são suficientes para manipular estes distúrbios.

1.4.3 Incertezas

Em processos de tratamento de esgotos existem também incertezas: incertezas na precisão das medições, incertezas na determinação dos parâmetros do sistema, onde alguns são variantes no tempo e dependentes da temperatura da água, e incertezas originadas pela impossibilidade de medir estados não acessíveis. Obviamente estas incertezas afetam o desempenho global do controlador a ser projetado.

1.5 Modelagem para controle do PLA

Sem dúvida, o ASM1 é uma ferramenta valiosa para o entendimento do processo e de suas interações em sistemas de lodos ativados, pois permite a simulação e a predição do comportamento integral do processo para análise operacional e dinâmica.

No entanto, o modelo contém 19 parâmetros (14 parâmetros cinéticos e 5 parâmetros estequiométricos). Muitos destes parâmetros são variantes no tempo e alguns deles podem variar consideravelmente em um período de tempo limitado. Fatores como a configuração da planta, condições de operação, dinâmicas da população de microrganismos, graus de inibição por compostos tóxicos, composição do afluente, temperatura, pH, etc., afetam os valores destes parâmetros (Jeppsson, 1996). Obviamente este modelo não é apropriado para o controle baseado em modelo, devido aos problemas de identificação deste grande número de parâmetros.

Se a finalidade principal do modelo é o controle, então a necessidade da formulação de modelos simples ou reduzidos é evidente. Muitos conceitos da teoria de controle são aplicáveis só em modelos de baixa ordem. Modelos de grande dimensão resultam em maiores requisitos computacionais, problemas de mal condicionamento e frequentemente em problemas numéricos do tipo *stiff*, devido à interação de dinâmicas lentas e rápidas.

Assim no caso em estudo, estratégias tradicionais de controle, baseadas em realimentação, possivelmente não sejam suficientes. Métodos baseados em controle *feed-forward*, adaptativo ou preditivo são adequados para este propósito. No entanto, todos estes métodos requerem modelos de processo que sejam simples em sua estrutura, robustos, unicamente identificáveis e possíveis de atualizar *on-line*. Desta forma, pode-se concluir que modelos de ordem reduzida devem ser utilizados para o projeto do controlador (Jeppsson, 1996; Weijers, 2000).

Grande ênfase tem sido dada à obtenção de modelos reduzidos do ASM1, com o objetivo de simplificar o cálculo da ação de controle. Em Weijers (2000) é feita uma revisão de diversas abordagens de redução de modelos para o ASM1, classificando-as segundo a quantidade de conhecimento embutido no modelo reduzido.

Dentre as abordagens para a redução de modelos, tem-se a identificação caixa preta, que utiliza uma pequena quantidade de conhecimento *a priori* e pode estimar os parâmetros do modelo reduzido com base nos dados entrada-saída. A identificação linear caixa preta pode ser aplicada a modelos com não linearidades moderadas ao redor de um ponto de operação. Para construir modelos reduzidos de sistemas não lineares, com validade em uma faixa ampla de operação, devem ser aplicadas técnicas de identificação não lineares.

O projeto de sinais de excitação, para redução do modelo através da identificação, constitui uma etapa muito importante (Weijers, 2000). Mediante a escolha apropriada das faixas de frequência dos sinais de excitação, os modelos caixa preta podem ser obtidos para enfatizar regiões de frequência de interesse para o controlador.

1.5.1 Sinais de Excitação Relevantes para Controle

Sinais de excitação podem ser projetados para enfatizar a região de frequência de relevância para o controle do processo e ao mesmo tempo para aproximar dinâmicas lineares ou não lineares do processo. Uma outra característica que pode ser inserida no sinal de excitação, que é útil para plantas de tratamentos de esgotos, é a escolha de uma forma e amplitude do sinal compatíveis com as restrições físicas dos atuadores.

Esta será a abordagem empregada neste trabalho, a redução de modelos dos processos de remoção de matéria orgânica numa planta de tratamento de esgotos por lodos ativados, mediante a identificação de modelos usando técnicas de projeto de sinais relevantes para controle. Para isto, sinais de excitação pseudo-aleatórios multi-nível serão projetados para testar o sistema em diferentes pontos de operação, possibilitando a identificação de dinâmicas não lineares ou a identificação de um comportamento linear em presença de não linearidades, mediante a especificação do conteúdo de harmônicos, conforme procedimento proposto em Barker e Zhuang (1997a). Esta abordagem permite a identificação de modelos diferentes, segundo o objetivo e grau de complexidade. Assim estes modelos podem servir para controle baseado em modelo e serem simples o suficiente para serem atualizados *on-line* via medições do processo. Outra aplicação dos modelos reduzidos consiste no estudo e entendimento do processo, onde seja interessante explorar o comportamento dinâmico de determinados fenôme-

nos, tais como a dinâmica da concentração de oxigênio dissolvido, da concentração de nitrato ou da concentração de amônio na remoção de nitrogênio.

1.5.2 Identificação de Modelos para Controle Preditivo

Se o objetivo do exercício de modelagem, mediante técnicas de identificação de sistemas, é o bom desempenho das predições do modelo para o projeto do controle preditivo baseado em modelo (MPC), então modelos de predição, associados a modelos um-passo-à frente, não são recomendados. É sabido que estes modelos de predição não dão boas predições para vários passos-à frente no caso de haver dinâmicas não modeladas ou presença de ruídos de medição. O Controlador Preditivo Generalizado (GPC), introduzido por Clarke et al. (1987), que pertence à sub-classe de controladores preditivos, atualiza os parâmetros do modelo do controlador mediante o algoritmo de mínimos quadrados recursivos. Neste caso, em presença de incertezas de modelagem entre a planta e o modelo, como as que existem quando são usados modelos reduzidos, o algoritmo de mínimos quadrados não fornecerá as melhores predições k -passos-à frente.

A primeira tentativa de utilizar predições k -passos-à frente, para identificar modelos a serem usados no controle preditivo, foi feita por Shook et al. (Shook et al., 1991; Shook et al., 1992), abordagem denominada identificação preditiva de horizonte estendido. No procedimento de identificação, na abordagem de Shook et al. (1991), é empregada uma função custo que considera uma média dos erros de predição ao longo de um horizonte de predição. Isto implica que o mesmo modelo será usado para prever as saídas do sistema verdadeiro em todos os estágios do horizonte de predição. Outra tendência, que resulta em modelos de predição múltiplos, foi inspirada na abordagem do controlador MUSMAR (Greco et al., 1984). Esta abordagem é denominada identificação de preditores múltiplos-passos-à frente, onde são identificados modelos de predição, um para cada passo de predição. Isto remove o problema da média dos erros de predição do trabalho de Shook et al. (1991).

Com o objetivo de modelar preditores que resultem num melhor desempenho do controlador preditivo, neste trabalho será realizado um estudo comparativo entre o modelo de predição para vários passos-à frente padrão, o modelo de predição de identificação preditiva de horizonte estendido e o modelo de predição de múltiplos-passos-à frente. O controlador preditivo adotado para a análise de desempenho em malha fechada é o Controlador Preditivo Generalizado, escolhido porque esta classe de controladores preditivos é apropriada para processos que apresentam não linearidades moderadas, com comportamento variante no tempo (Camacho e Bordons, 1995), que são características dos mecanismos biológicos existentes no processo de lodos ativados.

1.6 O MPC Aplicado à Remoção Biológica de Nitrogênio

Como visto, o processo de lodos ativados é um processo multivariável, não linear e que apresenta dinâmicas variadas com interações entre suas variáveis. Para obter um controle eficiente da remoção da matéria nitrogenada na planta de lodos ativados com pré-desnitrificação, têm sido propostas diferentes estratégias. Uma delas é a denominada estratégia hierárquica, que considera as velocidades das dinâmicas dos processos biológicos envolvidos no processo de remoção de nitrogênio (ver Tabela 1.1). Em consequência, a planta de lodos ativados pode ser vista como unidades relativamente independentes segundo a velocidade de resposta das dinâmicas.

Tab. 1.1: Escalas de tempo das dinâmicas em uma planta de lodos ativados (Olsson e Newell, 1999).

Velocidade	Processo envolvido
Lento (dias - semanas)	Crescimento de biomassa
Médio (horas - dias)	Dinâmicas de concentração Remoção de nutrientes
Rápida (minutos - horas)	Dinâmicas de vazão Oxigênio dissolvido

Usando a premissa de "dividir e conquistar", o processo de remoção de nitrogênio pode ser considerado como o controle na zona anóxica e controle na zona aeróbia. Com esta hipótese é possível melhorar significativamente o desempenho operacional de plantas de tratamento de esgotos através de estruturas de controle locais, relativamente simples, com controladores baseados em medições das concentrações dos componentes orgânicos. Esta idéia é alentada pelo contínuo melhoramento dos sensores para plantas de tratamento de esgotos, tais como os de medição de oxigênio dissolvido, de amônio e de nitrato. Esta é uma tendência evidenciada em congressos, revistas especializadas, teses (Jeppsson et al., 2002; Vrečko et al., 2002; Vrečko et al., 2003; Singman, 1999; Rehnström, 2000) e em implementações em plantas de tratamento de esgotos em escala real (Ingildsen, 2002). A implementação de controles locais do tipo PI, tem sido efetuada por esses autores no controle do processo de desnitrificação, usando controle da concentração de nitrato na zona anóxica e no controle do processo de nitrificação, usando o controle da concentração de amônio no reator aeróbio. Ingildsen (2002) relata uma diminuição do custo de aeração entre 20 e 30%.

O GPC é um forte candidato para o controle da remoção de nitrogênio dentro da estratégia hierárquica de plantas de tratamento de esgoto. Unifica em sua formulação as vantagens do controle adaptativo e do controle de horizonte estendido no referente à robustez perante distúrbios quando escolhido adequadamente seu observador polinomial T (Clarke e Mohtadi, 1989). Mas para a aplicação do GPC torna-se indispensável obter modelos de adequada complexidade das dinâmicas da concentração de amônio e de nitrato, cuja estrutura possibilite a atualização dos parâmetros a partir de medições *on-line* e que seja eficaz em predizer o futuro comportamento destas concentrações para períodos de tempo adequados. Por outro lado, existe a necessidade no projeto do controlador de desenvolver um procedimento sistemático para reduzir os modelos dos processos de nitrificação e desnitrificação, considerando que os mesmos sejam simultaneamente dependentes dos objetivos de controle e da informação *a priori* disponível, como ressaltado por Steffens et al. (1997).

Em consequência, para o controle do processo de remoção de nitrogênio, nesta tese é proposto a decomposição do mesmo no controle do processo de nitrificação na zona aeróbia e no controle do processo de desnitrificação na zona anóxica, tomando em consideração a atenuação dos acoplamentos existentes entre estes dois processos. Desta forma, os processos serão tratados como duas malhas de controle independentes, utilizando o controlador preditivo generalizado.

1.7 Objetivos da tese

O objetivo principal desta tese é o desenvolvimento de procedimentos para identificação de modelos lineares reduzidos, com características apropriadas para a melhoria do desempenho do controle preditivo generalizado, relativo à redução de incertezas de modelagem e à redução da variância do erro do preditor k -passos-à frente padrão, visando a aplicação particular de controle do processo de remoção de nitrogênio em uma planta de tratamento de esgotos por lodos ativados com pré-desnitrificação.

Para este fim são delineados a seguir os objetivos secundários:

- Estudo de técnicas e procedimentos de geração de sinais de excitação relevantes para controle. Como mencionado, o projeto de sinais de excitação para redução de modelos via identificação constitui uma etapa muito importante do experimento. Portanto, para a realização deste objetivo, o estudo é direcionado ao desenvolvimento de sinais de excitação que testem o sistema não linear em diferentes pontos de operação, de forma que possibilitem a identificação do modelo reduzido linear dos processos não lineares e ao mesmo tempo permitam a especificação das faixas de frequência de interesse para o projeto do controlador.
- Modelar preditores que resultem num melhor desempenho do controlador preditivo. Neste trabalho, será realizado um estudo comparativo entre o modelo de previsão por iteração sucessiva padrão e outros modelos preditores que reduzem a propagação do erro de previsão. A redução dos efeitos de incertezas do modelo reduzido na síntese do controlador preditivo orientará a seleção do modelo preditor.
- Utilizar a decomposição do controle do processo de remoção de nitrogênio no controle do processo de nitrificação, mediante o controle da concentração de amônio no último reator aeróbio e no controle do processo de desnitrificação, mediante o controle da concentração de nitrato na zona anóxica. Esta opção é suportada pela factibilidade de desacoplamento dos processos mediante a introdução de condições operacionais apropriadas e a consideração das constantes de tempo das dinâmicas dos processos.
- Verificar que a abordagem de controle do processo de remoção de nitrogênio, mediante controladores preditivos generalizados locais, aplicados aos processos de nitrificação e desnitrificação como duas malhas independentes, é uma alternativa válida frente a abordagens de controle multivariáveis do processo de remoção de nitrogênio. Neste sentido, os controladores locais deverão ser testados em sua capacidade de robustez para rejeitar distúrbios externos e em sua eficiência no rastreamento da referência.

1.7.1 Restrições

Em vista dos objetivos propostos e da complexidade que eles envolvem, são listados a seguir algumas restrições assumidas para a simplificação da abordagem de identificação e controle do processo de remoção de nitrogênio:

- Os modelos usados no projeto do controlador preditivo são discretos, lineares e invariantes no tempo.

- Para a formulação do controlador preditivo generalizado não são incluídas restrições nas variáveis de entrada nem de saída.
- A identificação dos modelos reduzidos será realizada *off-line* através de um algoritmo ortogonal ótimo de seleção de estrutura e estimação de parâmetros.
- Os modelos reduzidos usados na simulação para a identificação e síntese dos controladores são modelos de uma entrada e uma saída (SISO).

1.8 Estrutura da Tese

O desenvolvimento do presente trabalho está estruturado nos seguintes capítulos:

Capítulo 1: É exposta a relevância da área de tratamento de esgotos. As dificuldades e problemas no controle de plantas de lodos ativados são listadas, permitindo enunciar os objetivos e os métodos propostos para tratar o problema de controle.

Capítulo 2: São apresentados os mecanismos biológicos que governam os principais processos de oxidação da matéria orgânica, nitrificação e desnitrificação, e suas interações em plantas de lodos ativados. Também é mostrado o modelo de simulação a ser utilizado no presente estudo.

Capítulo 3: Aborda o procedimento para a síntese de sinais de identificação relevantes para controle, com forma apropriada para os atuadores, baseado em diretrizes que utilizam informação disponível da planta. A eficácia da metodologia é ilustrada através da aplicação em um processo de neutralização de pH. Finalmente a metodologia é aplicada também na obtenção de modelos parcimoniosos do processo de oxigênio dissolvido da planta de lodos ativados.

Capítulo 4: É apresentado um estudo de modelagem, via identificação, de preditores orientados ao projeto de controladores preditivos baseados em modelo. São analisadas duas técnicas, a identificação preditiva com múltiplos-passos-à frente e a identificação preditiva de horizonte estendido.

Capítulo 5: A formulação padrão do controlador preditivo generalizado é revista para facilitar o entendimento da modificação do GPC baseado no modelo preditor múltiplos-passos-à frente. A seguir é feita uma análise comparativa do desempenho dos controladores baseados nos modelos preditores, obtidos via as seguintes formas de identificação: iteração sucessiva padrão, horizonte estendido e múltiplos-passos-à frente.

Capítulo 6: Trata da aplicação do controlador preditivo, baseado na identificação relevante para controle de modelos preditores, no processo de remoção biológica de nitrogênio. São usadas duas estratégias: controle do processo de nitrificação na zona aeróbia e controle do processo de desnitrificação na zona anóxica.

Capítulo 7: Enumera as contribuições e as conclusões do presente trabalho, obtidas do estudo e implementação do controlador preditivo, baseado em identificação de modelos relevantes para controle, aplicado no modelo *benchmark* ASWWTP-USP. Em seguida, são listados alguns tópicos para futuros trabalhos.

Capítulo 2

Modelo do Processo de Lodos Ativados e Modelo de Simulação

Projetos de supervisão, monitoramento, controle e otimização de plantas de tratamento de lodos ativados, não podem ser desenvolvidos eficientemente sem um modelo que caracterize os fenômenos complexos próprios do processo de lodos ativados. Obviamente não pode ser controlado nem modelado o que não pode ser entendido. Partindo deste postulado é que se torna indispensável a apresentação dos mecanismos biológicos que governam os principais processos de oxidação da matéria orgânica, nitrificação e desnitrificação e de suas interações no modelo de lodos ativados IAWQ No. 1, modelo este utilizado como planta em nossos estudos de identificação e controle.

Sendo assim, na primeira seção deste capítulo são discutidos os mecanismos biológicos presentes em uma planta de tratamento de lodos ativados, tais como hidrólise, crescimento e decaimento da biomassa. A interação destes mecanismos com a biomassa nos processos de remoção da matéria carbonácea e nitrogenada são expostos nas duas seções seguintes, incluindo uma explicação dos processos de nitrificação e desnitrificação. A seção seguinte apresenta as hipóteses de simplificação do modelo IAWQ No. 1 junto com os processos dinâmicos que são descritos no modelo. Na última seção é apresentado o modelo de simulação *benchmark* ASWWTP-USP (Sotomayor et al., 2001), que é usado no presente trabalho para os experimentos de identificação e controle, incluindo as equações que descrevem as taxas de reação do processo. Testes dinâmicos e padrões de validação do modelo não são apresentados, mas podem ser encontrados nessa referência.

2.1 Mecanismos Biológicos

Nas plantas de lodos ativados existe uma interação de diversos mecanismos, sendo que alguns deles ocorrem de forma simultânea e outros de forma seqüencial. Certamente estes mecanismos começam a atuar no próprio sistema de coleta e interceptação de esgotos, mas o máximo de atividade se desenvolve na planta de tratamento.

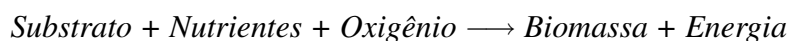
Os principais organismos envolvidos no processo de tratamento de lodos são as bactérias, fungos, algas e protozoários. Destes, as bactérias são as mais importantes na estabilização da matéria orgânica, porém, ao longo do presente trabalho, o termo bactéria denotará genericamente os microrganismos presentes no processo de lodos ativados, representando assim a biomassa. O gênero pre-

dominante de bactérias nos lodos ativados é determinado principalmente pela composição do esgoto bruto, o projeto da planta e as operações próprias de cada planta (Carstensen, 1994).

2.1.1 Microrganismos e suas atividades

Para que possam desempenhar as atividades de crescimento, locomoção, reprodução e outras, os microrganismos necessitam basicamente de energia, carbono e nutrientes (nitrogênio, fósforo, enxofre, potássio, cálcio, magnésio, etc).

No crescimento celular é utilizado o substrato (meio solubilizado no qual os microrganismos crescem) e os nutrientes, encontrados fora da membrana celular (produtos denominados exógenos). O processo é descrito em forma simplificada como:



Em termos de fonte de carbono, há fundamentalmente dois tipos de organismos: os denominados *heterótrofos*, cuja fonte de carbono é o carbono orgânico, e os *autótrofos*, cuja fonte de carbono é o carbono inorgânico (CO₂).

A maior parte das bactérias nos lodos ativados são heterótrofas e desenvolvem a função de remoção da matéria orgânica carbonácea ou remoção da Demanda Química de Oxigênio (DQO), definida no Apêndice A), apresentando uma alta taxa de crescimento celular, se comparada com a taxa de crescimento das bactérias autótrofas, cuja função é principalmente a remoção da matéria nitrogenada.

Os microrganismos absorvem os substratos e nutrientes muito mais rápido do que eles podem utilizar, produtos que não podem ser acumulados em grandes quantidades pelos microrganismos. Em lugar disso, os substratos e nutrientes são quimicamente modificados em pequenos grupos de moléculas (polisacáridos, lípidos e polifosfatos), que podem ser armazenados por um prolongado período de tempo sem um consumo de energia significativa.

A utilização do substrato disponível no meio pelas bactérias se dá em função do tamanho relativo das partículas dispersas a serem utilizadas. Basicamente, pode-se considerar as seguintes duas principais frações de matéria orgânica segundo sua biodegradabilidade

- Fração rapidamente biodegradável,
- Fração lentamente biodegradável.

Num esgoto doméstico típico, a maior parte da matéria orgânica na forma solúvel é rapidamente biodegradável (Von Sperling, 1996). O ciclo fundamental de produção e geração de biomassa é mostrado na Figura 2.1. Note como os processos se desenvolvem alguns em forma paralela e outros em forma seqüencial e como o esgoto bruto provê o substrato rapidamente biodegradável, o substrato lentamente biodegradável, assim como também os microrganismos nele presentes. Naturalmente produtos inertes estão presentes no esgoto bruto mas estes não sofrem alterações pela atuação dos mecanismos biológicos de hidrólise, crescimento e decaimento.

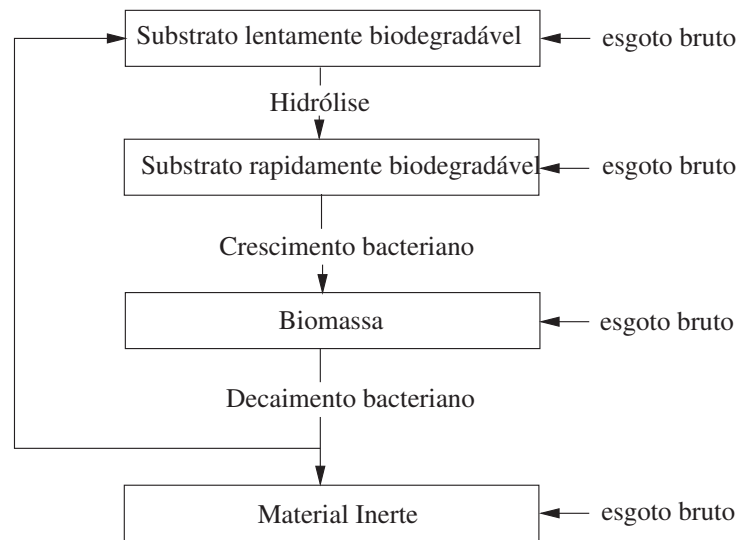


Fig. 2.1: Mecanismos biológicos do processo de lodos ativados (Carstensen, 1994).

2.1.2 Hidrólise

Nos microrganismos existem dois tipos de enzimas que suportam o processo de metabolismo celular: as endoenzimas, que auxiliam os microrganismos no consumo de matéria orgânica solúvel dentro da célula e as exoenzimas, que são lançadas ao meio externo.

Os compostos orgânicos de maiores dimensões ou lentamente biodegradáveis sofrem uma conversão para moléculas menores mediante uma reação catalisada pelas exoenzimas, de forma que possam ser absorvidos pela membrana celular. Esta reação enzimática é denominada *hidrólise*. O produto final da hidrólise se apresenta na forma rapidamente biodegradável, para ser consumido de forma similar à matéria solúvel dentro da célula.

A taxa do processo de hidrólise é lenta, se comparada com a taxa de crescimento da biomassa, sendo ela um fator limitante para o crescimento, principalmente quando o esgoto bruto consiste em moléculas grandes.

A hidrólise é um termo genérico para um grande número de processos bioquímicos diferentes, mas a taxa total do processo pode ser vista geralmente como uma expressão com cinética de primeira ordem:

$$\frac{dS_h}{dt} = K_h \cdot S_h \quad (2.1)$$

sendo S_h a concentração do substrato lentamente biodegradável e K_h a constante de tempo do processo. Uma expressão mais ampla, considerando os fatores limitantes é mostrada em Henze et al. (1987a).

2.1.3 Crescimento da biomassa

O crescimento da biomassa ocorre por fissão dos microrganismos e pode acontecer em poucos minutos sob condições favoráveis, sendo o tempo requerido para cada fissão denominado tempo de geração. Portanto, o crescimento é função do número de microrganismos (massa ou população) num dado instante.

A taxa de crescimento da biomassa é influenciada pelo nutriente limitante ou pelas concentrações de substrato, relação que pode ser descrita usando a função empírica de Monod (Monod, 1949). Assim, a influência da concentração de um nutriente limitante é caracterizada pela seguinte expressão conhecida como taxa de crescimento da biomassa:

$$\frac{dX_B}{dt} = \mu_{max} \frac{S_n}{S_n + K_s} \cdot X_B \quad (2.2)$$

sendo S_n a concentração de nutriente limitante, X_B a concentração de biomassa ativa (aquela responsável pela remoção biológica da matéria), μ_{max} a taxa de crescimento específica máxima da biomassa e K_s a concentração de saturação em que o nutriente limitante atinge $\mu = \frac{1}{2}\mu_{max}$ (Arceivala, 1981).

Diversos substratos limitantes para a taxa de crescimento podem ser modelados ao multiplicar o lado direito de (2.2) com o número apropriado de frações de Monod do substrato limitante, oxigênio ou concentração de nutrientes.

O crescimento da biomassa é relacionado à proporção de biomassa produzida, ΔX_B , e o nutriente ao substrato removido, $-\Delta S$. Este crescimento está caracterizado pelo coeficiente denominado produção de biomassa observada, Y_{obs} :

$$Y_{obs} = -\frac{\Delta X_B}{\Delta S} \quad (2.3)$$

Portanto, se S_n na relação (2.2) corresponde à concentração de nutriente limitante, então a taxa de consumo de nutrientes será modelada segundo a seguinte relação de Monod:

$$\frac{dS_n}{dt} = -\frac{\mu_{max}}{Y_{obs}} \cdot \frac{S_n}{S_n + K_s} \cdot X_B \quad (2.4)$$

Se vários nutrientes e substratos são usados para o crescimento dos microrganismos, o consumo de um determinado nutriente ou substrato é determinado dividindo a taxa de crescimento dos microrganismos com o coeficiente de produção observado individualmente para aquele nutriente ou substrato.

2.1.4 Decaimento da biomassa

Biomassa é perdida mediante decaimento, que incorpora um grande número de mecanismos incluindo metabolismo endógeno, morte, predação e lisis. O decaimento bacteriano é a transformação da biomassa ativa em substrato lentamente biodegradável, como ilustrado na Figura 2.1. Parte do decaimento bacteriano é considerado inerte, devido ao processo da hidrólise ser muito lento comparado

ao tempo de retenção do lodo em uma ETE típica (Carstensen, 1994).

O decaimento de biomassa é geralmente descrito como um processo com cinética de primeira ordem:

$$\frac{dX_B}{dt} = -b \cdot X_B \quad (2.5)$$

sendo b a taxa de decaimento ($b > 0$). A taxa de decaimento é assumida ser independente de fatores ambientais como a temperatura e das concentrações de oxigênio, de nutrientes e de substratos.

2.1.5 Fatores ambientais que afetam o crescimento de biomassa

Além de uma deficiência de nutrientes, o crescimento de biomassa pode ser inibido ou reduzido pelos seguintes fatores ambientais:

- Temperatura,
- pH,
- variações nas concentrações de substratos e nutrientes limitantes, e
- substâncias tóxicas ou inibitórias.

Tais fatores interferem ou inibem o sistema de enzimas ou o metabolismo geral da célula microbiana. Uma forma de modelar a influência destes fatores é definir a taxa de crescimento específica máxima como uma função dos mesmos (Carstensen, 1994):

$$\mu_{max} = f(T, pH, S_x, S_y, \dots) \quad (2.6)$$

onde T representa a temperatura do lodo ativado, pH mede a atividade do íon H^+ , S_x e S_y representam as concentrações de matéria limitante, tóxica ou inibitórias.

O efeito da temperatura no valor de μ_{max} para os organismos nitrificantes é significativa. Assim, para cada acréscimo de aproximadamente 7°C na temperatura, a taxa de crescimento dobra e, inversamente, cada queda de 7°C implica na redução da taxa de crescimento à metade (Von Sperling, 2000). De forma similar, a temperatura afeta a taxa de crescimento das bactérias desnitrificantes, estando a ocorrência da reação de desnitrificação em uma ampla faixa de temperatura, de 0 a 50°C . Embora a temperatura ótima fique em torno de 35 a 50°C .

O crescimento de biomassa é também sensível ao valor de pH no esgoto. Assim, este valor deveria estar entre valores de pH de 5 a 10. É importante notar que a própria nitrificação é responsável pelo decréscimo do pH. A bactéria heterotrófica consegue seu maior μ_{max} para valores de pH entre 7 e 9, enquanto as bactérias autotróficas preferem um meio alcalino ligeiramente maior, dentro de valores de pH 8 a 9.

Substâncias tóxicas podem inibir seriamente o crescimento das bactérias autotróficas nitrificantes, principalmente *Nitrosomonas*, que são as mais sensíveis. Não é o caso das bactérias heterotróficas

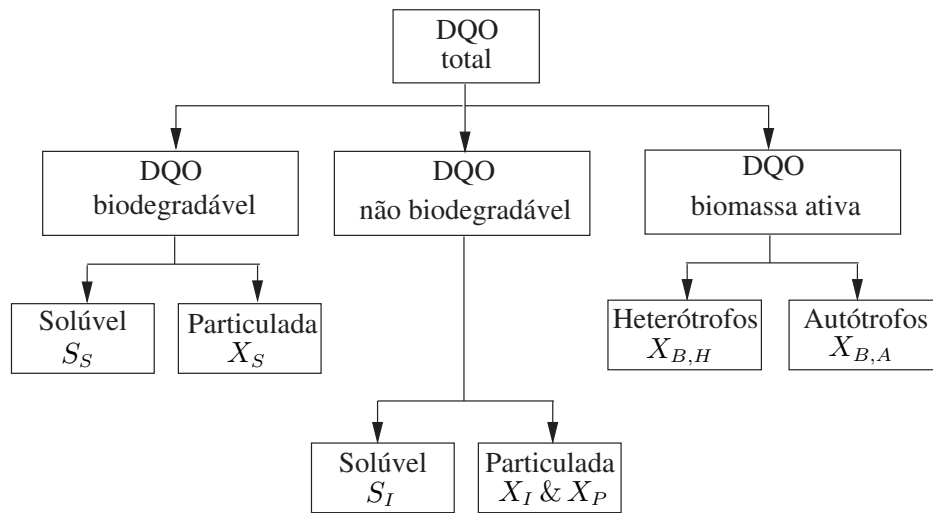


Fig. 2.2: Divisão da matéria orgânica carbonácea segundo sua biodegradabilidade.

responsáveis pela desnitrificação, pois encontram-se presentes em uma maior diversidade de espécies, o que reduz o impacto de algum agente inibitório específico. No entanto, caso haja a presença de substâncias tóxicas ou inibitórias, é bem provável que a desnitrificação seja bastante reduzida (ou anulada), pela simples razão da nitrificação ter sido inibida.

A inibição de uma determinada matéria pode ser modelada multiplicando a expressão da taxa de crescimento pela seguinte fração:

$$\frac{K_{S,I}}{K_{S,I} + S_{S,I}} \quad (2.7)$$

onde $S_{S,I}$ é a concentração de material inibitório e $K_{S,I}$ é uma constante de média-inibição apropriada.

2.2 Processo Aeróbio de Remoção de Matéria Carbonácea

Em estações de tratamento de esgotos é conveniente utilizar a divisão da matéria orgânica, estabelecida em Henze et al. (1987a), como suporte para apresentar os principais conceitos e mecanismos relacionados à interação da biomassa com o processo de remoção da matéria carbonácea. Neste contexto se adota a DQO como variável representativa do substrato orgânico, assim como a nomenclatura (S) para denotar as concentrações dos componentes solúveis e (X) para denotar as concentrações dos componentes particulados.

2.2.1 Divisão da matéria carbonácea

A matéria carbonácea presente nos esgotos é dividida quanto à sua biodegradabilidade em DQO biodegradável, DQO não biodegradável e biomassa, como mostrado na Figura 2.2.

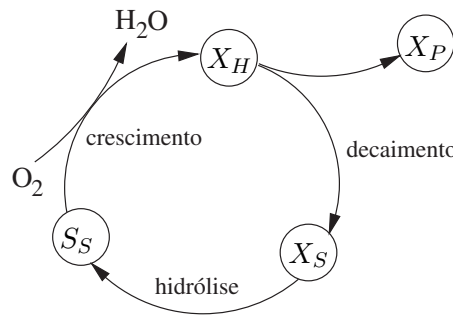


Fig. 2.3: Crescimento heterotrófico segundo a hipótese de morte-regeneração.

1. Matéria orgânica biodegradável: sofre alterações na sua passagem pelo sistema de tratamento e pode ser identificada em duas frações:
 - Matéria orgânica rapidamente biodegradável (S_S): apresenta-se usualmente na forma solúvel, onde moléculas simples podem ser rapidamente absorvidas pelas bactérias heterotróficas e metabolizadas para crescimento e fonte de energia.
 - Matéria orgânica lentamente biodegradável (X_S): apresenta-se na forma particulada e consiste de moléculas relativamente complexas que para serem usadas para metabolismo pelas bactérias, precisam passar antes pelo processo de hidrólise.
2. Matéria orgânica não biodegradável ou inerte: assume-se que passa pelo sistema de tratamento inalterada em sua forma, identificando-se também duas frações, em função de seu estado físico:
 - Matéria orgânica não biodegradável solúvel (S_I): não sofre transformações, deixando o sistema na mesma concentração em que entrou.
 - Matéria orgânica não biodegradável particulada (X_I): é envolvida pela biomassa, sendo removida conjuntamente com o lodo tratado.
3. Biomassa ativa: incluída na divisão devido à presença de microrganismos consumidores de oxigênio e também devido à hipótese da morte-regeneração, onde apesar de todo o substrato ter sido oxidado, ainda exista demanda de oxigênio para crescimento sobre o substrato originado pelos mecanismos de decaimento e hidrólise, como mostrado na Figura 2.3. A biomassa ativa por sua vez divide-se em:
 - Biomassa heterotrófica ($X_{B,H}$): utiliza o carbono orgânico como fonte de carbono e se desenvolve em condições aeróbias ou anóxicas.
 - Biomassa autotrófica ($X_{B,A}$): sua fonte de carbono é o gás carbônico (CO_2) e se desenvolve em condições anóxicas.

Note que é incluída a variável extra X_P , para modelar os produtos particulados inertes que resultam do decaimento da biomassa. A quantidade de nova biomassa formada desde o substrato liberado é sempre menor que a quantidade perdida.

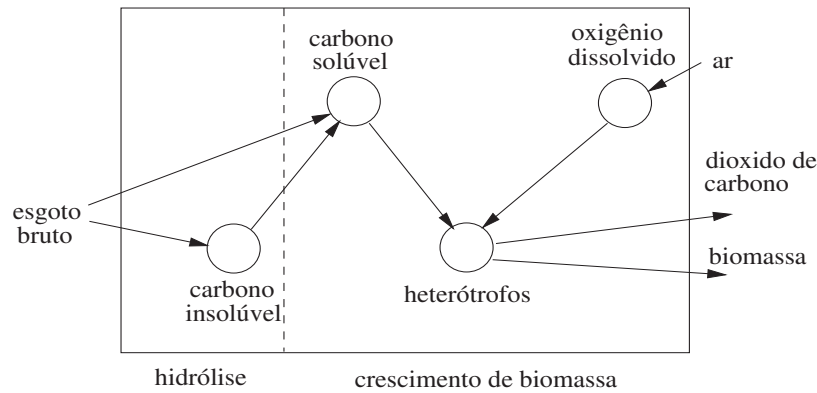


Fig. 2.4: Principais reações na remoção de carbono orgânico.

Pelo exposto acima, assume-se que a matéria carbonácea utilizada diretamente pelas bactérias esteja presente na forma solúvel, enquanto o carbono orgânico insolúvel deva ser primeiro convertido em carbono orgânico solúvel mediante o processo de hidrólise.

A remoção de carbono orgânico solúvel ocorre principalmente pela atuação das bactérias heterotróficas mediante crescimento aeróbio. Neste processo o carbono orgânico é reduzido a biomassa adicional ou a dióxido de carbono em forma solúvel e gasosa, e a outra fração do carbono orgânico é oxidado para prover energia.

Todo esse processo pode ser observado no esquema da Figura 2.4 e corresponde à seguinte reação bioquímica:



A reação mostrada em (2.8) é simplificada, pois na verdade outros componentes estão presentes. É o caso do amônio, que também é removido por causa do crescimento da biomassa heterotrófica (Jeppsson, 1996), embora em menor proporção em relação à remoção da matéria carbonácea. Pode-se notar também a conversão da matéria orgânica em produtos inertes, como gás carbônico e água, conseguindo com isto o objetivo de estabilização da matéria orgânica.

Note que na reação mostrada em (2.8) as concentrações do próprio substrato e do oxigênio impõem limitações ao processo de crescimento das bactérias heterotróficas. Desta forma, a taxa de consumo de substrato orgânico, rapidamente biodegradável sob condições aeróbias, é modelada pela seguinte relação de Monod:

$$\frac{dS_S}{dt} = -\frac{\mu_{max,H}}{Y_H} \cdot \frac{S_S}{S_S + K_S} \cdot \frac{S_O}{S_O + K_{O,H}} \cdot X_{B,H} \quad (2.9)$$

sendo:

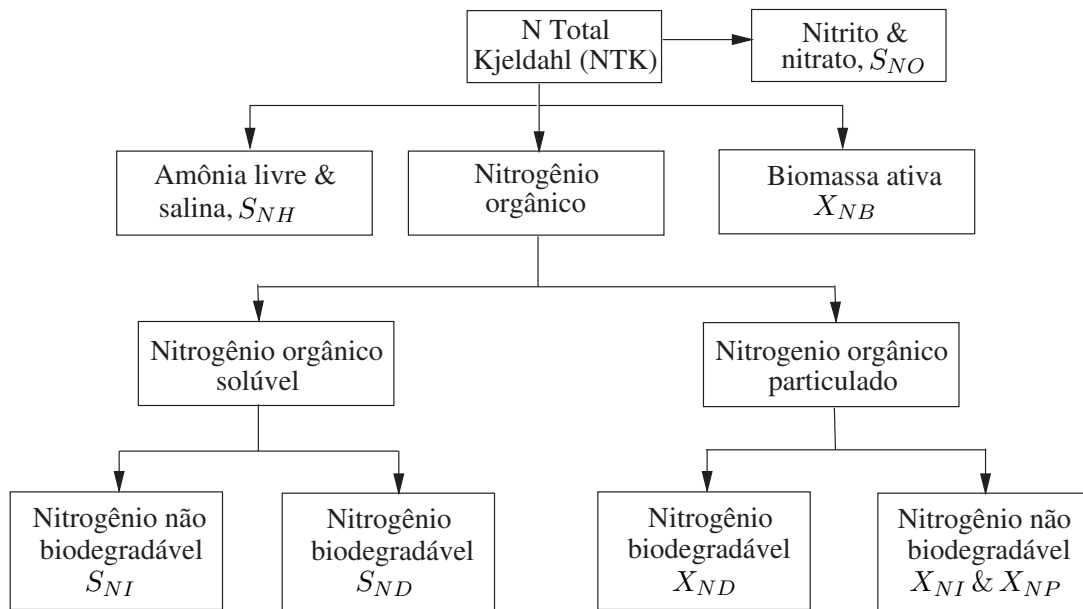


Fig. 2.5: Divisão da matéria nitrogenada segundo a medição do NTK.

- S_S = concentração de substrato rapidamente biodegradável.
 S_O = concentração de oxigênio dissolvido.
 $X_{B,H}$ = concentração de biomassa heterotrófica ativa.
 $\mu_{max,H}$ = taxa de crescimento específica máxima de biomassa heterotrófica.
 Y_H = coeficiente de produção de biomassa heterotrófica observada.
 $K_S, K_{O,H}$ = constantes de média-saturação apropriadas.

No caso de existência de nutrientes que imponham limitações à taxa de crescimento de bactérias heterotróficas durante condições aeróbias, as frações apropriadas de Monod são multiplicadas ao lado direito de (2.9).

2.3 Processo de Remoção de Matéria Nitrogenada

A matéria nitrogenada presente em um processo de tratamento de esgotos divide-se conforme à medição do Nitrogênio Total Kjeldahl¹ (NTK). Portanto, adotamos esta divisão em lugar de utilizar como base a DQO, que resulta sendo útil como indicador da matéria carbonácea oxidada (Jeppsson, 1996).

2.3.1 Divisão da matéria nitrogenada

Baseado no NTK, a matéria nitrogenada é dividida em nitrogênio orgânico, amônio, biomassa ativa, nitrito e nitrato (Jeppsson, 1996), tal como mostrado na Figura 2.5.

¹Ver definição no Apêndice A.

1. Nitrogênio orgânico: em forma similar à divisão da matéria carbonácea, divide-se em duas frações segundo sua biodegradabilidade
 - Nitrogênio orgânico solúvel: que por sua vez se divide em:
 - Nitrogênio orgânico solúvel não biodegradável (S_{NI}): esta fração é usualmente desprezível e pode ser desconsiderada.
 - Nitrogênio orgânico solúvel biodegradável (S_{ND}): presente em forma de compostos nitrogenados e compreende a matéria nitrogenada rapidamente biodegradável. É convertida pelas bactérias heterotróficas em amônia, através do processo de amonificação.
 - Nitrogênio orgânico particulado: que por sua vez se divide em duas frações:
 - Nitrogênio orgânico particulado biodegradável (X_{ND}): é convertida à forma solúvel (rapidamente biodegradável) através da hidrólise.
 - Nitrogênio orgânico particulado não biodegradável ou inerte: tal fração é associada à matéria orgânica carbonácea não biodegradável, sendo envolvida pela biomassa e removida com o lodo excedente. Reúne a matéria particulada orgânica inerte (X_{NI}) e os produtos particulados orgânicos inertes (X_{NP}).
2. Amônio livre e salino (S_{NH}): pode estar presente tanto na forma livre (NH_3), quanto na forma ionizada (NH_4^+), segundo o seguinte equilíbrio dinâmico:



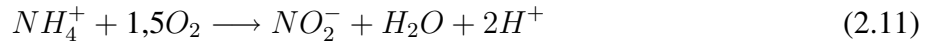
O amônio é resultado dos processos de hidrólise e amonificação descritos acima e também ingressa no esgoto afluente, dado que já no sistema de coleta e interceptação tem início estes processos. É utilizado pelas bactérias heterotróficas e autotróficas. Para o desenvolvimento do presente trabalho assume-se somente a existência do amônio na forma ionizada (íon amônio ou amônio), por ser esta a forma predominante na faixa de pH próxima à neutralidade.

3. Biomassa ativa (X_{NB}): é a fração da biomassa que é assumida ser nitrogênio e é incluída na divisão porque o decaimento de biomassa conduzirá à produção de nitrogênio orgânico biodegradável particulado, X_{ND} .
4. Nitrito e nitrato: produzidos no processo de nitrificação e são representadas numa única variável, S_{NO} .

2.3.2 Processo de Nitrificação

A nitrificação é um processo microbiológico de dois passos, onde amônio é transformado em nitrito e subseqüentemente este é transformado em nitrato. Neste processo a energia é obtida através da oxidação da amônio solúvel, de onde obtém-se o nutriente para as atividades metabólicas de crescimento da biomassa para um grupo especial de bactérias autotróficas, denominadas *nitrificadoras*. Estas bactérias utilizam o gás carbônico como principal fonte de carbono.

Se amônio é usado só como fonte de energia, o primeiro passo para a oxidação em nitrito (NO_2^-) é realizada através das bactérias autótrofas do gênero *Nitrosomonas*, conforme a seguinte reação:



Note que neste processo há uma grande quantidade de alcalinidade consumida, pois para cada mol de NH_4^+ que está sendo oxidado se produzem 2 moles de H^+ . Por esta razão algumas ETE's requerem a adição de algum agente alcalinizante, para manter assim o pH em níveis adequados (Jeppsson, 1996). Níveis baixos de pH diminuem a taxa de nitrificação (taxa de conversão de amônio a nitrato), provocando outros problemas como a corrosão de tubulações e aparição de efluentes agressivos.

O segundo passo na oxidação de nitritos em nitratos é:



Esta reação dá-se principalmente pela ativação de bactérias autótrofas, como as do gênero *Nitrobacter*. Em climas tropicais como o Brasil, a nitrificação é facilitada pelas temperaturas elevadas, ocorrendo quase que sistematicamente em sistemas de lodos ativados convencionais (Von Sperling, 2000).

Para fins de modelagem em processos de lodos ativados, o processo de conversão a nitratos é considerado como um processo de um passo. Sendo assim, os três componentes do lado esquerdo nas reações (2.11) e (2.12) limitarão o crescimento das bactérias nitrificantes. Mas para fins práticos, a taxa de remoção de amônio por nitrificação considera somente as concentrações de amônio e de oxigênio como substratos limitantes:

$$\frac{dS_{NH}}{dt} = -\frac{\mu_{max,A}}{Y_{A,NH}} \cdot \frac{S_O}{S_O + K_{O,A}} \cdot \frac{S_{NH}}{S_{NH} + K_{NH}} \cdot X_{B,A} \quad (2.13)$$

sendo

S_{NH}	=	concentração de amônio.
S_O	=	concentração de oxigênio dissolvido.
$X_{B,A}$	=	concentração de bactéria autotrófica ativa.
$\mu_{max,A}$	=	taxa de crescimento específica máxima da bactéria autotrófica.
$Y_{A,NH}$	=	coeficiente de produção de biomassa autotrófica observada.
$K_{NH}, K_{O,A}$	=	constantes de meia-saturação apropriadas.

Como no caso da remoção de amônio, a taxa de produção de nitratos também é considerada como um processo de um passo, obtida a partir da seguinte relação:

$$\frac{dS_{NO}}{dt} = \frac{\mu_{max,A}}{Y_{A,NO}} \cdot \frac{S_O}{S_O + K_{O,A}} \cdot \frac{S_{NH}}{S_{NH} + K_{NH}} \cdot X_{B,A} \quad (2.14)$$

Note que em ambas reações há o consumo de oxigênio livre, consumo que é geralmente referido como demanda nitrogenada.

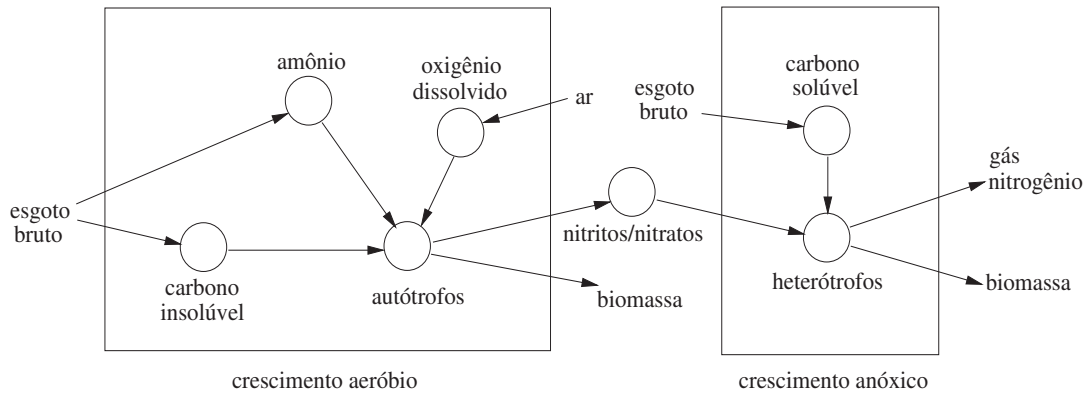
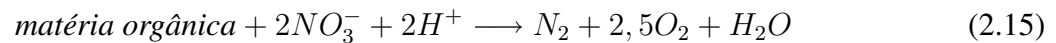


Fig. 2.6: Principais reações no processo de nitrificação e desnitrificação na remoção de matéria nitrogenada.

2.3.3 Processo de Desnitrificação

A desnitrificação é um processo microbiológico heterotrófico onde o nitrato é reduzido a nitrogênio gasoso. Utiliza o nitrato como agente oxidante em lugar de oxigênio. As condições em que este processo ocorre são denominadas anóxicas, devido à ausência de oxigênio e a que algumas bactérias heterotróficas são facultadas usar nitrato para oxidação.

O mecanismo de desnitrificação pode ser descrito mediante a reação microbial geral seguinte (Von Sperling, 2000):



Note que há economia de oxigênio dado que a matéria orgânica pode ser estabilizada na ausência de oxigênio, assim como consumo de H^+ implicando na economia de alcalinidade e no aumento da capacidade tampão do meio.

Em forma estrita, são três as concentrações limitantes para o crescimento das bactérias heterotróficas desnitrificantes: as concentrações de substrato rapidamente biodegradável, de nitrato e de amônio. No entanto, a quantidade requerida de amônio para crescimento celular é muito pouca porque a bactéria heterotrófica é capaz de usar nitrato na falta de amônio para este fim. Sendo assim, a concentração de amônio impõe uma limitação relativamente pequena, pelo que é desconsiderada. Em consequência, para a taxa de remoção de nitrato por desnitrificação, a seguinte expressão cinética é utilizada:

$$\frac{dS_{NO}}{dt} = -\frac{\mu_{max,H}}{Y_{H,NO}} \cdot \frac{S_S}{S_S + K_S} \cdot \frac{S_{NO}}{S_{NO} + K_{NO}} \cdot X_{B,H} \quad (2.16)$$

S_{NO}	=	concentração de nitrato.
S_S	=	concentração de substrato rapidamente biodegradável.
$X_{B,H}$	=	concentração de biomassa heterotrófica ativa.
$\mu_{max,H}$	=	taxa de crescimento específica máxima da bactéria heterotrófica.
$Y_{H,NO}$	=	coeficiente de produção de biomassa heterotrófica observada.
K_S, K_{NO}	=	constantes de meia-saturação apropriadas.

Uma fonte de carbono é também essencial para a ocorrência da desnitrificação. Pode ser em forma de carbono inteiramente disponível no esgoto, ou artificialmente adicionada na forma de metano ou qualquer outra forma rica em carbono. O processo total de remoção de matéria nitrogenada é esquematizada na Figura 2.6.

2.4 O Modelo IAWQ No. 1

O modelo do processo de lodos ativados IAWQ No. 1 (Henze et al., 1987a) é o resultado do trabalho de um grupo internacional cujo principal objetivo foi fazer a revisão de modelos existentes e encontrar um modelo matemático simples, mas que realize previsões realistas do desempenho dos processos biológicos de oxidação de carbono, nitrificação e desnitrificação em plantas de tratamento de esgotos.

Dentre os conceitos incorporados no modelo IAWQ No. 1 estão as hipóteses do substrato rapidamente e lentamente biodegradável e da morte e regeneração da biomassa. Introduz também funções de comutação modeladas como sendo funções do tipo Monod (Monod, 1949), funções matematicamente contínuas, permitindo estabilidade numérica nas simulações. Uma grande vantagem deste modelo é a apresentação matricial dos modelos cinéticos dos processos biológicos envolvidos, facilitando a leitura e entendimento destes processos.

O modelo IAWQ No. 1 é formado por 13 componentes ou variáveis de estado e por 19 parâmetros que descrevem 8 processos biológicos. As variáveis envolvidas no processo de lodos ativados que são consideradas pelo modelo são as concentrações do substrato rapidamente biodegradável (S_S), do nitrogênio orgânico biodegradável solúvel (S_{ND}), do nitrogênio-amônio (S_{NH}), de nitrato (S_{NO}), de oxigênio dissolvido (S_O), da alcalinidade (S_{ALK}), da biomassa heterotrófica ativa ($X_{B,H}$), da biomassa autotrófica ativa ($X_{B,A}$), de nitrogênio orgânico biodegradável particulado (X_{ND}), de produtos particulados por decaimento de biomassa (X_P), de substrato lentamente biodegradável (X_S) e da matéria orgânica inerte solúvel e particulada (S_I e X_I).

Processos Dinâmicos

Os diferentes processos incorporados no modelo IAWQ No. 1 são mostrados em forma esquemática na Figura 2.7, adaptada de Olsson (2002a). Cada um dos processos considerados no modelo estão denotados no esquema com números, descritos a seguir:

1. Crescimento aeróbio de heterótrofos: considera somente as concentrações de S_S e S_O como limitantes porque somente o substrato para crescimento aeróbio é assumido como rapidamente biodegradável. É aplicada a cinética de Monod. O amônio é usado como fonte de nitrogênio para síntese celular e também é assimilado dentro das células.

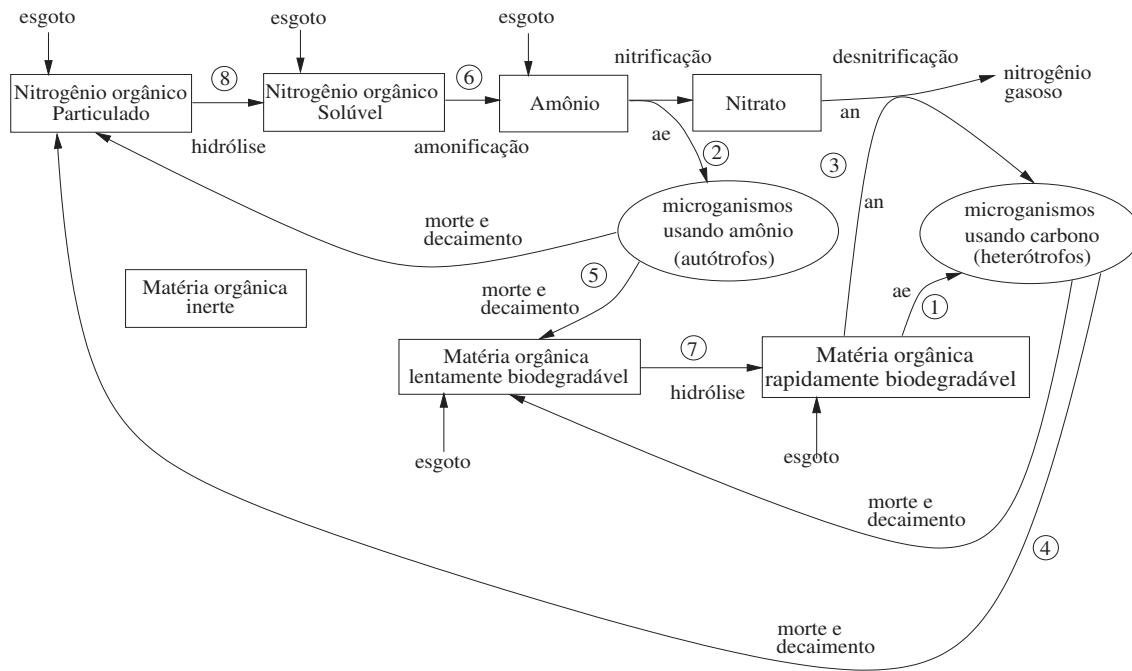


Fig. 2.7: Processos biológicos de oxidação de carbono, nitrificação e desnitrificação considerados pelo modelo IAWQ No. 1: (ae) denota o processo aeróbico e (an) denota o processo anóxico.

2. Crescimento aeróbico de autótrofos: usa amônio solúvel como fonte de energia para o crescimento dos nitrificadores, cuja demanda de oxigênio está relacionada à quantidade de amônio-nitrogênio oxidada. Considera a S_{NH} e S_O como concentrações limitantes. A taxa de crescimento também é modelada usando a cinética de Monod.
3. Crescimento anóxico de heterótrofos: neste crescimento são utilizadas as concentrações de S_S e S_{NO} como concentrações limitantes. S_{NO} é considerada por ser o nitrato o agente oxidante. Biomassa heterotrófica é formada e gás nitrogênio é produzido. É introduzido o fator η_g para corrigir a reduzida taxa de crescimento dos heterótrofos capazes de utilizar nitrato. A inibição de crescimento anóxico na presença de oxigênio é modelada pela função $K_{OH}/(K_{OH} + S_O)$. Amônio é utilizado como fonte de nitrogênio para síntese celular provocando mudanças na alcalinidade do meio.
4. Decaimento de heterótrofos: este processo é modelado conforme a hipótese de morte e regeneração. Decaimento converte biomassa em matéria orgânica lentamente biodegradável X_S , em nitrogênio orgânico particulado X_{ND} e em produtos particulados X_P . Assume-se que não existe perda de DQO porque não são utilizados agentes oxidantes.
5. Decaimento de autótrofos: o decaimento dos autótrofos é modelado da mesma forma que o decaimento dos heterótrofos.
6. Amonificação de nitrogênio orgânico solúvel: nitrogênio orgânico solúvel é convertido em amônio através de uma reação de primeira ordem mediada por heterótrofos ativos. Há consumo

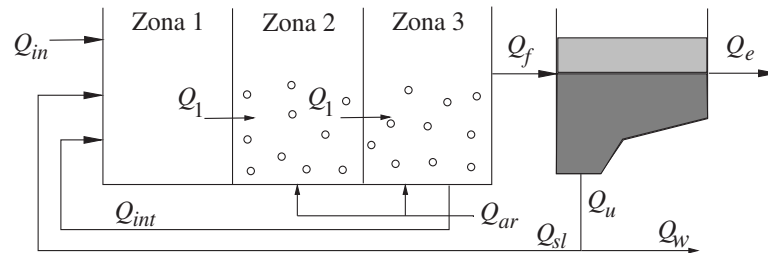


Fig. 2.8: Esquema do modelo *benchmark* de simulação ASWWTP-USP.

de íons hidrogênio, porém um eventual aumento de alcalinidade.

7. Hidrólise de matéria carbonácea particulada: substrato lentamente biodegradável, X_S é convertido em substrato rapidamente biodegradável, S_S . Hidrólise de partículas orgânicas ocorre somente em condições aeróbicas e anóxicas, sendo que em condições anóxicas a taxa de hidrólise é reduzida por um fator $\eta_h < 1$, resultando numa taxa menor que a de condições aeróbicas.
8. Hidrólise de nitrogênio orgânico particulado: esta taxa de hidrólise é proporcional à taxa de hidrólise do substrato lentamente biodegradável, porque o nitrogênio orgânico particulado é assumido ser uniformemente distribuído através do substrato lentamente biodegradável.

No diagrama da Figura 2.7 a matéria orgânica inerte recebe as contribuições do decaimento de biomassa heterotrófica, X_P e autotrófica, X_{NP} e X_{NI} . No entanto estas duas últimas não estão incluídas na representação matricial do modelo.

2.5 Modelo de Simulação

Como visto na seção 1.3.2, o modelo de simulação utilizado no presente trabalho é o modelo denominado modelo *benchmark* de simulação ASWWTP-USP (Sotomayor et al., 2001). A configuração do processo considerado neste modelo consiste de um bioreator em série com um decantador secundário, como observado na Figura 2.8.

O bioreator é composto por uma zona anóxica (zona 1 com volume de 13 m^3) e duas zonas aeróbicas (zona 2 e zona 3, com volumes de 18 m^3 e 20 m^3 respectivamente), enquanto que o decantador secundário tem um volume de 20 m^3 . Em condições de operação nominal e de estado estacionário, a vazão de afluente de esgoto é $Q_{in} = 4,17 \text{ m}^3/\text{h}$ com um valor médio de DQO total de 284 mg/l e nitrogênio total de $45,88 \text{ mg/l}$. O tempo de retenção hidráulica é de 17 horas (baseado no volume total, bioreator mais o decantador secundário). A vazão de recirculação interna ou retorno de licor misto (lodo mais água tratada, vazão rica em nitrato que vem da zona 3 à zona 1 do bioreator) é $Q_{int} = 2Q_{in}$. A vazão de recirculação externa ou recirculação de lodo (vazão de lodo que vem do decantador secundário à zona 1 do bioreator) é $Q_{sl} = 0,5Q_{in}$. A vazão de descarte de lodo é $Q_w = 25,8 \text{ l/h}$. As vazões de injeção de ar para a zona 2 e zona 3 são $Q_{ar} = 44 \text{ l/h}$ e $Q_{ar_3} = 33 \text{ l/h}$, respectivamente. Na zona 1, por ser uma zona anóxica, não há aeração.

Modelo do bioreator

Cada zona do bioreator é baseada no modelo IAWQ No. 1. As equações diferenciais que resultam do modelo correspondem às taxas de reação do processo para cada um dos 13 componentes considerados. A terminologia usada é conforme a utilizada em Henze et al. (1987a).

1. Concentração de biomassa heterotrófica: é afetada por três diferentes processos, crescimento aeróbio, crescimento anóxico e decaimento, conforme a hipótese de morte-regeneração, da seguinte maneira:

$$\frac{dX_{B,H}}{dt} = [\mu_H(\frac{S_S}{K_S + S_S})\{(\frac{S_O}{K_{O,H} + S_O}) + \eta_g(\frac{K_{O,H}}{K_{O,H} + S_O})(\frac{S_{NO}}{K_{NO} + S_{NO}})\} - b_H]X_{B,H} \quad (2.17)$$

2. Concentração de biomassa autotrófica: é simplesmente aquela fração de biomassa que não cresce em ambiente anóxico. Inclui decaimento:

$$\frac{dX_{B,A}}{dt} = [\mu_A(\frac{S_{NH}}{K_{NH} + S_{NH}})(\frac{S_O}{K_{O,A} + S_O}) - b_A]X_{B,A} \quad (2.18)$$

3. Concentração de substrato rapidamente biodegradável: é reduzido pelo crescimento de bactérias heterotróficas, em condições tanto aeróbias como anóxicas e aumentada mediante hidrólise do substrato lentamente biodegradável, conforme:

$$\begin{aligned} \frac{dS_S}{dt} = & [-\frac{\mu_H}{Y_H}(\frac{S_S}{K_S + S_S})\{(\frac{S_O}{K_{O,H} + S_O}) + \eta_g(\frac{K_{O,H}}{K_{O,H} + S_O})(\frac{S_{NO}}{K_{NO} + S_{NO}})\} \\ & + k_h\frac{X_S/X_{B,H}}{K_X + (X_S/X_{B,H})}\{(\frac{S_O}{K_{O,H} + S_O}) + \eta_h(\frac{K_{O,H}}{K_{O,H} + S_O})(\frac{S_{NO}}{K_{NO} + S_{NO}})\}]X_{B,H} \end{aligned} \quad (2.19)$$

4. Concentração de substrato lentamente biodegradável: é acrescida pelo reciclo de bactérias heterotróficas e autotróficas conforme a hipótese de morte e regeneração. É decrescida mediante o processo de hidrólise, modelada por:

$$\begin{aligned} \frac{dX_S}{dt} = & (1 - f_P)(b_H X_{B,H} + b_A X_{B,A}) - k_h\frac{X_S/X_{B,H}}{K_X + (X_S/X_{B,H})}\{(\frac{S_O}{K_{O,H} + S_O}) + \\ & \eta_h(\frac{K_{O,H}}{K_{O,H} + S_O})(\frac{S_{NO}}{K_{NO} + S_{NO}})\}X_{B,H} \end{aligned} \quad (2.20)$$

5. Concentração de produtos particulados inertes: originada pelo decaimento de biomassa heterotrófica e autotrófica, modelada por:

$$\frac{dX_P}{dt} = f_P(b_H X_{B,H} + b_A X_{B,A}) \quad (2.21)$$

6. Concentração de nitrogênio orgânico biodegradável particulado: é acrescida pelo decaimento de biomassa heterotrófica e autotrófica e decrescida pelo processo de hidrólise.

$$\frac{dX_{ND}}{dt} = (i_{XB} - f_P i_{XP})(b_H X_{B,H} + b_A X_{B,A}) - k_h \frac{X_{ND}/X_{B,H}}{K_X + (X_S/X_{B,H})} \left\{ \left(\frac{S_O}{K_{O,H} + S_O} \right) + \eta_h \left(\frac{K_{O,H}}{K_{O,H} + S_O} \right) \left(\frac{S_{NO}}{K_{NO} + S_{NO}} \right) \right\} X_{B,H} \quad (2.22)$$

7. Concentração de nitrogênio orgânico solúvel: afetada pelo processo de amonificação e hidrólise conforme a:

$$\frac{dS_{ND}}{dt} = \left[-k_a S_{ND} + k_h \frac{X_{ND}/X_{B,H}}{K_X + (X_S/X_{B,H})} \left\{ \left(\frac{S_O}{K_{O,H} + S_O} \right) + \eta_h \left(\frac{K_{O,H}}{K_{O,H} + S_O} \right) \left(\frac{S_{NO}}{K_{NO} + S_{NO}} \right) \right\} \right] X_{B,H} \quad (2.23)$$

8. Concentração de amônio: é reduzida pelo crescimento de todos os microrganismos que utilizam amônio como fonte de nitrogênio para incorporá-lo dentro da massa celular. A concentração é também decrescida pelo processo de nitrificação e aumentada como resultado da amonificação de nitrogênio orgânico solúvel é modelada por:

$$\frac{dS_{NH}}{dt} = \left[-i_{XB} \mu_H \left(\frac{S_S}{K_S + S_S} \right) \left\{ \left(\frac{S_O}{K_{O,H} + S_O} \right) + \eta_g \left(\frac{K_{O,H}}{K_{O,H} + S_O} \right) \left(\frac{S_{NO}}{K_{NO} + S_{NO}} \right) \right\} - k_a S_{ND} \right] X_{B,H} - \mu_A \left(i_{XB} + \frac{1}{Y_A} \right) \left(\frac{S_{NH}}{K_{NH} + S_{NH}} \right) \left(\frac{S_O}{K_{O,A} + S_O} \right) X_{B,A} \quad (2.24)$$

9. Concentração de nitrato: é aumentada pelo processo de nitrificação e reduzida pelo processo de desnitrificação. A equação dinâmica que descreve estes processos é:

$$\frac{dS_{NO}}{dt} = -\mu_H \eta_g \left(\frac{1 - Y_H}{2,86 Y_H} \right) \left(\frac{S_S}{K_S + S_S} \right) \left(\frac{K_{O,H}}{K_{O,H} + S_O} \right) \left(\frac{S_{NO}}{K_{NO} + S_{NO}} \right) X_{B,H} + \frac{\mu_A}{Y_A} \left(\frac{S_{NH}}{K_{NH} + S_{NH}} \right) \left(\frac{S_O}{K_{O,A} + S_O} \right) X_{B,A} \quad (2.25)$$

10. Concentração de oxigênio dissolvido: é reduzida pelo crescimento aeróbio de heterótrofos e biomassa autotrófica, modelada por:

$$\frac{dS_O}{dt} = -\mu_H \left(\frac{1 - Y_H}{Y_H} \right) \left(\frac{S_S}{K_S + S_S} \right) \left(\frac{S_O}{K_{O,H} + S_O} \right) X_{B,H} - \mu_A \left(\frac{4,57 - Y_A}{Y_A} \right) \left(\frac{S_{NH}}{K_{NH} + S_{NH}} \right) \left(\frac{S_O}{K_{O,A} + S_O} \right) X_{B,A} \quad (2.26)$$

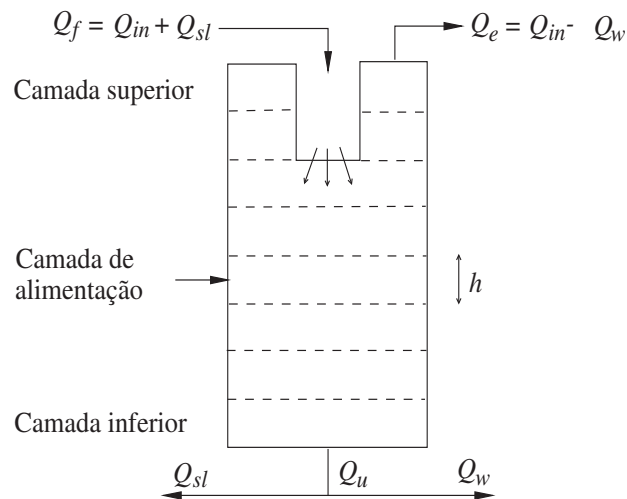


Fig. 2.9: Modelo do decantador segundo o princípio de múltiplas camadas.

Como visto na seção 2.2, a fração de matéria orgânica não biodegradável particulada (X_I) e de matéria orgânica não biodegradável solúvel (S_I), passam pelo sistema sem sofrer transformações. Portanto suas taxas de reação são consideradas nulas. A taxa de reação descrevendo a dinâmica de alcalinidade não é apresentada.

Os valores dos parâmetros para cada uma das taxas de reação, a caracterização do esgoto e as condições de operação nominal não são apresentadas mas podem ser obtidas em Sotomayor et al. (2001).

Modelo do Decantador

A função do decantador secundário é separar as fases sólida e líquida, separando os flocos de lodo da água tratada e condensando o lodo que irá retornar ao bioreator ou que poderá ser também descartado.

Uma característica importante para o controle do processo é o decantador ser considerado como um armazém de lodos ativados, que poderão ser retornados ao bioreator com o objetivo de manter uma adequada população microbiana. É um elemento de muita importância na determinação da dinâmica completa do processo de lodos ativados, dado que existe um forte acoplamento entre o bioreator e o decantador secundário.

Dentre os modelos de decantadores secundários, o modelo proposto por Takács et al. (1991) representa mais realisticamente o processo clarificação-sedimentação, sob condições dinâmicas e de estado estacionário e é o modelo adotado no simulador ASWWT-USP. O modelo de Takács et al. (1991) é um modelo unidimensional, de múltiplas camadas, isto é, considera o decantador secundário como um tanque com várias camadas horizontais, como mostrado na Figura 2.9. Cada camada modela fluxos de líquidos e sólidos (licor) que se propagam através do decantador em duas direções a partir do fluxo de alimentação (Q_f), denominada camada de alimentação. Um fluxo ascendente em direção à camada superior e, um segundo, descendente em direção à camada inferior. Isto significa que o fluxo ascendente (Q_e) é igual ao fluxo do efluente e que o fluxo descendente é igual à soma do fluxo

de recirculação de lodo (Q_{sl}) com o fluxo de lodo em excesso (Q_w). As relações entre os diferentes fluxos, conforme a Figura 2.9, são as seguintes:

$$\begin{aligned}Q_f &= Q_{in} + Q_{sl} \\Q_e &= Q_{in} - Q_w \\Q_u &= Q_{sl} + Q_w\end{aligned}\tag{2.27}$$

Adicionalmente, o modelo considera a sedimentação de sólidos dentro de cada camada que são colocados na parte externa, propagando-se da camada superior em direção à camada inferior. No modelo do decantador é assumido que cada camada tem mistura perfeita.

2.6 Conclusões

Os conceitos apresentados neste capítulo foram basicamente divididos em duas partes. Na primeira foram introduzidos, de forma qualitativa e progressiva, os mecanismos biológicos que governam o processo de lodos ativados em estações de tratamento de esgotos. Fenômenos tais como crescimento da biomassa, hidrólise e decaimento foram apresentados para dar sustento teórico à posterior apresentação das relações bioquímicas presentes nos processos de remoção de matéria orgânica nitrogenada. Desta forma, fica clara a formulação das taxas de remoção de substrato orgânico na remoção aeróbia de carbono orgânico, da taxa de remoção de amônio por nitrificação, da taxa de produção de nitratos e da taxa de remoção de nitrato por desnitrificação. Deve-se destacar que a nitrificação não resulta na remoção de nitrogênio, mas uma conversão na sua forma, de amônio a nitrato. Estes conceitos servem como fundamento teórico para a apresentação do modelo IAWQ No. 1, onde se introduz as variáveis de estado ou concentrações envolvidas em processos de lodos ativados e que são usadas ao longo do presente trabalho nos projetos de identificação e controle.

O esquema que resume os processos biológicos de remoção de matéria carbonácea e nitrogenada será particularmente útil, pois facilita a compreensão das taxas de reação usadas pelo modelo de simulação, o ASWWTP-USP, cuja breve descrição constituiu a segunda parte deste capítulo. O modelo de simulação tem sido uma ferramenta muito útil para estudar os vários aspectos dos processos de lodos ativados, como o conhecimento dos modos de operação do processo e o comportamento dinâmico das variáveis envolvidas. Este simulador é composto por uma zona anóxica e duas zonas aeróbias em uma configuração de pre-desnitrificação. Cada um dos bioreatores consideram os processos dinâmicos do modelo IAWQ No. 1. Finalmente foi ressaltada a importante função do decantador secundário dentro da dinâmica completa do processo de lodos ativados, elemento que também é incluído no modelo de simulação.

Capítulo 3

Projeto de Sinais de Identificação para Controle

Sabendo que um dos objetivos do presente trabalho é a busca de modelos matemáticos de ordem reduzida apropriados para controle, aborda-se neste capítulo o procedimento para a síntese de sinais de identificação relevantes para controle.

Ao longo deste capítulo são utilizados conceitos relativos à identificação não paramétrica pertinentes ao projeto de sinais de excitação, apesar de serem usados modelos paramétricos para caracterizar nossos modelos reduzidos. É importante enfatizar que se um sinal é bem projetado para identificar um modelo não paramétrico de um sistema, então este é também bem projetado para identificar um modelo paramétrico do sistema (Godfrey, 1993).

3.1 Introdução

Nos últimos anos vem se incrementando o reconhecimento do papel que sinais de perturbação convenientemente projetados podem exercer na identificação de sistemas (Godfrey et al., 2003). Este é o caso de sinais de excitação que sejam suficientemente informativos e que possuam excitação numa faixa de frequência relevante para controle, tanto na identificação de sistemas lineares quanto não lineares.

Do ponto de vista teórico, sinais de entrada aleatórios são interessantes desde que possuam funções de autocorrelação de natureza ruído branco (definidas na subseção 3.2.1), especialmente quando não é possível obter conhecimento sobre os parâmetros e a estrutura da planta a ser modelada. No entanto, o uso de sinais aleatórios apresenta dificuldades de ordem prática (Braun, 2001; Aguirre, 2000). Por exemplo, o *span* (amplitude da faixa de operação física do elemento final de controle), pode ser restrito pelo atuador ou então, o comprimento do sinal de entrada poderia não excitar a largura de faixa de interesse para controle.

Sinais de entrada pseudo-aleatórios (por exemplo sinais determinísticos com propriedades similares às de sinais aleatórios), podem ser projetados como alternativa no caso de se contar com um conhecimento *a priori* do sistema a identificar, facilitando a especificação de uma faixa de frequência, que terá dependência direta com o controlador a ser projetado usando o modelo obtido. Restrições na amplitude de entrada, deslocamento no *span* do atuador e o comprimento do sinal podem ser conhecidos

antes de aplicar o sinal à planta (Braun et al., 1999). Por esta razão, sinais pseudo-aleatórios de posse destas características de projeto são denominados também sinais determinísticos (Braun, 2001). Outra vantagem é que o tempo requerido para a identificação é menor quando comparado com o tempo necessário para sinais aleatórios (Godfrey, 1993). Estes sinais têm também função de autocorrelação do tipo ruído branco e podem ser reproduzidos repetidas vezes para períodos pré-especificados (Davies, 1970), por serem de natureza pseudo-aleatória.

Um dos tipos de sinal pseudo-aleatório mais popular na identificação de sistemas lineares é o sinal binário pseudo-aleatório, devido à sua facilidade de geração e transdução elétrica. Contudo, estes sinais podem apresentar limitações na identificação do comportamento não linear de um sistema. A limitação pode ser ilustrada supondo um sistema não linear de segunda ordem $y = u^2$, que testado com um sinal binário u com níveis 1 e -1 resultará em uma saída y constante. Como consequência, o interesse pelo uso de sinais pseudo-aleatórios multi-níveis para identificação de sistemas acaba sendo natural. Estes sinais testam um sistema em vários níveis de operação, provendo a possibilidade de identificar um comportamento não linear ou ainda identificar o comportamento linear em presença de não linearidades, reduzindo o efeito de distorções devido à presença de não linearidades (Barker, 1993). Ao mesmo tempo, por serem sinais do tipo pseudo-aleatório, possuem também função de autocorrelação do tipo ruído branco.

A desvantagem que pode ser argumentada contra estes sinais é com respeito à complexidade envolvida em sua geração, baseada na Teoria de Campos Finitos, especificamente em Campos de Galois. Portanto, neste capítulo apresentam-se os fundamentos para a geração de sinais pseudo-aleatórios multi-níveis, expondo o procedimento para o projeto destes sinais baseados no conhecimento *a priori* do sistema. Estas diretrizes possibilitam projetar sinais pseudo-aleatórios multi-níveis numa faixa de frequência útil para o controle considerando a aplicação como sendo apropriada para atuadores em plantas industriais, denominadas *plant-friendly*.

O sentido do termo *plant-friendly* na identificação de sistemas para indústria de processos está baseado na necessidade de experimentos informativos que reúnam requisitos práticos. Rivera et al. (2003) o define como sendo os sinais de testes que gerem dados tal que produzam um modelo adequado para tempos de experimentos aceitáveis, mantendo a variação da entrada e da saída do sinal dentro de restrições definidas pelo usuário. Entre as restrições por eles incluídas estão:

- Manter desvios os menores possíveis tal que a variabilidade na qualidade do produto seja reduzida.
- Implementar um sinal da menor duração possível tal que reduza o custo de engenharia associado à duração do experimento e que ao mesmo tempo reduza a quantidade de produtos usados para o ensaio de identificação.
- Manter os desvios da entrada tão pequenos quanto necessários, tal que não forcem as restrições físicas dos atuadores.

Para mostrar como podem ser gerados sinais que reúnam estes requisitos, é mostrado o procedimento para o projeto de sinais de identificação para controle e a aplicação em experimentos de identificação. Portanto, o presente capítulo é estruturado da seguinte forma. É realizada uma revisão de sinais de excitação relativos à identificação de sistemas lineares e não lineares, assim como é introduzido o conceito de redução de distorções por não linearidades na identificação de sistemas lineares.

Índices de desempenho para medir a qualidade do sinal de excitação são apresentados em seguida. Na seção seguinte é explicado o procedimento de geração de sinais multi-níveis. Para ilustrar a eficácia da metodologia é analisada a aplicação desta abordagem no processo de neutralização de pH, que se caracteriza por ser um sistema com características estáticas não lineares. Finalmente, a identificação da dinâmica do processo de oxigênio dissolvido é realizada, processo que é de vital importância para a remoção aeróbia da matéria orgânica carbonácea em plantas de tratamento de esgotos por lodos ativados. É mostrada a eficácia dos sinais de excitação multi-níveis com supressão de harmônicos para a identificação de modelos cuja aplicação seja o projeto de controladores baseados em modelo.

3.2 Tipos de sinais de excitação

Como visto anteriormente, os sinais de excitação podem ser classificados em dois grandes grupos: sinais aleatórios e sinais pseudo-aleatórios. A seguir apresenta-se uma breve revisão de sinais de excitação na identificação de sistemas, junto com a informação para a obtenção dos programas computacionais geradores destes tipos de sinais.

3.2.1 Sinais de excitação aleatórios

Nesta categoria estão os sinais denominados ruído branco, isto é, sinais com função de autocorrelação nula para todos os atrasos τ fora da origem, de média nula e potência espectral constante para uma determinada faixa de frequências:

$$E[u(k)u(k + \tau)] = \begin{cases} 0 & , \tau \neq 0 \\ \sigma^2 & , \tau = 0 \end{cases} \quad (3.1)$$

onde o operador $E[\cdot]$ denota a esperança matemática. A razão da utilização de sinais ruído branco como sinais de excitação é porque eles são persistentemente excitantes em diferentes frequências. Dois dos tipos de sinais estocásticas mais populares na identificação de sistemas são:

- Ruído branco binário (RBB): composto por variáveis binárias tomando valores $\pm a$ da amplitude do sinal. A variância do sinal RBB é a^2 . Apesar de ser um sinal relativamente fácil de implementar, requer um tempo de experimentação longo para que suas propriedades tenham significado. Um exemplo deste sinal é baseado em cadeias de Markov (Pearson et al., 1998).
- Ruído branco gaussiano (RBG): cuja função de distribuição de probabilidade é gaussiana $\mathcal{N}(0, \sigma^2)$, sobre o intervalo $\pm a$ cuja forma está parcialmente caracterizada por sua função de autocorrelação.

Na Figura 3.1 são mostrados sinais de ruídos branco binário e gaussiano, respectivamente, para um comprimento do sinal $N_s = 64$. Do ponto de vista de aplicação em plantas como em estações de tratamento de esgotos, existem dois tipos de problemas. O primeiro é que o ruído branco gaussiano é continuamente distribuído. Como tal, em um dado instante t , a entrada irá comutar para um novo valor, porém este movimento constante resultará em um significativo deslocamento do atuador. O segundo problema é que estes sinais precisam de registros de dados longos para atingir a propriedade de

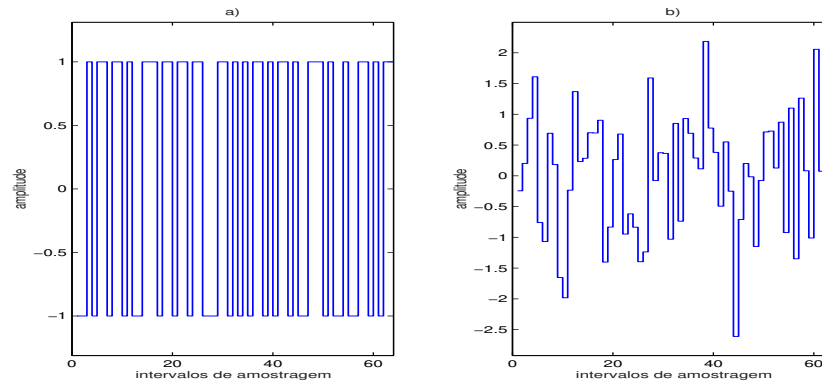


Fig. 3.1: Sinais de excitação estocásticos: a) Ruído branco binário; b) Ruído branco gaussiano.

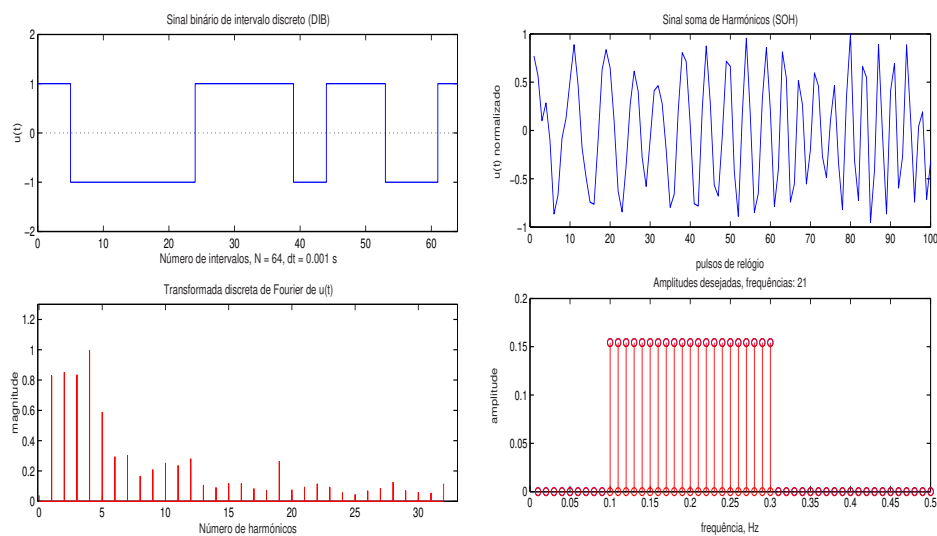
autocorrelação nula e assim proporcionar estimativas de parâmetros razoavelmente precisas. Estimativas precisas requerem longos períodos de medição o que pode no final prejudicar economicamente o custo de modelagem do processo de tratamento.

3.2.2 Sinais de excitação pseudo-aleatórios

Sinais pseudo-aleatórios periódicos têm potência espectral fixa e a escolha de projeto reside principalmente em o sinal ser binário ou multi-nível. Podem ser divididos em duas categorias:

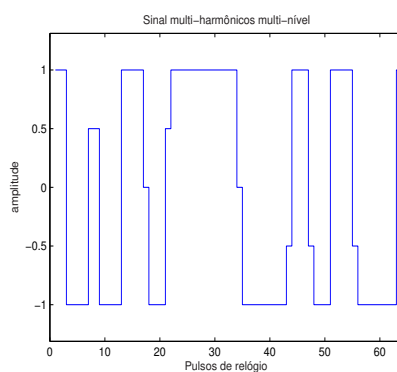
- Sinais de espectro ótimo: sinais de espectro ótimo são projetados para colocar o espectro de potência próximo ao especificado, onde a proporção da potência do sinal contida em harmônicos especificados é maximizada. A esta categoria pertencem:
 - Sinal binário de intervalo discreto (DIB - *discrete interval binary*): o objetivo neste caso é forçar o máximo de potência possível dentro de harmônicos especificados. Os níveis da seqüência do sinal são ± 1 , como visto na Figura 3.2(a).
 - Sinal multi-seno (SOH - *sum of harmonics*): pode tomar diferentes valores dentro de um intervalo máximo e mínimo especificado. Aqui são especificados aqueles harmônicos com uma potência requerida juntamente com um padrão de amplitude desejado, como pode ser visto na Figura 3.2(b).
 - Sinal com multi-harmônicos multi-nível (MLMH - *multi-level multi-harmonics*): podem especificar um número de níveis ≥ 2 e reúnem as vantagens dos sinais DIB e SOH anteriores sem suas desvantagens. Sinais DIB são binários, tem uma amplitude pico-a-pico bem definida, mas uma quantidade significativa da potência total aparece em harmônicos não especificados. Sinais SOH têm uma potência maior em harmônicos especificados, mas uma maior amplitude pico-a-pico.

Sinais DIB e SOH podem ser projetados usando o “Frequency Domain Identification Toolbox - FDIDENT” da Mathworks Inc., através das funções `dibs` e `msinclip`. Sinais MLMH podem ser projetados usando a função em Matlab `multilev` como mostrado em Cormack et al.



(a) Sinal binário de intervalo discreto (DIB), com amplitudes máximas de potência em $k = 1, 2, 3, 4$ e 5 .

(b) Sinal soma de harmônicos (SOH), com potência especificada no intervalo 0,1 a 0,3Hz.



(c) Sinal multi-harmônicos de 5 níveis.

Fig. 3.2: Sinais de espectro ótimo.

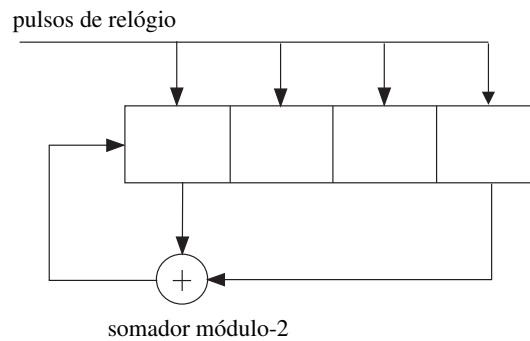


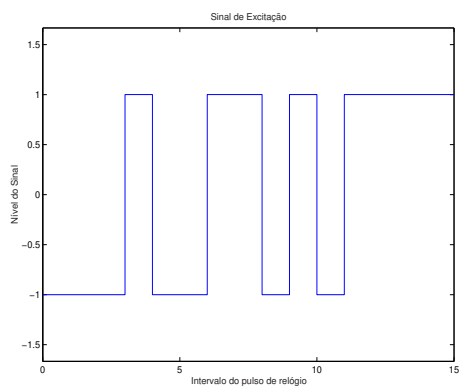
Fig. 3.3: Esquema do gerador PRBS de 4-estágios (comprimento máximo de 15 dígitos).

(1995), no qual ademais é realizado um estudo comparativo para a identificação de sistemas lineares com sinais de espectro ótimo. Apesar da função `multilev` não pertencer ao “FDIDENT Toolbox”, esta pode ser obtida livremente através do site (último acesso 21/05/2004):

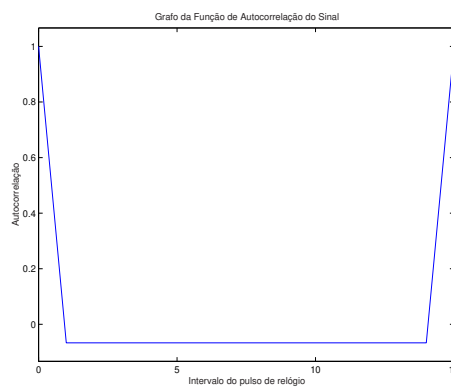
<http://www.eng.warwick.ac.uk/EED/DSM/multilev.htm>.

Um exemplo de sinal MLMH é mostrado na Figura 3.2(c), para um comprimento $N_s = 64$ e 5 níveis, gerado com a função `multilev`. Sinais de espectro ótimo não permitem especificar algumas supressões de harmônicos, como apreciado nas Figuras 3.2(a) e 3.2(b).

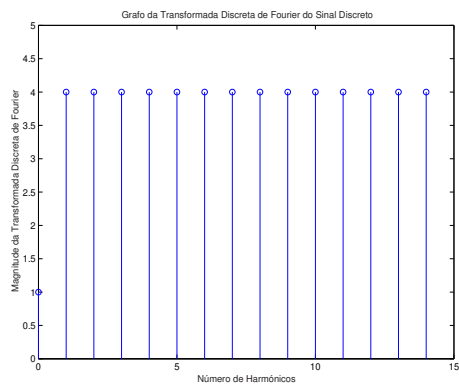
- Sinais pseudo-aleatórios baseados em seqüências m : diferentemente dos sinais de espectro ótimo, estes sinais pseudo-aleatórios têm um espectro de potência fixo. São gerados a partir de seqüências de comprimento máximo (seqüências m), baseadas em Campos de Galois e podem ser tanto binárias como multi-nível. Este tipo de sinal possui propriedades de autocorrelação e correlação cruzada próximas do ruído branco, porém possuem a propriedade de serem persistentemente excitantes conforme a definição (B.3.1). Adicionalmente, estes podem ser projetados tal que seu espectro de potência seja uniforme com a possibilidade de indicar quais harmônicos serão suprimidos. A esta categoria de sinais pertencem:
 - Sinal pseudo-aleatório binário (PRBS - *pseudo-random binary signal*): o sinal PRBS está baseado na seqüência binária pseudo-aleatória de comprimento máximo $N_s = 2^{n_r} - 1$, com n_r um inteiro positivo tal que $N_s = 3, 7, 15, 31, 63, 127$, etc. É especificado pelo número de estágios do registro de deslocamento (n_r) e pelo pulso de relógio ou tempo de comutação (T_{sw}), que é o tempo mínimo entre mudanças de níveis do sinal. O sinal PRBS é facilmente gerado usando registros de deslocamento de n_r -estágios com realimentação ao primeiro estágio de uma soma módulo-2. Na Figura 3.3 é mostrado um gerador PRBS de 4-estágios, cujo sinal pode ser visto na Figura 3.4(a). Note na Figura 3.4(b) que este sinal possui função de autocorrelação do tipo ruído branco e que apresenta todos os harmônicos com potência uniforme, tal como mostrado na Figura 3.4(c), após a aplicação de um dispositivo segurador de ordem zero (ZOH).



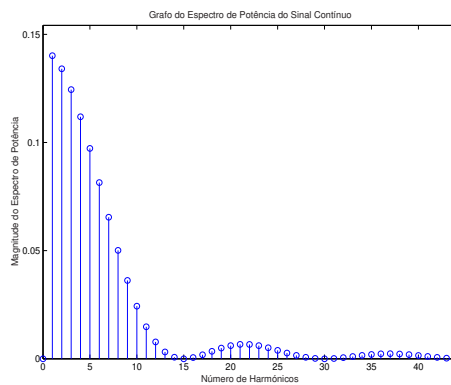
(a) Sinal pseudo-aleatório binário para $N_s = 15$ e $n_r = 4$.



(b) Função de autocorrelação.



(c) Espectro de potência de $\{u_i\}$.



(d) Espectro de potência de $u(t)$ (após o ZOH).

Fig. 3.4: Exemplo de sinal pseudo-aleatório binário.

- Sinal pseudo-aleatório binário com repetição inversa (IPRBS): a versão de repetição inversa é obtida invertendo todos os outros membros da seqüência para dar uma nova seqüência com período $2N_s$ onde uma metade da seqüência será o negativo da outra metade. A nova seqüência satisfaz a propriedade:

$$\{u_i\} + \{u_{i+N_s/2}\} = 0 \quad , \forall i \quad (3.2)$$

portanto é antisimétrica com propriedade de repetição inversa. Sua função de autocorrelação é aproximadamente nula:

$$R_{uu}(k) + R_{uu}(k + N_s/2) = 0 \quad , \forall k \quad (3.3)$$

exceto para uns poucos valores de k dentro do período N_s :

$$R_{uu}(k) = 0 \quad , \forall k \neq rN_s/2 \quad , r = 0, 1 \quad (3.4)$$

Caracteriza-se por ter harmônicos múltiplos de 2 suprimidos, como será visto na subseção 3.4.2. Este tipo de sinal tem a propriedade de reduzir efeitos de termos não lineares na identificação de sistemas lineares (Godfrey, 1993), no entanto sinais IPRBS requerem de um tempo de experimentação maior devido ao período do sinal ser $2N_s$.

- Sinal pseudo-aleatório quasi-binário (*near-binary pseudo-random signals*): Embora sinais PRBS baseados em seqüências m sejam os mais conhecidos, existem outros tipos de sinais com funções de autocorrelação idênticas ou quasi-idênticas ao PRBS, com possibilidade de supressão de harmônicos (Tan e Godfrey, 2001):
 - * Sinal binário de resíduo quadrático (QRB - *quadratic-residue binary*), com período $N = 4n - 1$ e primo, isto é $N = 3, 7, 11, 19$, etc.
 - * Sinal ternário de resíduo quadrático (QRT - *quadratic-residue ternary*), de três níveis com período $N = 4n \pm 1$ primo ímpar, isto é $N = 3, 5, 7, 11, 13$, etc. com n inteiro positivo.
 - * Sinal binário de Hall (HAB - *Hall binary*), com período $N = 4n^2 + 27$ e primo, isto é $N = 31, 43, 127$, etc.
 - * Sinal binário primo twin (TPB - *twin-prime binary*), com período $N = n(n + 2)$ quando n e $(n + 2)$ são primos, isto é $N = 15, 35, 143$, etc.
- Sinais pseudo-aleatório multi-nível (MLPRS - *multi-level pseudo-random signals*): originados também a partir de seqüências m , existem para um números de níveis q , igual a um primo ou a potência de um primo ($q = 2, 3, 4, 5, 6, 8, 9, 11, \dots$, etc). O comprimento da seqüência é $N_s = q^{n_r} - 1$, com n_r um inteiro. Convertendo os elementos do campo de Galois em níveis apropriados consegue-se um sinal pseudo-aleatório multi-nível (MLPRS) de média nula. São gerados de forma similar ao sinal PRBS, com n_r registros de deslocamento e somadores módulo- n_r .

Uma das principais razões para considerar sinais baseados em seqüências m com $q > 2$, é que estes podem contribuir para melhores estimativas que os sinais binários, pois testam um sistema em diferentes pontos de operação e dão a possibilidade de identificar um

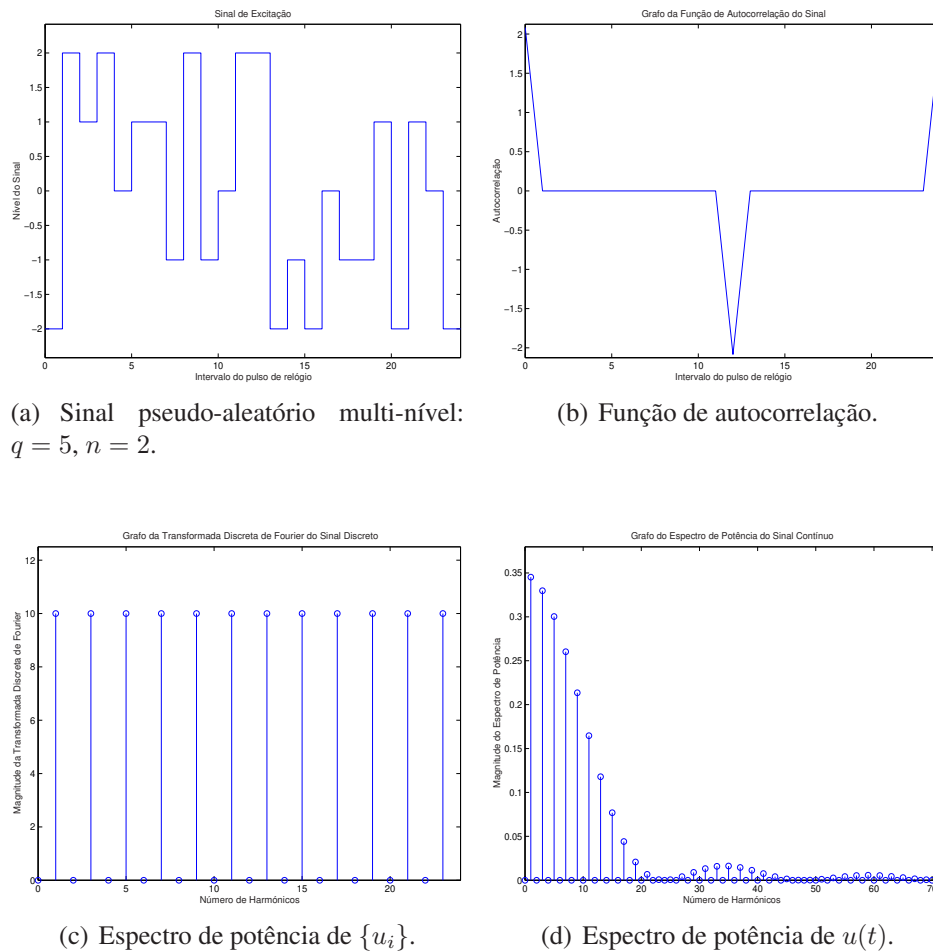


Fig. 3.5: Exemplo de sinal pseudo-aleatório multi-nível (com supressão de harmônicos múltiplos de 2).

comportamento não linear ou ainda identificar um comportamento linear em presença de não linearidade, como será visto na seção 3.4. Também sinais MLPRS podem ser projetados com a propriedade de repetição inversa da equação (3.3) para um comprimento N_s .

O sinal PRBS pode ser facilmente gerado mediante a função `m1bs` do `FDIDENT` Toolbox para um $N_s = 3$ até $N_s = 1,0737 \times 10^9$, isto é o N_s máximo possível para um registro de deslocamento de 30 estágios. A função `m1bs` não pode gerar sinais PRBS com outra combinação para a realimentação de estágios. O pacote `prs` (Tan e Godfrey, 2001) em Matlab[®], permite gerar todos os tipos de sinais pseudo-aleatórios binários mencionados, inclusive com outras combinações. Ele pode ser obtido livremente através do site (último acesso 21/05/2004):

<http://www.eng.warwick.ac.uk/EED/DSM/prs.htm>.

O software Galois[®] (Barker e Zhuang, 1997b) escrito em C++, gera sinais MLPRS para $q \leq 128$, com a possibilidade de indicar os harmônicos a serem suprimidos no sinal de excitação. O soft-

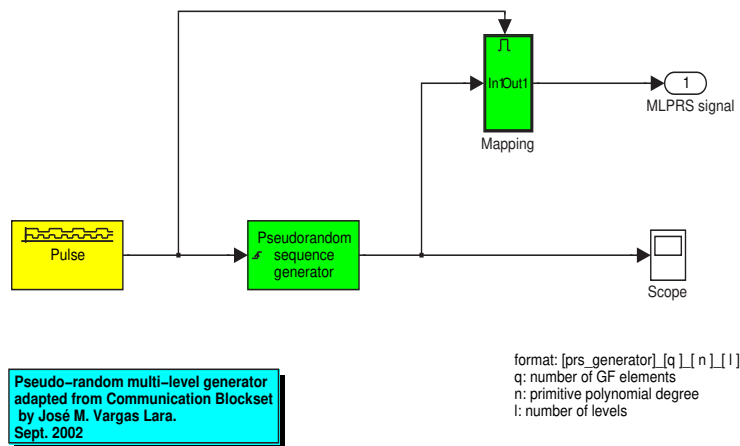


Fig. 3.6: Gerador de sinais MLPRS em Simulink.

ware pode ser disponibilizado livremente mediante contato com o primeiro autor. O gerador de sinais MLPRS com supressão de harmônicos (Lara e Milani, 2003), permite gerar sinais de q número de níveis (q primo), para serem usados no ambiente Simulink do Matlab[®], usando os mapeamentos de Barker e Zhuang (1997b). Um exemplo de sinal MLPRS de 5 níveis com comprimento $N_s = 25$ e com supressão de harmônicos pares (do tipo repetição inversa), pode ser visto na Figura 3.5. O bloco gerador de sinais MLPRS em Simulink é mostrado na Figura 3.6, o qual será usado posteriormente nos experimento de identificação da dinâmica do processo de oxigênio dissolvido e do processo de remoção de matéria nitrogenada a partir do modelo *benchmark* ASWWTP-USP. Os fundamentos de como estes sinais são gerados assim como a aplicação das diretrizes de projeto são revistos na seção 3.3 e na seção 3.6, respectivamente.

3.3 Geração de sinais pseudo-aleatórios multi-níveis

Os fundamentos teóricos para a geração de sinais pseudo-aleatórios multi-níveis residem na teoria de campos finitos, especificamente em Campos de Galois (Zierler, 1959). Estes sinais são gerados a partir de seqüências de período máximo usando recorrências lineares. Como visto anteriormente, estas são referidas como seqüências de comprimento máximo, denotadas como seqüências m , porque são formadas por sub-seqüências que diferem umas de outras, formando uma única seqüência cujo comprimento será o máximo possível. O desenvolvimento teórico para a geração de sinais de seqüência m é detalhado em Barker (1993). Portanto aqui só será apresentada a parte essencial da teoria que serve de fundamento para o projeto de sinais de excitação para identificação de sistemas.

3.3.1 Campos de Galois

Um Campo de Galois é uma coleção finita de elementos que satisfazem os requisitos da teoria de campos finitos, em que o número de elementos de campo pode ser um primo ou uma potência de um primo.

O campo finito mais simples é o campo primo $GF(p)$ de p elementos de campo a_0, a_1, \dots, a_{p-1} , cujos elementos de campo devem ser inteiros módulo- p tal que satisfaçam os requisitos de um campo primo:

$$a_i = i, \quad i = 0, 1, \dots, p-1, \quad p \text{ primo.} \quad (3.5)$$

O campo primo $GF(p)$ realiza suas operações de álgebra de campos mediante a aritmética módulo- p de soma e multiplicação:

$$\begin{aligned} a_i + a_j &= i + j, & \text{módulo-}p \\ a_i a_j &= ij, & \text{módulo-}p \end{aligned} \quad (3.6)$$

A potência de um primo $q = p^l$ (l positivo inteiro) origina a extensão do campo denotada por $GF(q)$, definida quando os q elementos de campo b_0, b_1, \dots, b_{q-1} são polinômios de grau $l-1$ no campo primo $GF(p)$, com coeficientes que são elementos de $GF(p)$, considerando todas as possíveis combinações. As operações de álgebra de campos não são mais de aritmética módulo- p , mas podem ser implementadas através de tabelas como as mostradas em Barker (1993), com valores de q até 27.

Por simplicidade adota-se o campo de número primos ($l = 1$), que será denotado equivalentemente por $GF(q)$. Isto permite a geração de seqüências m mediante a operação de soma e multiplicação módulo- q utilizando registros de deslocamento, como será visto na seção 3.3.2. Cabe ressaltar que apesar desta abordagem para geração de sinais não ser recente (Yarmolik e Demidenko, 1988), não houve difusão da sua aplicação como sinais de excitação devido à falta de recursos computacionais e hardware iniciais. Hoje em dia existem esses recursos, no entanto é surpreendente notar poucas aplicações na sua utilização como sinais para identificação de processos industriais (Braun, 2001).

3.3.2 Seqüências m em $GF(q)$

Uma seqüência m , $\{x_i\}$, em $GF(q)$ é gerada através da relação de recorrência:

$$x_i = - \sum_{r=1}^{n_r} a_r x_{i-r}, \quad \forall i \quad (3.7)$$

O registro de deslocamento de n_r -estágios mostrado na Figura 3.7, com conexões que resultam em uma seqüência de comprimento máximo, tem uma equação característica correspondente ao polinômio primitivo módulo- q ¹. Um polinômio primitivo é aquele que não é possível efetuar sua fatoração em pelo menos dois polinômios irredutíveis de graus inferiores e que não é fator de $1 \oplus x^r$ para qualquer $r < q^{n_r} - 1$, onde n_r é o grau do polinômio. Este polinômio primitivo $f(D)$ de grau n_r em $GF(q)$, cujos coeficientes a_r tomam valores inteiros dentro do intervalo 0 a $q-1$, é (Barker, 1993):

$$f(D)x_i = \sum_{r=0}^{n_r} a_r D^r x_i, \quad a_0 = 1, a_{n_r} \neq n_r \quad (3.8)$$

¹Define a operação de adição da seguinte forma : diz-se que o inteiro r é a soma dos inteiros s e t módulo q se r for o resto da divisão de $s + t$ por q , operação denotada por \oplus .

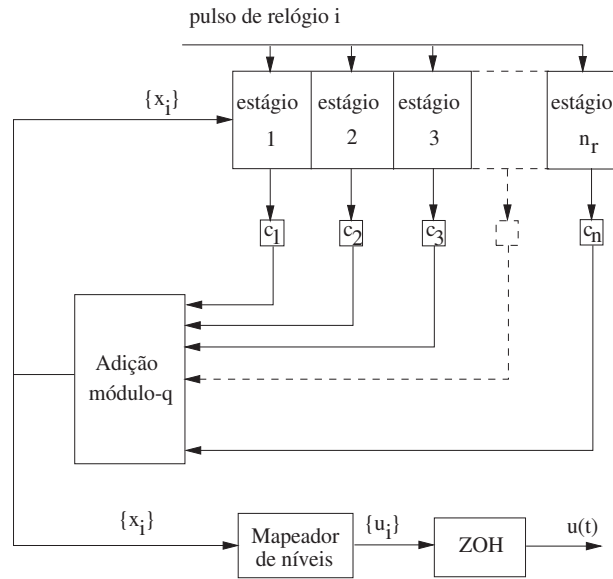


Fig. 3.7: Esquema do gerador do sinal de seqüência m multi-nível (Braun et al., 1999).

em que D corresponde ao operador de atraso. A equação (3.8) é a equação característica da seqüência $\{x_i\}$. Para $q > 2$ a subtração módulo- q não é a mesma que a adição como no caso binário, portanto a realimentação tem que ser revisada. Se $f(D)$ é primitivo, módulo- q , a correspondente equação característica revisada, com conexões lógicas retornando ao primeiro estágio do registro de deslocamento, é dada por Godfrey (1993):

$$X = c_1DX + \dots + c_{n_r-1}D^{n_r-1}X + c_{n_r}D^{n_r}X \quad , \text{módulo-}q \quad (3.9)$$

onde os coeficientes c_r são calculados mediante a relação:

$$c_r = (q - a_r), \quad r = 1, 2, \dots, n_r \quad (3.10)$$

3.3.3 Sinal de seqüência m multi-nível

Uma seqüência pseudo-aleatória, $\{u_i\}$, é gerada mapeando os elementos da seqüência $\{x_i\}$ da equação (3.7) em números reais apropriados, determinando o número de níveis m ($m < q$) e os harmônicos a serem suprimidos. Um segurador de ordem zero, com tempo de comutação T_{sw} é finalmente aplicado à seqüência $\{u_i\}$, dando origem ao sinal de seqüência m multi-nível MLPRS, $u(t)$, de acordo com:

$$u(t) = u_i \quad , i \cdot T_{sw} \leq t \leq (i + 1) \cdot T_{sw} \quad (3.11)$$

Na Figura 3.7 são mostrados os elementos do gerador do sinal de seqüência m e o fluxo de dados até a obtenção do sinal $u(t)$ em função do tempo. Note que para cada pulso de relógio as saídas do registro são multiplicadas com os ganhos relacionados aos coeficientes do polinômio primitivo, c_r , a seguir somadas para gerar a seqüência seguinte que realimentará o primeiro estágio do registro de

deslocamento. As operações de soma e produto podem ser efetuadas segundo as tabelas de somas e produtos de Barker (1986). Os registros de deslocamento devem ser inicializados com valores dentro do Campo de Galois diferentes de zero, pois de outro modo gerariam uma seqüência nula.

3.4 Análise de harmônicos

Sinais pseudo-aleatórios multi-níveis podem ser usados no projeto de identificação tanto de sistemas lineares quanto não lineares, dado que permitem a caracterização de sistemas lineares em presença de não linearidades ou mesmo a caracterização de não linearidades, mediante uma cuidadosa escolha dos harmônicos no sinal de excitação.

3.4.1 Conteúdo de harmônicos de sinais pseudo-aleatórios multi-níveis

Quando um sinal pseudo-aleatório multi-nível $u(t)$ é projetado como sinal de excitação na identificação de sistemas, é a análise do conteúdo de harmônicos quem dita as vantagens sobre outros tipos de sinais de excitação. Retornando à Figura 3.7, que mostra o processo de geração do sinal de seqüência m multi-nível. No processo de geração, a seqüência de comprimento máximo $\{x_i\}$ é mapeada na seqüência $\{u_i\}$ multi-nível segundo um conteúdo de harmônicos especificado pelo usuário, para depois ser convertido pelo segurador de ordem zero (ZOH) no sinal $u(t)$.

O conteúdo de harmônicos do sinal $u(t)$, para um período $N\Delta t$ (por simplicidade de notação nesta sub-seção adota-se $N_s = N$), é obtido através de Séries de Fourier:

$$u(t) = \sum_{k=-\infty}^{\infty} C_k e^{\frac{2\pi jkt}{N\Delta t}} \quad (3.12)$$

em que o coeficiente C_k define a magnitude e a fase do k -ésimo harmônico de $u(t)$ com frequência $k/N\Delta t$ Hz. Este coeficiente está relacionado à transformada discreta de Fourier $\{U_k\}$ da seqüência pseudo-aleatória $\{u_i\}$:

$$U_k = \frac{1}{N} \sum_{i=0}^{N-1} u_i e^{-j\frac{2\pi ki}{N}} \quad (3.13)$$

através da resposta em frequência do ZOH, que converte $\{u_i\}$ em $u(t)$:

$$C_u(k) = \frac{\text{sen}(\pi k/N)}{\pi k/N} e^{-j\frac{\pi k}{N}} \cdot U_k/N \quad (3.14)$$

em que a potência do sinal pseudo-aleatório $u(t)$ é definida por $|C_u(k)|^2$. Se a seqüência de comprimento máximo $\{u_i\}$ é real e de período N , então a transformada discreta de Fourier, U_k , tem a seguinte propriedade:

$$|U_k| = |U_{N-k}| \quad (3.15)$$

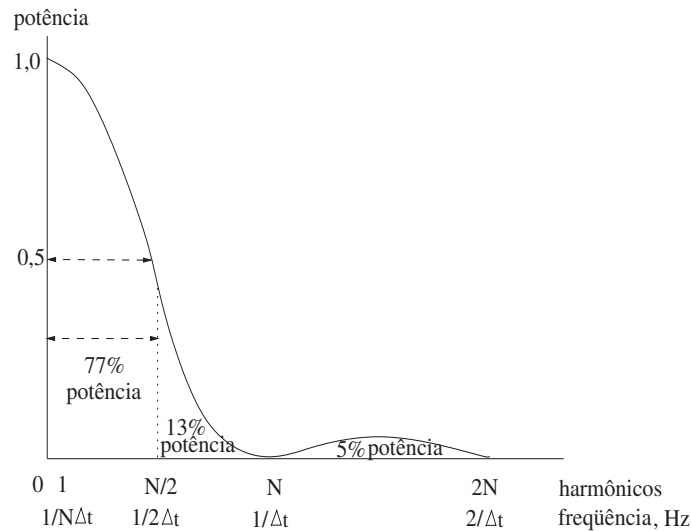


Fig. 3.8: Espectro de potência após o ZOH.

e se as amplitudes e potências dos primeiros $N/2$ harmônicos do sinal $u(t)$ são uniformes, então as amplitudes e potências dos harmônicos presentes também serão uniformes.

A equação (3.14) também mostra que a amplitude e a potência do k -ésimo harmônico do sinal de excitação contínuo $u(t)$ são atenuados pelo fator $|\text{sen}(\pi k/N)/(\pi k/N)|$. Neste caso, se as magnitudes de U_k são uniformes, então 77% da potência total do sinal de excitação $u(t)$ estará contida nos primeiros $N/2$ harmônicos, 13% nos seguintes $N/2$ harmônicos e 10% nos remanescentes (Barker e Godfrey, 1999). Porém, para identificação de sistemas só os primeiros $N/2$ harmônicos do sinal pseudo-aleatório $u(t)$ serão de interesse, como apreciado graficamente na Figura 3.8 que representa o espectro de potência após a correção inserida pelo ZOH.

O requisito de considerar os primeiros $N/2$ harmônicos resulta do fato de que os sinais são amostrados a intervalos T_{sw} (denotados aqui por Δt), ocorrendo harmônicos em frequência múltiplos da frequência fundamental $1/N\Delta t$. Então a faixa de frequência do sistema sob teste não deve exceder o $N/2$ -ésimo harmônico ou a frequência $f = 1/2\Delta t$, a denominada frequência de Nyquist (Shanon), para que o efeito *aliasing* seja evitado. Portanto, a atenção deve ser focada sobre a seqüência da transformada de Fourier discreta $\{U_k\}$, para determinar o conteúdo de harmônicos do sinal pseudo-aleatório $u(t)$ e da seqüência pseudo-aleatória $\{u_i\}$.

3.4.2 Supressão de harmônicos por mapeamento

Em geral, o processo de projeto de um sinal pseudo-aleatório para identificação de sistemas, mediante análise de harmônicos, envolve a escolha de números reais que sejam elementos do campo de Galois $GF(q)$ e que mapeados na seqüência $\{u_i\}$, todos os harmônicos que sejam múltiplos de um número primo ν estarão ausentes, isto é (Barker, 1993):

$$U_{i\nu} = 0 \quad , \forall i \quad (3.16)$$

Esta equação pode ser sempre satisfeita se ν é um fator de $(q - 1)$ quando:

$$\sum_{r=0}^{\nu-1} u(g^{h+r(q-1)/\nu}) = 0 \quad , h = 0, 1, \dots, (q-1)/\nu - 1 \quad (3.17)$$

onde g é um elemento primitivo e suas potências $0, g^0, g^1, \dots, g^{q-2}$ geram todos os elementos do campo. A equação (3.17) mostra que se ν é um fator de $(q - 1)$, então este é também fator dos múltiplos de $(q - 1)$. Esta envolve só aqueles números reais mapeados a partir de elementos não nulos de $GF(q)$ com o requisito:

$$u(0) = 0 \quad (3.18)$$

Se os mapeamentos satisfazem as relações (3.17) e (3.18), então a seqüência pseudo-aleatória apresentará a seguinte propriedade:

$$\sum_{r=0}^{\nu-1} u(i + rN/\nu) = 0 \quad (3.19)$$

Exemplo 3.4.1. Supondo que é requerido projetar um sinal pseudo-aleatório em que todos os harmônicos múltiplos de 2 e 3 sejam suprimidos, então o menor campo finito $GF(q)$ em que a seqüência m pode ser gerada será:

$$(q - 1) = \nu_1 \times \nu_2 = 2 \times 3$$

resultando $q = 7$. Um dos elemento primitivos de $GF(7)$ é 3, portanto as equações (3.17) e (3.18) proverão os requisitos para mapear os elementos de campo $GF(7)$. Um mapeamento que satisfaz estes requisitos é

$$\begin{aligned} u(0) &= 0; & u(3^2) &= u(2) = 2; & u(3^5) &= u(5) = -2; \\ u(3^0) &= u(1) = 1; & u(3^3) &= u(6) = -1; \\ u(3^1) &= u(3) = 3; & u(3^4) &= u(4) = -3; \end{aligned}$$

□

Para projetar um sinal com supressão de harmônicos múltiplos de 2, 3 e 5, será necessário um campo finito de pelo menos 31 elementos, $GF(31)$. Mapeamentos para gerar sinais pseudo-aleatórios, tanto para supressão de harmônicos múltiplos de 2 quanto para supressão de harmônicos múltiplos de 2 e 3, podem ser consultados em Barker e Zhuang (1997a).

Sinais com supressão de harmônicos múltiplos de 2

Sinais pseudo-aleatórios com supressão de harmônicos múltiplos de 2 podem ser projetados para o campo $GF(q)$ com qualquer $q \geq 2$ ímpar que satisfaça os requisitos (3.17) e (3.18), por exemplo com $q = 2, 3, 5, 7, 9$ ou 11. Sinais pseudo-aleatórios com harmônicos múltiplos de 2 suprimidos possuem a característica de repetição inversa cuja propriedade de simetria e função de autocorrelação

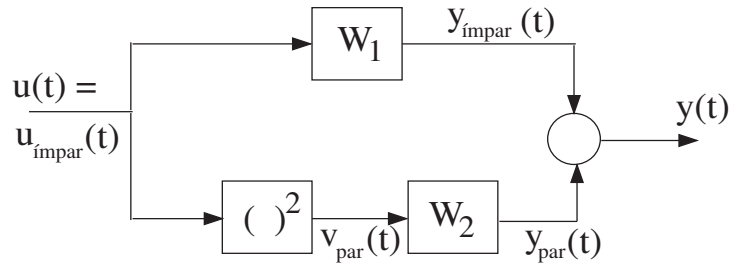


Fig. 3.9: Modelo de Hammerstein.

correspondem às relações (3.2) e (3.3) mostradas anteriormente.

Note que sinais com supressão de harmônicos múltiplos de 2 são um caso particular de (3.16), em que $\nu = 2$:

$$U_{2i} = 0, \forall i \quad (3.20)$$

correspondentes à supressão de harmônicos múltiplos de 2. Em geral, as magnitudes $|U_k|$ do sinal pseudo-aleatório tem a seguinte propriedade no domínio da frequências (Barker e Zhuang, 1997b):

$$|U_k| = |U_{\nu+k}| = |U_{\nu-k}| \quad (3.21)$$

cuja equivalência para supressão de harmônicos múltiplos de 2 é:

$$|U_k| = |U_{2+k}| = |U_{2-k}| \quad (3.22)$$

As equações (3.2) e (3.20) garantem que harmônicos pares sejam suprimidos em $u(t)$, mas isto não garante que os harmônicos ímpares sejam uniformes. Para isto é necessário o requisito adicional:

$$|U_k| = |U_1|, \forall k \text{ ímpar} \quad (3.23)$$

Sinais pseudo-aleatórios com harmônicos múltiplos de 2 suprimidos são especialmente úteis quando o sistema sob teste é não linear. Para este caso, o sinal de saída $y(t)$ apresentará componentes que estão relacionados ao sinal de excitação contínua $u(t)$ mediante relações não lineares. Sejam os componentes devido a não linearidades de ordem ℓ denotados por $y_\ell(t)$, então pode-se escrever:

$$y(t) = \sum_{\ell=0}^{\infty} y_\ell(t) = \sum_{\ell=0}^{\infty} y_{2\ell}(t) + \sum_{\ell=0}^{\infty} y_{2\ell+1}(t) \quad (3.24)$$

Desta forma, componentes pares do sinal de saída podem ser separados dos componentes ímpares. Para ilustrar a relação (3.24), aplica-se um sinal com supressão de harmônicos múltiplos de 2 no modelo de Hammerstein da Figura 3.9, como realizado por Barker et al. (2000). Pode-se apreciar como os componentes do sinal pseudo-aleatório de repetição inversa (harmônicos múltiplos de 2 suprimidos) são separados em componentes pares e ímpares. Note que os blocos W_1 e W_2 representam subsistemas lineares enquanto o bloco que antecede W_2 representa a não linearidade quadrática.

Mas o mais importante, Barker et al. (2000) mostraram que sinais pseudo-aleatórios de 3 níveis obtidos de $GF(3)$ são apropriados como sinais de excitação para identificar a parte linear e quadrática do modelo de Hammerstein. Esta conclusão está baseada na observação de que sinais com menos de 3 níveis tendem a produzir uma saída constante da não linearidade quadrática, porém resultam não sendo adequadas para o objetivo de identificação mencionado. Outra observação é que sinais com mais de 3 níveis e harmônicos múltiplos de 2 suprimidos não necessariamente produzem harmônicos de amplitude uniforme.

Estudos de caso na obtenção de modelos lineares a partir da identificação de processos químicos (não lineares) usando sinais pseudo-aleatórios com supressão de harmônicos pares são mostrados em Srinivasan e Rengaswamy (1999), demonstrando que estes sinais têm um melhor desempenho na identificação quando comparados com o obtido usando sinais PRBS. Os modelos lineares obtidos usando sinais com supressão de harmônicos múltiplos de 2 mostraram também melhorar o desempenho de controladores preditivos (MPC) aplicados a esses processos.

3.5 Medida de qualidade para um sinal de excitação

Os requisitos de um sinal de excitação dependem da aplicação pretendida. Assim, na identificação de sistemas lineares a persistência da excitação tem que ser garantida junto com a magnitude da amplitude, que deve ser pequena o suficiente para evitar os efeitos de não linearidades e ao mesmo tempo ser grande o suficiente para distinguir incertezas de medição ou outras fontes de erro (Godfrey et al., 1999). A isto agrega-se o requisito de maior potência possível para harmônicos especificados, no caso de redução de distorções por efeitos de não linearidades.

Na identificação de sistemas não lineares é necessário considerar além da persistência da excitação, a resposta do modelo para diferentes amplitudes da entrada (Pearson e Ogunnaike, 1997). Portanto interessa a "forma" do sinal (que deve resultar em uma função de autocorrelação similar ao ruído branco) e também como as amplitudes estão distribuídas para excitar diferentes pontos de operação dentro do intervalo máximo e mínimo do sinal.

A estes requisitos, pode-se adicionar o de sinais com poucos níveis para aplicações em atuadores de plantas industriais. Um índice de desempenho do sinal de excitação que considera as restrições físicas em atuadores de plantas industriais é o denominado índice *friendliness* (Doyle et al., 1999). É definido para qualquer seqüência $\{u_k\}$ em forma de porcentagem:

$$f = 100 \times \left(1 - \frac{n_T}{N - 1}\right) \quad (3.25)$$

onde N é o comprimento do sinal e n_T representa o número de transições do sinal de entrada (por exemplo em situações em que $u_k \neq u_{k-1}$). Conforme esta medida, uma seqüência constante é "100% apta" para atuadores, enquanto que uma seqüência que varia a qualquer instante é "0% apta" para atuadores. É possível obter uma expressão para o caso de sinais estocásticos. Assim f resulta "0% apta" para sinais ruído branco gaussiano, demonstrando as dificuldades de ordem prática para estes sinais.

Outros índices de desempenho que proporcionam uma medida do grau de dispersão do sinal é o PIPS (*Performance index for perturbation signals*) e o Fator de Pico (Barker e Zhuang, 1997b).

3.6 Critérios e parâmetros para o projeto de sinais MLPRS

O projeto de sinais pseudo-aleatórios multi-níveis deve satisfazer dois requisitos: o sinal deve ser persistentemente excitante e, relacionado com a identificação não linear, o sinal deve ter forma e amplitude adequadas para excitar dinâmicas não lineares.

Na prática, estes requisitos são difíceis de implementar em plantas industriais porque os atuadores impõem restrições físicas na forma e amplitude dos sinais de excitação, sendo preferível um sinal multi-nível com o menor número de níveis (Barker e Godfrey, 1999). Um tratamento teórico deste problema, baseado no conhecimento *a priori* do sistema a identificar, provê parâmetros de projeto para sinais de excitação (Gaikwad e Rivera, 1996; Braun et al., 1999). Desta forma os sinais podem ser projetados com características de sinais *plant-friendly* (ver seção 3.1), excitar a faixa de frequência relevante para controle e satisfazer requisitos de sinais de excitação para a identificação de sistemas. Sinais com tais características serão doravante denominados sinais relevantes para controle.

3.6.1 Faixa de frequência relevante para excitação do sistema

A potência do sinal pseudo-aleatório multi-nível é obtida a partir da equação (3.14):

$$|C_u(k)|^2 = \frac{|U_k|^2}{N_s^2} \left| \frac{\text{sen}(\pi k/N_s)}{\pi k/N_s} e^{(-j\pi k/N_s)} \right|^2 \quad (3.26)$$

que define a faixa de frequência útil para a excitação do sistema. Como visto na seção 3.4.1, se as magnitudes de U_k são uniformes, então o $N_s/2$ -ésimo harmônico não deve ultrapassar a largura de faixa do sistema. Na verdade, o espectro de potência do sinal consegue sua largura de faixa na frequência de potência metade, que corresponde aos primeiros $0,443N_s$ harmônicos (Barker e Godfrey, 1999) e que contém 72% da potência total. Em consequência, o limite de frequência superior do sinal de período $N_s T_{sw}$ (aqui T_{sw} é também o tempo mínimo entre mudanças de níveis do sinal), não deve ultrapassar $2\pi(0,443)/T_{sw}$. Se o espaçamento entre harmônicos é $1/N_s T_{sw}$, então o primeiro ponto em frequência correspondem a $2\pi/N_s T_{sw}$. Portanto, a largura de faixa do sinal de excitação é (Gaikwad e Rivera, 1996):

$$\frac{2\pi}{N_s T_{sw}} \leq w_s \leq \frac{2,78}{T_{sw}} \quad (3.27)$$

3.6.2 Faixa de frequência relevante para controle

O comportamento dinâmico de processos industriais é caracterizado por apresentar uma ampla gama de respostas dinâmicas para diferentes partes do processo. Os processos de tratamento de esgotos não são a exceção. Por exemplo, processos de precipitação química podem responder na casa de segundos, enquanto o processo de remoção de nutrientes pode ter constantes de tempo na ordem de minutos a horas. Para propósitos de controle do processo pode não ser importante modelar os modos mais rápidos do sistema, que corresponde à região de frequência alta. Portanto, interessa modelar o processo até a frequência de corte, que corresponde à frequência em que o módulo da resposta em frequência de malha fechada é 3 dB abaixo de seu valor na frequência zero, na qual o processo atinge sua largura de faixa.

Se uma estimativa *a priori* das constantes de tempo dominantes do processo é obtida, então pode ser definida a seguinte largura de faixa relevante para controle (Gaikwad e Rivera, 1996):

$$\frac{1}{\beta_s \tau_{dom}^l} \leq w_s \leq \frac{\alpha_s}{\tau_{dom}^c} \quad (3.28)$$

onde τ_{dom}^c é a constante de tempo dominante mais curta e τ_{dom}^l é a constante de tempo dominante mais longa. α_s e β_s são parâmetros de projeto usados para especificar a região de frequência importante para controle.

3.6.3 Parâmetros para o projeto do sinal

Para o projeto do sinal relevante para controle é necessário uma estimativa da seguinte informação *a priori*: a velocidade de resposta em malha fechada, as constantes de tempo dominantes, a ordem da não linearidade do modelo candidato e as amplitudes do sinal que sejam apropriadas para os atuadores. É com esta informação que podem ser calculados os parâmetros de projeto, listados a seguir:

- *Largura de faixa do sinal de excitação, w_s* : definida segundo a largura de faixa para controle como:

$$w_{inf} = \frac{1}{\beta_s \tau_{dom}^l} \leq w_s \leq \frac{\alpha_s}{\tau_{dom}^c} = w_{sup} \quad (3.29)$$

aqui β_s é o fator relacionado ao tempo de acomodação do processo e especifica quanta informação de baixa frequência estará presente no sinal. Valores de β_s iguais a 3, 4 ou 5 correspondem a 95%, 98% ou 99% do tempo de acomodação, respectivamente. Por exemplo, um valor de $\beta_s = 5$ enfatizará informação de frequência mais baixa. α_s determina o conteúdo em alta frequência do sinal e é uma medida de quão rápida é a velocidade de resposta em malha fechada em relação à resposta em malha aberta. Por exemplo, um valor de $\alpha_s = 2$ implica que a velocidade de resposta em malha fechada é duas vezes mais rápida que a de malha aberta. Valores típicos destes parâmetros são $\alpha_s = 2$ ou 3 e $\beta_s = 5$.

- *Número de níveis, m* : é determinado pelo grau de não linearidade estimado e pela supressão de harmônicos requerida. Assim, o número de níveis mínimo deve ser maior que o grau de não linearidade em pelo menos uma unidade ($m \geq \ell + 1$), como recomendado em Barker e Zhuang (1997a). O número de elementos de campo, q , não necessariamente é igual ao número de níveis, m , que depende do mapeamento adotado para supressão de harmônicos, como observado nos mapeamentos com supressão de harmônicos de Barker e Zhuang (1997b).
- *Tempo de comutação, T_{sw}* : o tempo mínimo entre mudanças de níveis do sinal é escolhido, comparando as frequências de interesse para controle da equação (3.28) com as frequências úteis para excitação do processo da equação (3.27), como sendo:

$$T_{sw} \leq \frac{2,78 \cdot \tau_{dom}^c}{\alpha_s} \quad (3.30)$$

- *Comprimento mínimo do sinal*: comparando a região de frequência útil para controle com a região de frequência relevante para excitação do sistema das equações (3.28) e (3.27), é obtida uma relação para o comprimento mínimo do sinal, de período NT_{sw} :

$$N = q^{n_r} - 1 \geq \frac{2\pi\beta_s\tau_{dom}^l}{T_{sw}} \quad (3.31)$$

cujo número de estágios do registro de deslocamento, n_r , deve ser escolhido para satisfazer esta relação. Note que o limite de baixa frequência pode ser obtido a partir das equações (3.29) e (3.31), como sendo:

$$w_{inf} \geq \frac{2\pi}{T_{sw}(q^{n_r} - 1)} \quad (3.32)$$

- *Número de harmônicos não suprimidos*: o número de harmônicos presentes dentro da largura de faixa de interesse é dado por Braun et al. (2001):

$$\# \text{ harm.} = \frac{2,78(q^{n_r} - 1)\gamma}{2\pi} \quad (3.33)$$

onde γ é definido segundo os harmônicos a serem suprimidos. Assim, para suprimir harmônicos múltiplos de 2, temos $\gamma = 1/2$. Para suprimir harmônicos múltiplos de 2 e 3, temos $\gamma = 1/3$ e para suprimir harmônicos múltiplos de 2, 3 e 5, temos $\gamma = 4/5$. Como visto na seção 3.4, a supressão de harmônicos possibilita a identificação de não linearidades ou a redução de distorções por não linearidades na identificação de sistemas lineares. Conforme Braun (2001), para a identificação de dinâmicas lineares interessa que o maior número de harmônicos estejam presentes, garantindo assim a persistência da excitação. Neste caso, recomenda-se usar sinais com todos os harmônicos presentes, como mostrado em Godfrey et al. (1999).

- *Tempo de duração de um período*, T_{cyc} : Para qualquer sinal baseado em seqüências m , o T_{cyc} é definido como:

$$T_{cyc} = T_{sw}(q^{n_r} - 1) \quad (3.34)$$

- *Escolha do mapeamento*: a função de autocorrelação discreta ($R_{uu}(k)$), o espectro do sinal de entrada e os harmônicos a serem suprimidos são fortemente influenciados pela escolha do mapeamento da seqüência m multi-nível. No presente trabalho são adotados os mapeamentos recomendados em Barker e Zhuang (1997a), para Campos de Galois de 3, 5, 7, 9, 11 e 13 elementos, com supressão de harmônicos múltiplos de 2, assim como para 7 e 13 elementos com supressão de harmônicos múltiplos de 2 e 3.

- *Deslocamento do sinal para múltiplas entradas*, D_s : em caso de vários sinais de entrada, requer-se introduzir sucessivos atrasos entre um sinal e outro, para serem descorrelacionados. Para um sistema com n_u entradas, cada entrada poderá ser deslocada em relação à anterior pelo

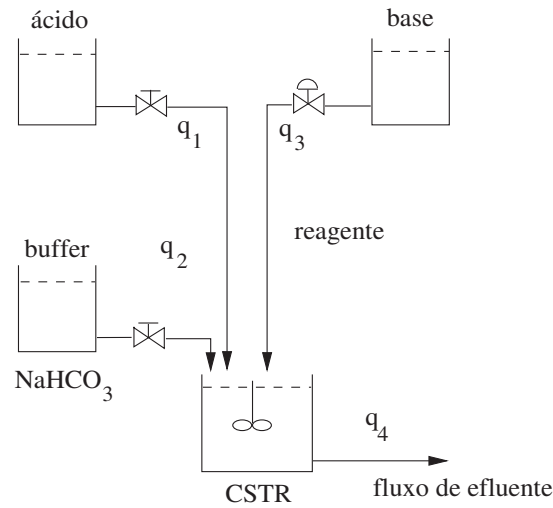


Fig. 3.10: Esquema do reator de neutralização de pH

atraso D_s , segundo $[D_s, 2D_s, 3D_s, \dots, (n_u - 1)D_s]$, onde D_s é escolhido segundo:

$$D_s \geq \beta_s \tau_{dom}^l \quad (3.35)$$

Para o caso de várias entradas injetadas simultaneamente o comprimento do sinal será:

$$N_s = \max\left(\frac{2\pi\beta_s\tau_{dom}^l}{T_{sw}}, \frac{n_u\beta_s\tau_{dom}^l}{T_{sw}}\right) \quad (3.36)$$

3.7 Exemplo de aplicação da metodologia

Nesta seção é mostrada a efetividade da metodologia para projeto de sinais pseudo-aleatórios multi-nível na identificação da dinâmica de um processo de neutralização de pH, tal como apresentado em Lara e Milani (2003).

O processo de neutralização de pH é com freqüência um processo difícil de identificar devido à sua característica inerentemente não linear, pois o ganho do processo pode variar em grandes ordens de magnitude para modestas faixas no valor de pH. O diagrama esquemático simplificado do processo de neutralização de pH é mostrado na Figura 3.10. O processo consiste de um fluxo de ácido ($q_1 = 0,996$ l/min), fluxo de uma solução tampão ou buffer ($q_2 = 0,033$ l/min) e um fluxo de base ($q_3 = 0,936$ l/min), componentes estes que são misturados em um reator. O modelo utilizado baseia-se na teoria das invariantes de reações de Gustafsson e Waller (1983), implementado em Sotomayor (1997).

Os sinais de excitação foram gerados pelo Gerador MLPRS, implementado em Simulink[®], que permite gerar sinais especificando o número de elementos do campo de Galois, a equação característica revisada e o mapeamento com a supressão de harmônicos desejada. Os diferentes polinômios primitivos e os mapeamentos podem ser obtidos através do software Galois[®] (Barker e Zhuang, 1997b).

Tab. 3.1: Condições de operação nominal do sistema (Henson e Seborg, 1997).

$[q_1] = 0,003 \text{ M HNO}_3$	$W_{a1} = 0,003 \text{ M}$
$[q_2] = 0,03 \text{ M NaHCO}_3$	$W_{b1} = 0 \text{ M}$
$[q_3] = 0,003 \text{ M NaOH},$	$W_{a2} = -0,03 \text{ M}$
$5 \times 10^{-5} \text{ M NaHCO}_3$	$W_{b2} = 0,03 \text{ M}$
$K_{a1} = 4,47 \times 10^{-7} \text{ mol/l}$	$W_{a3} = -3,05 \times 10^{-3} \text{ M}$
$K_{a2} = 5,62 \times 10^{-11} \text{ mol/l}$	$W_{b3} = 5,00 \times 10^{-5} \text{ M}$
$K_W = 1,00 \times 10^{-14} \text{ (mol/l)}^2$	$W_{a4} = -4,32 \times 10^{-4} \text{ M}$
$h_4 = 14,0 \text{ cm}$	$W_{b4} = 5,28 \times 10^{-4} \text{ M}$
$A_4 = 207 \text{ cm}^2$	$\text{pH}_4 = 7$

3.7.1 Condições experimentais

O objetivo aqui é identificar um modelo para fins de controle, onde distúrbios na concentração do fluxo de ácido, c_{q1} , devem ser controlados manipulando a vazão de base. Portanto, os sinais de entrada para a identificação do processo de neutralização de pH são o fluxo de base, q_3 e a concentração do fluxo de ácido, c_{q1} , assumida como uma variável de distúrbio medida. A incerteza na medição de pH é simulada mediante ruído branco com potência 0,01 e desvio padrão 0,05. O atuador foi modelado como uma função de transferência de primeira ordem com ganho unitário e constante de tempo 0,1 min. O período de amostragem para medição é igual ao tamanho de passo do método de integração numérica, 0,1 min. As condições de operação nominal e os parâmetros do modelo podem ser vistos na Tabela 3.1, baseados no modelo de Henson e Seborg (1997).

3.7.2 Projeto dos sinais de excitação

O projeto dos sinais pseudo-aleatórios multi-níveis começa estimando as constantes de tempo dominantes. Para isto, as entradas são perturbadas utilizando uma série de degraus positivos e negativos de $\pm 5\%$ em c_{q1} , o que origina demandas máximas de reagente de $\pm 40\%$ em relação ao valor nominal de q_3 . Estas demandas máximas determinam as amplitudes "informativas" no sinal de excitação q_3 . τ_{dom}^l foi estimada em 2,325 min e o τ_{dom}^c foi estimado em 0,255 min, com base na resposta no domínio do tempo, obtida aplicando sucessivos degraus com as amplitudes descritas.

É assumido $\alpha_s = 2$, para um tempo curto de acomodação em malha fechada, em relação ao tempo de acomodação em malha aberta. Para colocar ênfase na informação de baixa frequência, é assumido $\beta_s = 7$, correspondente a um tempo de acomodação maior que 99%. A largura de faixa do sinal é calculada segundo equação (3.29), como sendo $0,061 \text{ rad/min} \leq w_s \leq 7,84 \text{ rad/min}$. O passo seguinte é calcular o tempo de comutação mínimo usando a equação (3.30), resultando um $T_{sw} = 0,4 \text{ min}$. Note que o tempo de comutação deve ser um múltiplo do tempo de amostragem do sinal ($T_s = 0,1 \text{ min}$).

A análise de baixa frequência é resumida na Tabela 3.2, conforme a equação (3.32), para vários pares (q, n_r) . A partir da Tabela 3.3, pode ser verificado que o sinal com $q = 7$ e $n_r = 3$ reúne

Tab. 3.2: Análise do limite de baixa frequência para n_r e q .

rad/min		q				
		3	5	7	11	13
n_r	2	2,618	0,872	0,436	0,175	0,125
	3	0,806	0,169	0,0612	0,0157	0,00953
	4	0,262	0,169	0,0087		
	5	0,0865	0,0067			
	6	0,0288				

Tab. 3.3: Número de harmônicos não suprimidos presentes dentro da largura de faixa de interesse com $\gamma = 1/2$.

# de harm.		q				
		3	5	7	11	13
n_r	2	2	5	10	26	37
	3	6	27	75	290	479
	4	17	136	524		
	5	53	682			
	6	159				

os requisitos de baixa frequência com o máximo número de harmônicos presentes. São incluídos só os resultados para os sinais com o maior número de harmônicos ($\gamma = 1/2$). Portanto, é escolhido $GF(7)$ com um polinômio primitivo de grau 3. Dado que é necessário aplicar dois sinais de excitação, o atraso do primeiro sinal em relação ao segundo é calculado através da equação (3.35), resultando $D_s \geq 11,63$ min. O máximo comprimento do sinal resulta $N_s = 340,86$, usando a equação (3.36).

Para um $GF(7)$ com grau do polinômio primitivo igual a 3, o pacote Galois[®] proporciona uma lista de polinômios primitivos que geram seqüências de comprimento máximo. Assim sendo, escolhe-se entre outros o seguinte polinômio primitivo:

$$f(x) = 1 + 2x + 4x^2 + 2x^3 \quad (3.37)$$

com a equação (3.9) e (3.10), a equação característica revisada resulta:

$$X = 5DX + 3D^2X + 5D^3X \quad (3.38)$$

onde o máximo grau do polinômio determina o número de estágios do registro de deslocamento, seus expoentes determinam as conexões de realimentação e seus coeficientes determinam os ganhos nas respectivas saídas, segundo o esquema da Figura 3.7. Estes valores são então usados no Gerador MLPRS para projetar sinais de excitação com supressão de múltiplos de 2 ($\gamma = 1/2$) e com supressão de harmônicos múltiplos de 2 e 3 ($\gamma = 1/3$). Os sinais pseudo-aleatórios multi-níveis são escalonados apropriadamente antes de serem aplicados à planta, isto com a finalidade de que os sinais projetados

possuam as amplitudes de operação do fluxo de base e da concentração do fluxo de ácido.

3.7.3 Algoritmo de identificação

O procedimento de identificação não linear é desenvolvido utilizando um método ortogonal para a resolução do problema de mínimos quadrados, o Método de Gram-Schmidt Modificado (Chen et al., 1989), que combina a estimação de parâmetros com a seleção da estrutura. O modelo adotado aqui é uma formulação estendida do modelo não linear autoregressivo de média móvel e entrada externa (NARX -*Non linear AutoRegressive and eXogeneous input*):

$$\mathbf{y}(t) = \mathbf{b} + \mathbf{f}^\ell[\mathbf{y}(t-1), \dots, \mathbf{y}(t-N_y), \mathbf{u}(t-1), \dots, \mathbf{u}(t-N_u), \mathbf{d}(t-1), \dots, \mathbf{d}(t-N_d)] + \mathbf{e}(t) \quad (3.39)$$

onde \mathbf{b} representa o vetor com os termos de polarização, \mathbf{y} , \mathbf{u} e \mathbf{d} representam os vetores com os termos para a saída, entrada e distúrbio medido, respectivamente. N_y , N_u e N_d representam os atrasos máximos da saída, entrada e distúrbio medido. O expoente ℓ , caracteriza o grau de não linearidade do modelo e \mathbf{e} é um vetor representando o erro de modelagem e as incertezas do modelo.

A estrutura do modelo NARX inclui termos não lineares de parâmetros lineares. Portanto é possível a resolução numérica do problema de mínimos quadrados para obter modelos polinomiais de diversas ordens. Um resumo do algoritmo ortogonal multivariável de Gram-Schmidt Modificado é apresentado no Apêndice C, que corresponde a uma extensão do algoritmo para sistemas de uma entrada e uma saída implementado em Lara (1999).

3.7.4 Avaliação de desempenho

Usando os sinais de excitação MLPRS para supressão de harmônicos múltiplos de 2 e para supressão de harmônicos múltiplos de 2 e 3, com $m = 3, 5$ e 7 níveis, são identificados modelos fixando graus de não linearidade $\ell = 1, 2$ e 3 no algoritmo de identificação não linear. A avaliação de desempenho utiliza os seguintes índices para avaliar o ajuste da saída do modelo identificado com respeito à saída da planta de pH: Somatório de Erro Quadrático (RMS), Erro Máximo (EM) e o Critério de Informação de Akaike (AIC).

Os menores índices de erro são obtidos para o sinal com $q = 7$, $n_r = 3$, $m = 3$ e com harmônicos múltiplos de 2 suprimidos. Os resultados para os graus de não linearidade $\ell = 1, 2$ e 3 podem ser verificados na parte superior da Tabela 3.7. Na Figura 3.11 mostra-se a saída da planta e o sinal de excitação pseudo-aleatório multi-nível projetado para a vazão de base atrasada em $D_s \geq 11,63$ min respeito de c_{q1} , a entrada de distúrbio medido. Este atraso é calculado segundo a relação (3.35), usado para excitar o sistema com entradas descorrelacionadas.

A Figura 3.12 mostra a potência espectral modificada após o ZOH. Verifica-se que 77% da potência total do sinal de excitação está contida na largura de faixa do sistema, $0,061 \leq w \leq 7,754$ rad/min, como indicado com o quadrado de traço contínuo.

Os parâmetros e os regressores do modelo mínimo resultante do algoritmo de identificação não linear são mostrados na Tabela 3.4, para grau de não linearidade $\ell = 2$, atrasos na saída, entrada e distúrbio medido $N_y = N_u = N_d = 2$ respectivamente e para $d = 1$. Note também, a partir da Tabela 3.7, que os melhores resultados são obtidos para os graus de não linearidade $\ell = 1$ e 2 . A validação

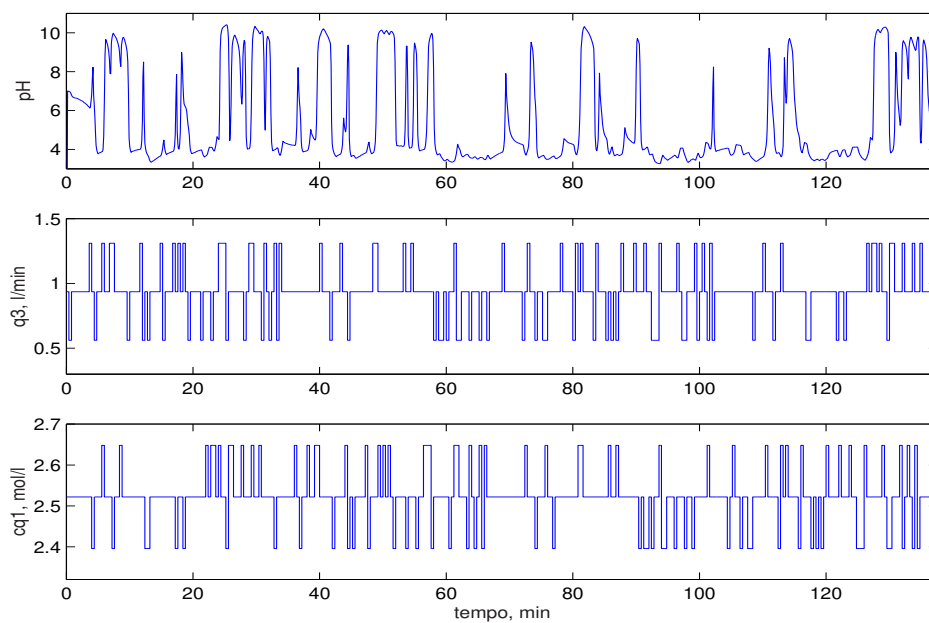


Fig. 3.11: Saída da planta e sinais de excitação para $q = 7$, $n_r = 3$ e $m = 3$.

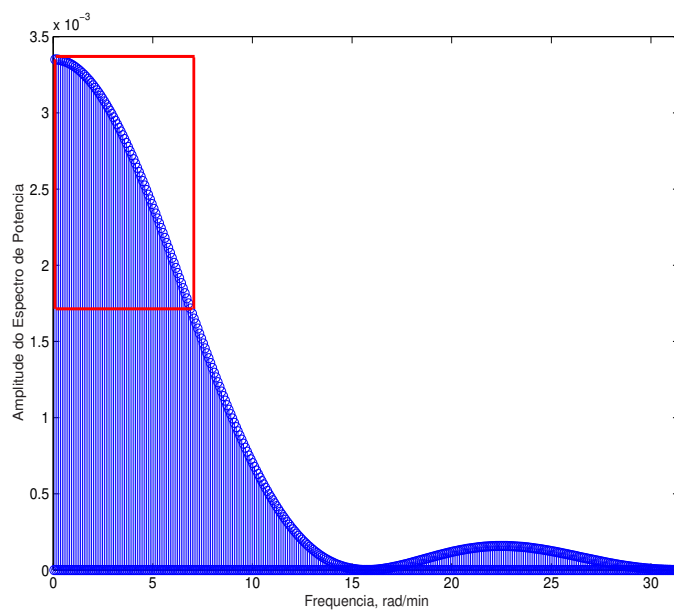


Fig. 3.12: Espectro de potência do sinal MLPRS para $q = 7$, $n_r = 3$ e $m = 3$.

Tab. 3.4: Parâmetros do modelo NARX não linear.

elemento	termo	θ_i
1º	1	-0,0002
2º	$u(t - 2)$	0,0161
3º	$y(t - 1)$	1,5245
4º	$y(t - 2)$	-0,5502
5º	$d(t - 2)$	0,0267
6º	$u(t - 3)$	-0,1698
7º	$y(t - 1)d(t - 2)$	0,0929
8º	$y(t - 1)u(t - 3)$	0,0744
9º	$y(t - 2)u(t - 3)$	0,0136
10º	$y(t - 1)y(t - 2)$	0,0067
11º	$y(t - 1)^2$	0,0130
12º	$u(t - 2)d(t - 1)$	0,1061

do modelo NARX estimado e a saída da planta para $\ell = 2$, apresentados na Figura 3.13, mostra a boa aderência que é obtida com a aplicação do sinal MLPRS.

O modelo identificado também é submetido a um outro procedimento de validação com um novo conjunto independente de dados, como sendo perturbações simétricas do tipo degrau positivos e negativos nas duas entradas (q_3 e c_{q1}). As perturbações são aplicadas simultaneamente depois do período de identificação, isto é, para o intervalo de validação $137,6 \text{ min} < t < 205 \text{ min}$. Estas perturbações simulam situações realistas em processos de pH. Verifica-se também a boa aderência entre a saída da planta com a saída do modelo não linear identificado para estas duas novas excitações, como observado na Figura 3.14.

Comparação com sinais PRBS

Para mostrar as vantagens de sinais de excitação MLPRS na identificação de sistemas não lineares, seu desempenho é comparado com o desempenho de sinais PRBS, usando o mesmo algoritmo de identificação. Neste caso, $n_r = 8$ satisfaz o requisito do limite de baixa frequência. O comprimento máximo do sinal é $2^{n_r} - 1$, resultando um $T_{cyc} = 102 \text{ min}$ e um atraso entre sinais $D_s = 16, 28 \text{ min}$.

Em forma similar, pode ser obtida a seguinte equação característica:

$$X = D^4 X + D^5 X + D^6 X + D^8 X \quad (3.40)$$

A partir dos índices de erro das Tabelas 3.5 e 3.6, verifica-se o desempenho inferior dos sinais PRBS em relação aos sinais MLPRS das Tabelas 3.7 e 3.8, especialmente quando é considerada

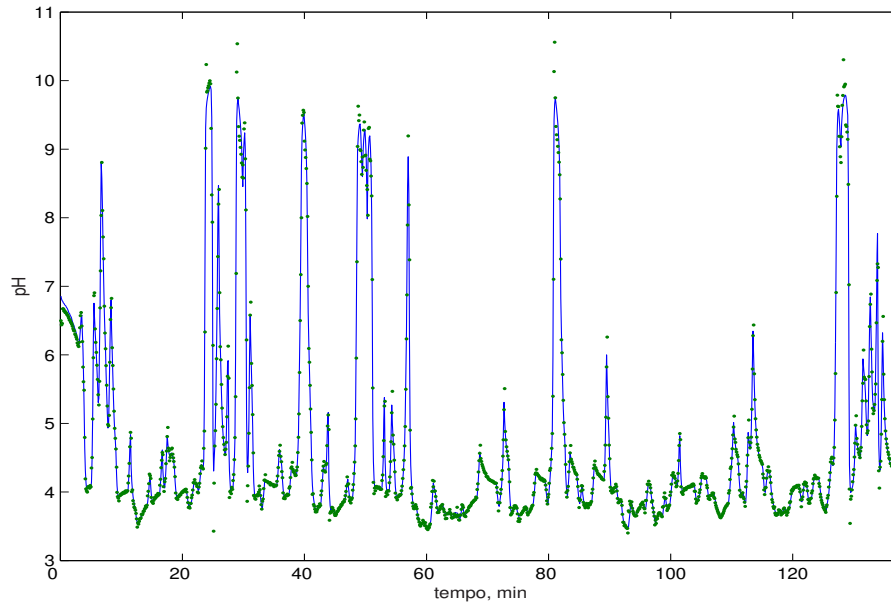


Fig. 3.13: Saída do modelo identificado (pontos) e saída da planta (traço contínuo).

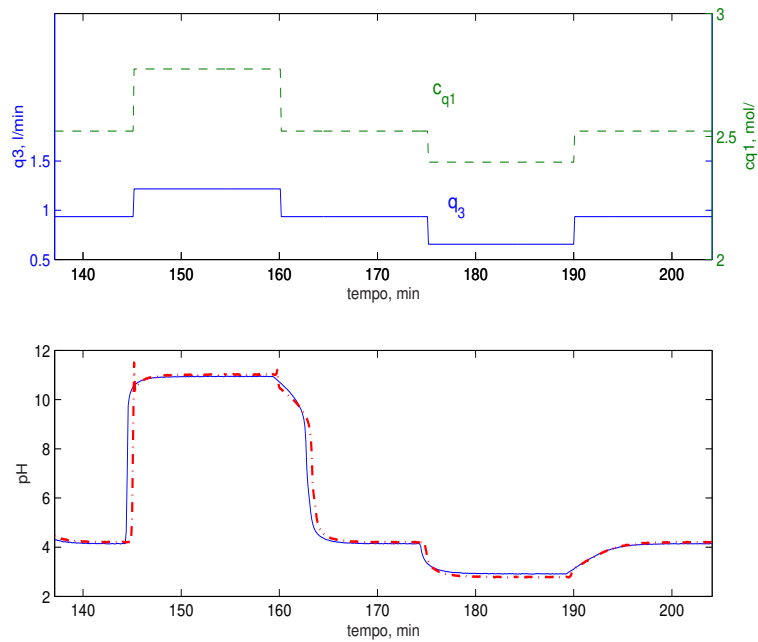


Fig. 3.14: Conjunto independente de dados para c_{q1} e q_3 (grafo superior). Validação entre a saída da planta (-) com a saída do modelo identificado (-) (gráfico inferior).

Tab. 3.5: Índices de erro usando o sinal PRBS sem considerar a dinâmica do atuador.

grau ℓ	Erro RMS	Erro Máximo	AIC
1	0,1233	0,0152	$-4,20 \times 10^3$
2	0,1201	0,0144	$-1,05 \times 10^3$
3	0,0447	0,0020	$-1,46 \times 10^3$

Tab. 3.6: Índices de erro usando sinais PRBS considerando a dinâmica do atuador.

grau ℓ	Erro RMS	Erro Máximo	AIC
1	0,1242	0,0154	$-1,04 \times 10^3$
2	0,1219	0,0149	$-1,04 \times 10^3$
3	0,0506	0,0026	$-1,42 \times 10^3$

a dinâmica do atuador. Evidentemente os sinais PRBS não conseguem excitar adequadamente a dinâmica do processo de pH.

Comparação com sinais aleatórios

Nesta seção é comparado o desempenho do sinal MLPRS com o desempenho de dois sinais aleatórios: sinal aleatório de distribuição uniforme e ruído branco filtrado. A comutação mínima para ambos sinais é 0,4 min para o mesmo comprimento do sinal MLPRS (137,6 min). Os resultados para o sinal MLPRS e para ambos tipos de sinais estocásticos estão resumidos na Tabela 3.7 e na Tabela 3.8. É verificado na Tabela 3.7, que o desempenho do sinal multi-nível pseudo-aleatório é relativamente superior. No entanto, na Tabela 3.8, onde considera-se a dinâmica do atuador, os índices de erros são significativamente menores, especialmente para o grau de não linearidade $\ell = 2$, razão pela qual é escolhido o modelo identificado com este grau de não linearidade.

Conclusões do Experimento

Foi mostrado que sinais de excitação pseudo-aleatórios de seqüência m multi-nível podem ser projetados utilizando diretrizes que, considerando o conhecimento *a priori* do sistema físico, conseguem atender com eficácia requisitos de sinais de identificação relevantes para controle. Características importantes como comprimento mínimo do sinal, persistência da excitação e potência espectral do sinal foram projetadas em uma aplicação para um processo de neutralização de pH, demonstrando sua capacidade de excitar dinâmicas não lineares através da escolha adequada do número de níveis e mapeamento dos sinais, que possam reduzir o efeito de distorções não lineares.

Os melhores resultados foram obtidos com o sinal de excitação com harmônicos múltiplos de 2 suprimidos, para graus de não linearidade de primeira e segunda ordem, demonstrando que o sistema

Tab. 3.7: Índices de erro entre o sinal pseudo-aleatório e os sinais aleatórios sem considerar a dinâmica do atuador.

	grau ℓ	Erro RMS	Erro Máximo	AIC ($\times 10^3$)
sinal MLPRS	1	0,0207	$4,28 \times 10^{-4}$	-2,62
	2	0,0186	$3,44 \times 10^{-4}$	-2,67
	3	0,0262	$1,83 \times 10^{-4}$	-9,86
Sinal aleatório distr. unif.	1	0,0436	0,019	-8,56
	2	0,0412	0,0017	-8,68
	3	0,0232	$5,3 \times 10^{-4}$	-10,2
Sinal ruído branco filtrado	1	0,0458	0,0021	-8,48
	2	0,0446	0,0020	-8,48
	3	0,0308	$9,5 \times 10^{-4}$	-93,80

Tab. 3.8: Índices de erro entre o sinal pseudo-aleatório e os sinais aleatórios considerando a dinâmica do atuador.

	grau 1	Erro RMS	Erro Máximo	AIC ($\times 10^4$)
sinal MLPRS	1	0,0047	$2,2 \times 10^{-5}$	-1,45
	2	0,0081	$6,6 \times 10^{-5}$	-1,31
	3	0,0520	$2,7 \times 10^{-5}$	-1,43
sinal aleat. distr. unif.	1	0,0419	0,0018	-8,66
	2	0,0406	0,016	-8,74
	3	0,0224	$5,0 \times 10^{-4}$	-1,02
sinal aleat. ruído branco filt.	1	0,0457	0,0021	-0,84
	2	0,0450	0,0020	-0,85
	3	0,0298	$8,9 \times 10^{-4}$	-0,95

possui majoritariamente dinâmicas não lineares de segunda e primeira ordem.

3.8 Identificação do processo de remoção de matéria carbonácea

O comportamento dos processos biotecnológicos que ocorrem em um bioreator apresentam uma complexidade sem paralelos na indústria química. Dentre os problemas na modelagem de sistemas de tratamento de esgotos estão a estrutura complexa do modelo e o grande número de estados e parâmetros a serem identificados (Jeppsson, 1996). Daí a importância na busca de modelos reduzidos quando o objetivo de modelagem é o controle baseado em modelo.

Dentre todos os processos de remoção de matéria orgânica em plantas de tratamento de esgotos o processo de lodos ativados é o mais utilizado. Este processo reproduz em uma unidade previamente projetada os fenômenos biológicos de autodepuração que ocorrem na natureza quando um efluente é descartado em um corpo receptor. Em outras palavras, no projeto de plantas de lodos ativados se busca a transformação da matéria orgânica em produtos mineralizados inertes, basicamente mediante a oxidação de carbono orgânico solúvel desenvolvida majoritariamente por bactérias heterotróficas aeróbias e, se dado o caso, também mediante a oxidação da matéria nitrogenada desenvolvida pelas bactérias autotróficas.

O conhecimento da dinâmica do processo de transferência de oxigênio dentro do bioreator é de fundamental importância. A manutenção do nível adequado de concentração de oxigênio dissolvido permite que se desenvolva o crescimento aeróbio dos microrganismos heterotróficos e autotróficos responsáveis pela remoção do substrato rapidamente biodegradável da matéria carbonácea. Portanto, o processo de oxigênio dissolvido constitui um indicador importante no processo de tratamento por lodos ativados. Se o objetivo é a remoção da matéria nitrogenada, será também importante garantir a condição anóxica do bioreator de pre-desnitrificação, isto é, a baixa concentração de oxigênio dissolvido no meio. Assim será possível que os microrganismos utilizem o nitrato como substância oxidada.

Devido a esta motivação, a presente seção tem por objetivo obter um modelo reduzido do processo de remoção de matéria carbonácea caracterizada pela dinâmica da concentração de oxigênio dissolvido.

3.8.1 Identificação do Processo de Oxigênio Dissolvido

O balanço completo da concentração de oxigênio dissolvido nos reatores aeróbios para o modelo ASWWTP-USP e que descreve o comportamento dinâmico para as zonas aeróbias k , sendo $k = 2, 3$, resulta na seguinte equação diferencial ordinária:

$$\frac{dS_{O,k}}{dt} = \frac{Q_{k-1}S_{O,k-1} - Q_k dS_{O,k}}{V_k} + r_{S_{O,k}} + K_l a (Q_{ar})_k \cdot (S_{O,sat} - S_{O,k}) \quad (3.41)$$

onde

- $S_{O,k}$: concentração de oxigênio dissolvido da zona aeróbia k .
 $r_{S_{O,k}}$: taxa de consumo de oxigênio da zona aeróbia k .
 $K_{l,a}(Q_{ar})$: função transferência de oxigênio.
 Q_{ar} : vazão de injeção de ar.
 Q_{k-1}, Q_k : vazão de licor misto de entrada e saída respectivamente da zona aeróbia k .
 V_k : volume da zona aeróbia k .

entende-se o licor misto como sendo o lodo mais água tratada e $r_{S_{O,k}}$, o negativo da concentração de oxigênio dissolvido mostrada em (2.26), representando o consumo de oxigênio pelos microrganismos.

A taxa de consumo de oxigênio é um indicador biológico muito importante e proporciona informação a respeito do estado atual das reações bioquímicas do processo de remoção de matéria orgânica. O que torna o problema não trivial é que $r_{S_{O,k}}$ é variante no tempo e $K_{l,a}$ resulta ser uma função não linear da Q_{ar} . Abordagens de estimação da dinâmica do oxigênio dissolvido existem, como a estimação simultânea de $r_{S_{O,k}}$ e $K_{l,a}$ mediante a introdução de perturbações do tipo liga-desliga no fluxo de ar (Suescon et al., 1998; Marsili-Libelli, 1990; Marsili-Libelli e Vaggi, 1997) e a estimação baseada no Filtro de Kalman usando as medições da concentração de oxigênio dissolvido e da vazão de injeção de ar (Lindberg, 1997).

O objetivo aqui é identificar um modelo não linear do tipo caixa-preta da dinâmica do oxigênio dissolvido, utilizando como entrada a vazão de injeção de ar e como saída a concentração de oxigênio dissolvido. Para um resultado confiável na identificação, a operação deve ser excitada de tal forma que todos os modos de operação de interesse sejam excitados. Isto cria requisitos para a amplitude e a duração destes sinais, assim como para a excitação da região de frequência que enfatize a largura de banda do sistema e não de ruídos de medição geralmente presentes em altas frequências. Os sinais de excitação multi-níveis são projetados a seguir seguindo o procedimento descrito na seção 3.6, para que satisfaçam estes requisitos. Desta forma, também é possível identificar um modelo linear parcimonioso mediante a análise do conteúdo de harmônicos do sinal de excitação, para que este possa ser utilizado, por exemplo, em projetos de controladores lineares adaptativos.

3.8.2 Projeto dos sinais de excitação da planta

Para o experimento de identificação os sinais devem estar delimitados em amplitude dentro dos intervalos de operação das variáveis de entrada. Os intervalos de operação são determinados pelas restrições operacionais do processo e pelas restrições físicas dos atuadores. A remoção biológica da matéria carbonácea ocorre para concentrações de oxigênio dissolvido maiores a 1,0 mgO₂/l, enquanto a formação de nitritos requer uma concentração superior a 2,0 mgO₂/l e a formação de nitratos acima de 5,5 mgO₂/l. Em consequência, se utiliza uma amplitude da vazão de injeção de ar na zona 3 do bioreator (Q_{ar3}) entre 0,01 e 0,09 m³/h, que origina variações na concentração de oxigênio dissolvido entre 0,5 e 5,0 mgO₂/l. Na zona 2, assume-se uma concentração constante igual 2 mgO₂/l. É adicionado ruído aditivo na medição da concentração de oxigênio dissolvido, um sinal aleatório de variância 0,01 e média nula.

Realizam-se então ensaios para excitar o espaço entrada-saída do processo mediante uma série de degraus positivos e negativos em Q_{ar3} , dentro da amplitude de operação assumida. As magnitudes do degrau são de $\pm 20\%$ do valor nominal (0,033 m³/h), com duração de 2 horas para cada degrau. A taxa máxima de amostragem dos sinais é de 0,01 h, que corresponde aproximadamente a $\frac{1}{10}$ da constante

Tab. 3.9: Análise do limite de baixa frequência para n_r e q .

rad/min		q				
		3	5	7	11	13
n_r	2	0,1309	0,0436	0,0218	0,0087	0,0062
	3	0,0403	0,0084	0,0031	0,0008	0,0005
	4	0,0131	0,0017	0,0004		
	5	0,0043	0,0003			

de tempo associada a uma mudança de 100% no valor nominal da concentração de oxigênio, isto é, 2,0 mgO₂/l. Esta constante de tempo é denotada por τ_n . São obtidas assim a constante de tempo dominante mais curta, $\tau_{dom}^c = 0,0837$ h e a constante de tempo mais longa, $\tau_{dom}^l = 0,1778$ h.

Selecionou-se o valor de $\alpha_s = 2$, assumindo que no processo de oxigênio dissolvido o tempo de acomodação em malha fechada seja mais rápido que o tempo de acomodação em malha aberta. Observou-se que depois de $4\tau_n$ o processo varia muito pouco, em relação à velocidade, para 1, 2 ou $3\tau_n$. Como se deseja a informação em baixa frequência, basta ter a informação de 98% do valor estacionário, e conseqüentemente seleciona-se o valor de $\beta_s = 4$. Portanto, a largura de banda determinada pela dinâmica do processo, à qual o sinal deverá se ajustar, resulta sendo $0,0234 \text{ rad/min} \leq w_s \leq 0,398 \text{ rad/min}$. O número de níveis deve ser $m \geq 3$, se o grau de não linearidade mínimo esperado for 2 ($m \geq \ell + 1$, segundo visto na sub-seção 3.6.3).

O tempo de comutação do sinal é calculado mediante a expressão (3.30), resultando $T_{sw} = 0,1$ h. A partir deste valor, conclui-se que o comprimento mínimo do sinal deve ser $N \geq 44,7$. Da mesma forma que no exemplo de aplicação, a escolha do sinal é feita mediante a análise de limite de baixa frequência, calculado para diferentes pares de valores do campo de Galois vs. o número de registros de deslocamento (q, n_r), como mostrado na Tabela 3.9. O par de valores que satisfaz o requisito do limite de baixa frequência da expressão (3.32) é (7, 2). Da análise do conteúdo de harmônicos dos sinais testados para $\gamma = 1/2, 1/3$ e $4/5$, pode-se verificar que o sinal com supressão de harmônicos múltiplos de 2 ($\gamma = 1/2$), possui maior quantidade de harmônicos na largura de faixa de interesse. Para comparação são mostrados os resultados para $\gamma = 1/2$ e $1/3$ nas Tabelas 3.10 e 3.11, respectivamente.

Em conseqüência, é selecionado o sinal com $q = 7$ e $n_r = 2$, que satisfaz os requisitos de largura de banda e de comprimento mínimo do sinal. Este sinal pode ser mapeado para supressão de harmônicos múltiplos de 2, como sugerido em Barker e Zhuang (1997a):

$$\begin{aligned} m = 3 : & \quad u(0) = 0; \quad u(1) = 0; \quad u(2) = 0; \quad u(3) = 1; \quad u(4) = -1; \quad u(5) = 0; \quad u(6) = 0; \\ m = 5 : & \quad u(0) = 0; \quad u(1) = 1; \quad u(2) = -2; \quad u(3) = 2; \quad u(4) = -2; \quad u(5) = 2; \quad u(6) = -1; \\ m = 7 : & \quad u(0) = 0; \quad u(1) = 2; \quad u(2) = -3; \quad u(3) = 6; \quad u(4) = -6; \quad u(5) = 3; \quad u(6) = -2; \end{aligned}$$

Com o par de valores (7, 2) o tempo de duração de um período do sinal de excitação multi-nível resulta ser de 4,8 h. Mediante o software Galois[®] escolhe-se o polinômio primitivo:

$$f(x) = 1 + x + 3x^2$$

Tab. 3.10: Número de harmônicos não suprimidos para $\gamma = 1/2$.

# de harm.	q					
	3	5	7	11	13	
n_r	2	2	5	11	27	37
	3	6	27	77	294	486
	4	18	138	531		
	5	34	691			

Tab. 3.11: Número de harmônicos não suprimidos para $\gamma = 1/3$.

# de harm.	q					
	3	5	7	11	13	
n_r	2	1	4	7	18	25
	3	4	18	50	196	324
	4	12	92	354		
	5	36	461			

de onde é obtida a equação característica revisada:

$$X = 6DX + 6D^2X + 4D^3X$$

Os sinais de excitação são escalados apropriadamente para terem as amplitudes dos regimes de operação descritos anteriormente, depois de serem gerados pelo bloco Gerador MLPRS implementado no Simulink[®].

3.8.3 Identificação e validação dos modelos

Os dados de entrada e saída são normalizados antes de serem usados pelo Método Ortogonal de Gram-Schmidt Modificado (Chen et al., 1989), que como resultado fornecerá os modelos não lineares ARX estendidos. É assumido que os sinais foram previamente processados para detecção e eliminação de "outliers" e tendências (Olsson e Newell, 1999).

A escolha do modelo mais apropriado é crucial para uma identificação de sucesso. Esta escolha está em função do tamanho e da complexidade do modelo. A melhor estrutura do modelo será então um compromisso entre flexibilidade e parcimônia (Ljung, 1999). Por outro lado, esta escolha dependerá também da aplicação do modelo.

Uma vez que a estrutura do modelo tenha sido escolhida, são realizados ensaios para diferentes ordens de atrasos N_y , N_u e d no algoritmo de identificação, dado que não se tem conhecimento *a priori*, da ordem do modelo. As saídas dos modelos candidatos são validadas mediante a utilização dos seguintes índices de erro: Somatório do Erro Quadrático (RMS), Erro Máximo (EM), Critério de Informação de Akaike (AIC) e de Bayes (BIC).

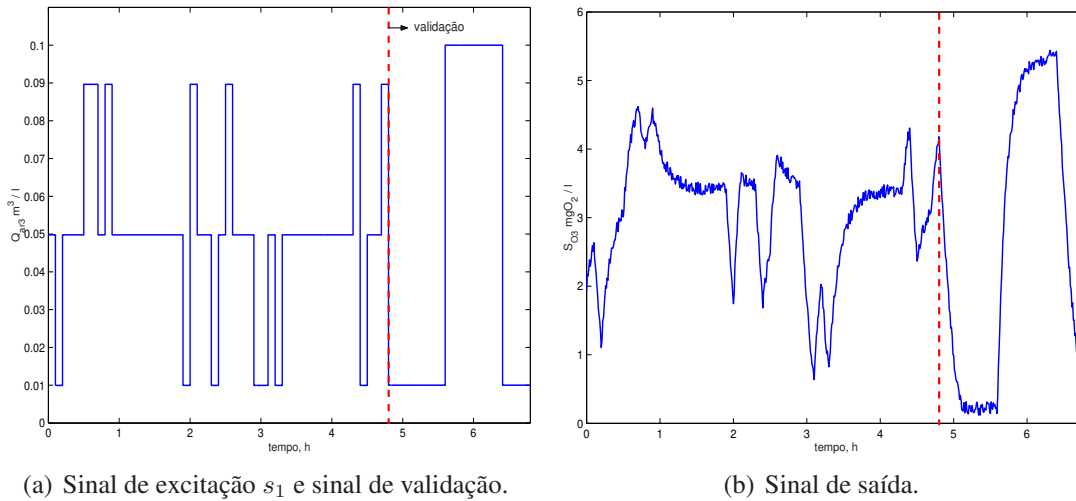


Fig. 3.15: Vazão de injeção de ar e concentração de oxigênio dissolvido na zona 3.

Tab. 3.12: Índices de erro para grau de não linearidade $\ell = 2$.

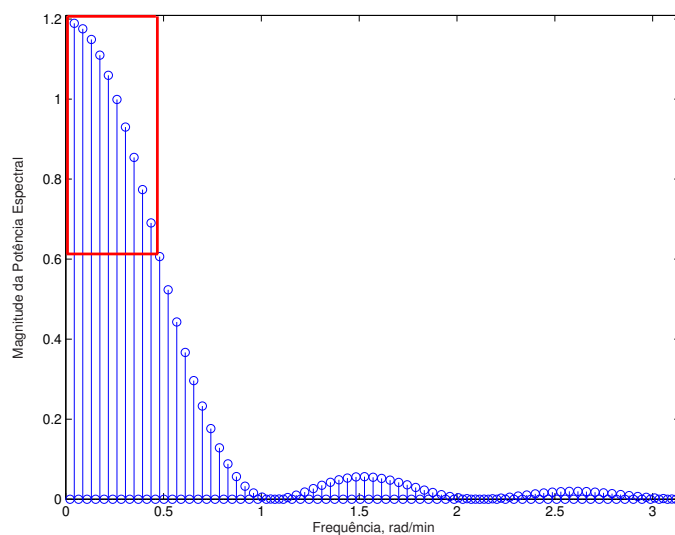
sinal multi-nível	# termos	Erro RMS	Erro Máximo	AIC	BIC
s_1	6	0,0186	$3,47 \times 10^{-4}$	$-3,70 \times 10^3$	$-3,68 \times 10^3$
s_2	17	0,0025	$6,00 \times 10^{-6}$	$-5,55 \times 10^3$	$-5,51 \times 10^3$

Para os ensaios de validação foram usados sinais MLPRS projetados para $m = 3, 5$ e 7 níveis dentro do $GF(7)$ projetado anteriormente. Os modelos identificados com os menores índices de erro resultaram pertencer às seguintes ordens de atrasos: $N_y = 6$, $N_u = 6$ e $d = 0$. Na Tabela 3.12 são mostrados os índices de erro para o grau de não linearidade 2. Aqui s_1 denota o sinal de excitação com $m = 3$ níveis e s_2 o sinal com $m = 5$ níveis. Na Figura 3.15 pode ser apreciada a realização do sinal s_1 e a respectiva resposta da concentração de oxigênio dissolvido. Nesta figura pode ser apreciada também a seqüência do sinal de excitação usada para a validação assumindo $Q_{in} = 0$. Esta seqüência liga-desliga simula os processos de aeração e de utilização de oxigênio pelos microrganismos (associada a r_{S_O}) na planta de lodos ativados.

Note também na Tabela 3.12 que o modelo obtido com o sinal s_2 apresenta a melhor capacidade de ajuste. Como mencionado anteriormente, a escolha do modelo é fortemente influenciada pela aplicação. Assim o modelo obtido com o sinal s_1 será adequado para aplicações de controle não linear baseado em modelo. Este modelo perde um pouco na precisão da predição, mas é um modelo parcimonioso. Enquanto que o modelo obtido com o sinal s_2 , de 17 termos, será adequado para estudos de entendimento e treinamento de operadores, onde a complexidade do modelo não impõe restrições computacionais como acontece no algoritmo de controle baseado em modelo. Na Tabela 3.13 podem ser apreciados os elementos e os parâmetros do modelo obtido com o sinal de excitação s_1 . Na Figura 3.16 é mostrado o espectro de potência depois do segurador de ordem zero do sinal MLPRS, com a região relevante para controle realçada.

Tab. 3.13: Parâmetros do modelo NARX não linear com o sinal s_1 .

elemento	termo	θ_i
1º	1	0,0059
2º	$u(t - 1)$	0,0840
3º	$y(t - 2)$	0,9388
4º	$u(t - 1)^2$	-0,039
5º	$y(t - 2)u(t - 1)$	-0,0682
6º	$y(t - 1)^2$	0,0193

Fig. 3.16: Espectro de potência do sinal com $q = 7$, $n_r = 2$.

Tab. 3.14: Índices de erro para grau de não linearidade $\ell = 1$.

sinal multi-nível	# termos	Erro RMS	Erro Máximo	AIC4	BIC
s_1	5	0,0227	$5,15 \times 10^{-4}$	$-3,52 \times 10^3$	$-3,51 \times 10^3$

Tab. 3.15: Parâmetros do modelo NARX linear com o sinal s_1 .

elemento	termo	θ_i
1º	1	0,0068
2º	$u(t-1)$	0,0849
3º	$y(t-1)$	1,8083
4º	$y(t-2)$	-0,8168
5º	$u(t-2)$	-0,0739

Segundo o relatório "Identification of the Dynamic Processes in WWTP" (Ayesa et al., 1994), um dos problemas da identificação de modelos em plantas de tratamento de esgotos é encontrar um compromisso entre a capacidade de rastreamento do modelo e sua complexidade, que deve ser a menor possível. Utilizando esta premissa é que pode ser escolhida uma representação linear que mostre a melhor capacidade de rastreamento para a concentração de oxigênio dissolvido. Este modelo identificado é mostrado na Tabela 3.15, obtido com o sinal de excitação s_1 . Note que este modelo apresenta índices de erro ligeiramente maiores, mas ele pode ser usado em aplicações de controle linear, onde seus parâmetros lineares são possíveis de atualizar mediante algoritmos de mínimos quadrados recursivos. Na Figura 3.17 pode-se apreciar o desempenho do modelo linear identificado, note a boa qualidade na aderência da concentração do OD, a pesar de este ser um modelo linear.

Como mencionado anteriormente, para efeitos de validação foram reservados dados de entrada e saída a partir das 4,8 h do tempo de simulação. Assume-se uma operação do tipo batelada no bioreator ($Q_{in} = 0$), onde este é arejado a partir das 5,6 até as 6,8 horas. Em seguida, a aeração é desligada, sendo observada a porção de consumo de oxigênio dos microrganismos até as 6,8 h do tempo de simulação. Na Figura 3.18 é incluída a validação para os modelos NARX linear e não linear. Note como o rastreamento do modelo linear perde um pouco na precisão das predições, mas consegue acompanhar a dinâmica da concentração do oxigênio dissolvido, mesmo sendo a sequência de excitação diferente da sequência utilizada para identificação.

Outra forma de validar o modelo é mediante a comparação com o modelo obtido por Jerônimo et al. (2000), na identificação da dinâmica da concentração de OD usando as mesmas condições operacionais do modelo *benchmark* ASWWTP-USP. O modelo NARX para o grau de não linearidade $\ell = 2$ obtido por eles é mostrado na Tabela 3.16.

A avaliação na qualidade das predições é comparado de forma quantitativa na Tabela 3.17. Os erros absoluto e máximo são mostrados para o modelo linear e não linear obtido com o sinal s_1 das

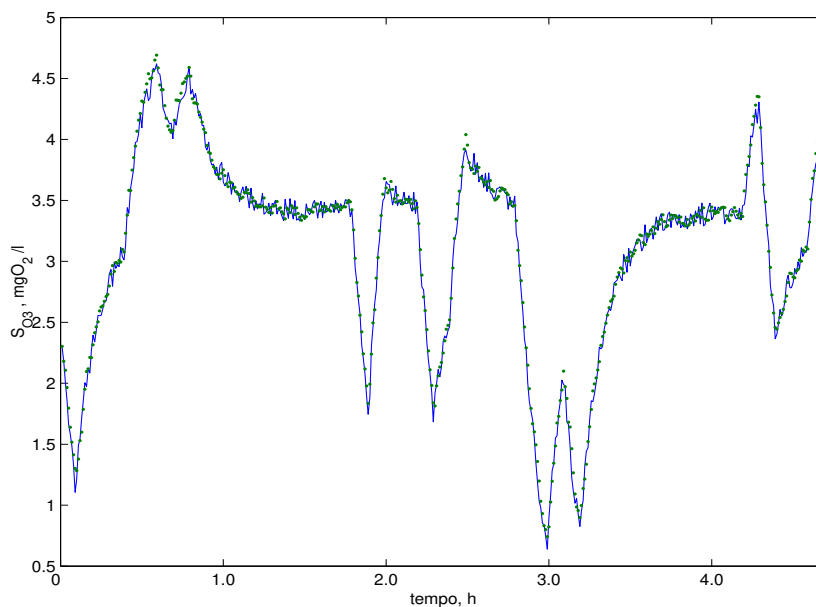


Fig. 3.17: Concentração do oxigênio dissolvido (traço contínuo) e saída do modelo identificado NARX linear (pontos) com o sinal s_1 .

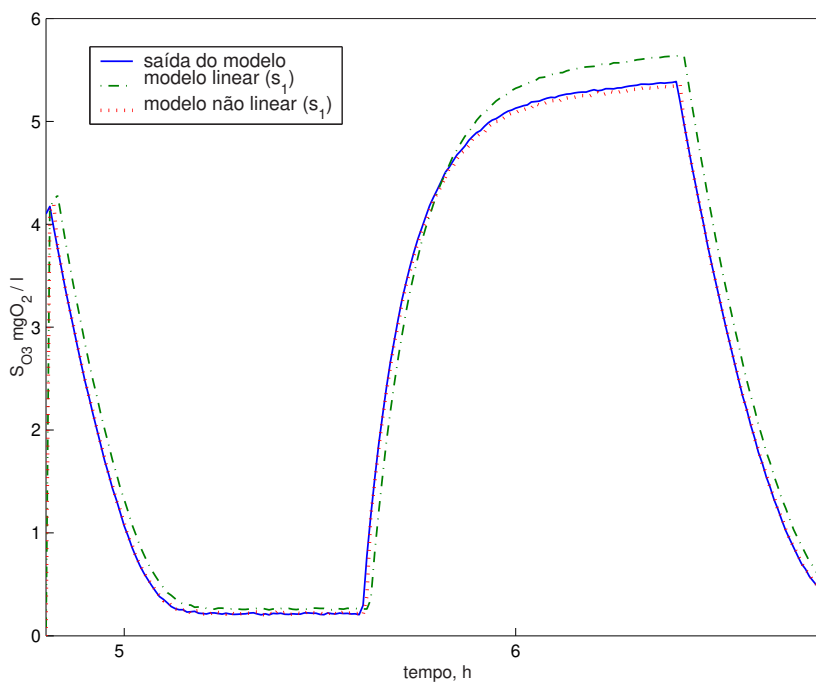


Fig. 3.18: Validação com os modelos identificados linear e não linear.

Tab. 3.16: Parâmetros do modelo NARX não linear (Jerônimo et al., 2000).

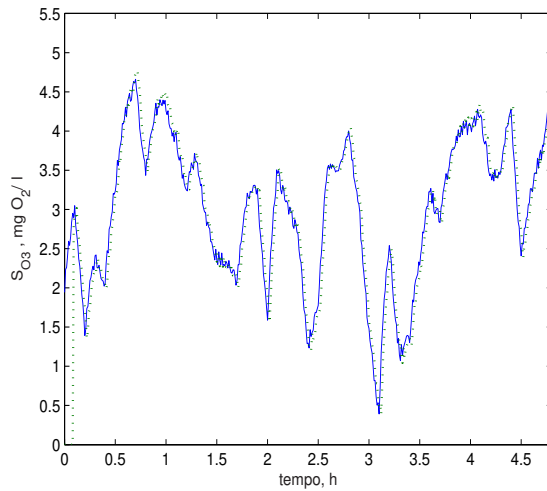
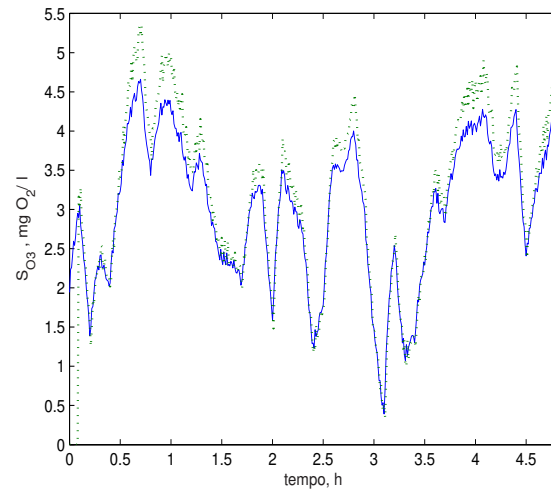
elemento	termo	θ_i	elemento	termo	θ_i
1º	1	0,022017	12º	$y(t-3)y(t-4)$	-0,002273
2º	$y(t-1)$	1,211622	13º	$u(t-1)$	0,113371
3º	$y(t-2)$	-0,301441	14º	$u(t-1)^2$	-0,079402
4º	$y(t-3)$	-0,01070	15º	$y(t-1)u(t-1)$	-0,431740
5º	$y(t-4)$	-0,011728	16º	$y(t-2)u(t-1)$	0,317758
6º	$y(t-5)$	0,014093	17º	$u(t-1)u(t-2)$	0,024843
7º	$y(t-6)$	-0,010819	18º	$y(t-1)y(t-4)$	0,153247
8º	$y(t-7)$	0,010092	19º	$u(t-2)$	-0,041326
9º	$y(t-8)$	-0,007407	20º	$u(t-2)^2$	0,044846
10º	$y(t-9)$	0,001671	21º	$y(t-4)u(t-3)$	-0,019077
11º	$y(t-3)y(t-4)$	-0,099591			

Tab. 3.17: Índices de erro para um conjunto independente de dados.

modelo	#	Erro	Erro
NARX	termos	RMS	Máximo
linear (s_1)	5	0.3991	0.1593
não linear (s_1)	6	0.3679	0.1354
não linear*	21	0.4587	0.2104

Tabelas 3.13 e 3.15, conjuntamente com o modelo da Tabela 3.16 (Jerônimo et al., 2000), indicado pelo asterisco. A geração de dados entrada-saída dos modelos foram realizada usando as mesmas condições experimentais, mas para um sinal MLPRS de 7 níveis, com o intuito de usar dados independentes dos usados na identificação. Cabe ressaltar que esse modelo possui um total de 21 termos para o grau de não linearidade 2, contra os 6 termos obtidos usando o sinal MLPRS. Note que o modelo linear identificado com o sinal s_1 apresenta menores índices de erro que aqueles do modelo de Jerônimo et al. (2000). Este resultado demonstra a eficácia na obtenção de modelos parcimoniosos quando são usadas as directrices de projeto de sinais MLPRS juntamente com o algoritmo ortogonal de Gram Schmidt Modificado para a estimação e seleção de estrutura mínima.

Uma comparação qualitativa da qualidade das predições dos modelos reduzidos é mostrada na Figura 3.19, onde é mostrada a predição da saída do modelo da concentração de OD quando usado o sinal s_1 e a predição do modelo não linear de Jerônimo et al. (2000). Pode-se verificar novamente uma melhor qualidade nas predições para o modelo obtido com o sinal s_1 , a pesar de este ser um

(a) Predição do modelo linear com s_1 (pontos).

(b) Predição do modelo não linear (pontos) (Jerônimo, 2000).

Fig. 3.19: Validação dos modelos da dinâmica de OD.

modelo linear.

3.9 Conclusão

Foi mostrado que sinais de excitação pseudo-aleatórios de seqüência m multi-nível podem ser projetados utilizando diretrizes que, considerando o conhecimento *a priori* do sistema físico, conseguem atender com eficácia requisitos de sinais de identificação relevantes para controle.

Características importantes como comprimento mínimo do sinal, persistência da excitação e potência espectral do sinal foram projetadas para dois processos com dinâmicas difíceis de modelar: o processo de neutralização de pH e a concentração de Oxigênio Dissolvido numa planta de lodos ativados. Demonstra-se então a capacidade dos sinais de comprimento máximo multi-nível, com característica de repetição inversa, de excitar dinâmicas não lineares de processos industriais, através da escolha adequada do número de níveis e mapeamento dos sinais, com vista a reduzir o efeito de distorções não lineares. Verifica-se também que o algoritmo ortogonal de identificação de modelos NARX (Chen et al., 1989) proporciona modelos de estrutura mínima.

No caso do processo de neutralização de pH, os melhores resultados foram obtidos com o sinal de excitação m multi-nível com 3 elementos no campo de Galois e com harmônicos múltiplos de 2 suprimidos, satisfazendo o requisito de mínimo número de níveis para o grau de não linearidade verificada. A comparação feita com a identificação mediante sinais aleatórios e considerando a dinâmica do atuador ressaltam a utilidade de sinais com características *plant-friendly*.

Para a identificação do processo de concentração de Oxigênio Dissolvido mostrou-se que o sinal de excitação m multi-nível com 7 elementos no campo de Galois e com harmônicos múltiplos de 2 suprimidos, excita adequadamente a largura de banda do processo para a identificação do modelo linear e não linear. Cabe ressaltar que modelos parcimoniosos foram obtidos para o sinal com

3 níveis, resultando em potenciais modelos para aplicações de controle baseado em modelo. Estes modelos mostraram serem eficazes na predição da dinâmica do OD na zona aeróbia quando foram comparados com o modelo obtido por Jerônimo et al. (2000), para as mesmas condições experimentais e com o mesmo modelo de simulação (o ASSWTP-USP). Portanto, o procedimento de geração de sinais MLPRS e o procedimento de identificação descritos serão adotados para a obtenção dos modelos reduzidos a serem usados no controle do processo de remoção de nitrogênio, como mostrado posteriormente no Capítulo 6.

Capítulo 4

Identificação de Modelos para Controle Preditivo

Uma característica importante num modelo de controle avançado é que este descreva com precisão as características dinâmicas que sejam essenciais no projeto do controlador. No contexto do controle preditivo baseado em modelo, um modelo explícito e medições obtidas do processo devem ser utilizadas para prever o comportamento futuro do processo, sendo portanto imprescindível um bom desempenho nas previsões do modelo. No entanto, a maioria dos modelos usados dentro do esquema de controle preditivo são identificados com algoritmos na forma de mínimos quadrados. Estes modelos são usados para obter a previsão k -passos-à frente da variável de saída, propagando repetidamente a estimação um-passo-à frente um número de k passos no futuro. Neste caso, visto que erros tendem a se acumular de uma iteração a outra, não são necessariamente obtidas boas previsões para vários passos à frente, caso existam dinâmicas não modeladas ou presença de ruídos de medição.

No presente capítulo é apresentado um estudo de modelagem para preditores orientados ao projeto de controladores preditivos baseados em modelo via identificação. É mostrado que o modelo preditor de múltiplos-passos-à frente, que identifica um modelo para cada passo de previsão, reduz a propagação de erros em preditores por iteração sucessiva padrão (um-passo-à frente). A vantagem desta técnica na melhoria do desempenho de modelos preditores é mostrada também comparando-a com outra técnica de modelagem, com o método de identificação preditiva de horizonte estendido. A análise realizada utilizando um modelo reduzido do processo, usado aqui para ilustrar incertezas não modeladas, mostra a eficácia do modelo preditor de múltiplos-passos-à frente para curtos horizontes de previsão. Será mostrado que para horizontes longos o erro de previsão do modelo preditor de identificação preditiva de horizonte estendido é tão pequeno quanto ao observado pelo modelo preditor de múltiplos-passos-à frente, os que são bem menores ao observado pelo modelo preditor por iteração sucessiva padrão.

4.1 Identificação e controle

Técnicas de identificação de sistemas têm sido usadas desde que o controle por realimentação começou a ser aplicado no controle de processos. Constituem algumas aplicações a análise de res-

posta transitória, no caso do ajuste para o controlador PID ¹, como também a análise de resposta em frequência e a síntese de Bode-Nichols clássica, no caso do projeto de controladores lineares.

O advento da teoria de controle moderna ao redor de 1960 e o uso explícito de modelos paramétricos possibilitou o desenvolvimento de uma grande variedade de técnicas de estimação de parâmetros em sistemas dinâmicos. Com isto, foi possível a introdução destes modelos na síntese de controladores baseados na estimação de sistemas lineares. Desta forma, a identificação de sistemas se converteu, em meados dos sessenta, em uma linha de pesquisa estabelecida dentro da área do controle automático (Ljung, 1996). Como consequência, na década de setenta houve um grande interesse na análise e desenvolvimento de novos métodos de identificação. A consolidação da identificação de sistemas como teoria se deu entre 1970 e 1985, onde se estabeleceu a importância do projeto do experimento de identificação e a influência das condições experimentais para a obtenção de "bons" modelos, como mostrado por exemplo em Goodwin e Payne (1977). Este desenvolvimento foi motivado principalmente para esses modelos serem usados em esquemas de controle baseado em modelo. Era de se esperar que as duas linhas de pesquisa, a identificação e o projeto de controle, explorassem as inter-relações existentes quando o modelo é orientado para controle.

O projeto de identificação e o projeto de controle tiveram desenvolvimentos em duas direções separadas e com escassas abordagens que explorem adequadamente estas relações. As contribuições para o projeto de identificação orientado ao controle foram poucas, devido ao fato de por um longo tempo a identificação ter sido vista como uma busca por um "sistema real", em lugar de uma metodologia para produzir uma "aproximação da realidade" quando a aplicação pretendida do modelo é o projeto de controladores. Com exceção da contribuição dada no texto de Ljung (1987), onde foram estabelecidos os fundamentos para um projeto formal quando o objetivo do modelo é orientado segundo a aplicação: simulação, predição ou controle.

Para o propósito de controle, modelos aproximados ou reduzidos de processos complexos (como são os processos biológicos de tratamento de esgotos), podem oferecer um bom desempenho em malha fechada se o modelo possui precisão em frequências relevantes para controle. Portanto, a identificação para controle (denominada assim na área de identificação de sistemas) ou o projeto de controle baseado no experimento (denominado assim desde o ponto de vista do projeto de controle), aborda a questão de como produzir diretrizes para o projeto de identificação no caso em que o objetivo de identificação seja o projeto de controladores baseado em modelos (Gevers, 1997).

4.2 Identificação de modelos para controle

Pode-se levantar agora a seguinte questão: Quando um modelo será "bom" para controle? Cabe ressaltar que um modelo para controle não necessariamente deve ser uma descrição precisa do sistema verdadeiro para se obter um alto desempenho no controlador, mas sim que ele descreva com grande precisão as características dinâmicas essenciais para o projeto do controlador. Um modelo será "bom" neste sentido, se (Blondel et al., 1997):

- O controlador derivado do modelo estabiliza a planta, e
- este controlador consegue um desempenho sobre a planta atual que seja próximo ao desempenho conseguido sobre o modelo.

¹Refere-se às siglas do controlador de ação proporcional, integral e derivativo.

Por exemplo, no caso de controladores PID, é usual aproximar modelos de grande ordem por modelos de primeira ordem do seguinte tipo:

$$G(s) = \frac{K}{\tau s + 1} e^{-\theta_d s} \quad (4.1)$$

onde os parâmetros a serem identificados a partir de medições feitas da curva de reação são: o ganho estático do processo, K , a constante de tempo, τ e o tempo morto ou atraso de transporte, θ_d . Este procedimento de modelagem baseado em análise gráfico é amplamente utilizado para sistemas com regimes de operação ao redor da frequência de corte do sistema em malha fechada. Segundo Seborg (1999), aproximadamente 90% de processos industriais ainda utilizam alguma forma do controle PID. No entanto, é importante ressaltar que este procedimento de identificação é efetivo quando a planta apresenta uma relação sinal-ruído muito pequena, ou em condições livres de perturbações (Rivera e Flores, 2000). Em presença significativa de ruído ou de distúrbios que levem ao sistema a um regime de operação diferente em que o controlador foi ajustado, o método pode derivar em resultados errôneos.

Uma outra forma de identificar modelos para controle, mas no domínio da frequência, é conseguida mediante a introdução de uma etapa de pré-filtragem no experimento de identificação do modelo, como explicado na seguinte seção.

4.2.1 Especificação de modelos para controle usando pré-filtragem

A pré-filtragem de dados dos erros de predição introduzem características interessantes no problema de identificação e que podem ser utilizadas no problema da obtenção de modelos orientados para controle. A seguir, é apresentada uma expressão para a função objetivo do problema de identificação que ilustra como o pré-filtro (entre outros elementos) afeta a estimativa dos parâmetros.

Considere a planta como sendo descrita por um sistema linear invariante no tempo (Ljung, 1999):

$$\mathcal{S} \quad : \quad y(t) = G(q^{-1})u(t) + H(q^{-1})e(t) \quad (4.2)$$

Aqui $u(t)$ é o sinal de entrada ou de controle, $y(t)$ é o sinal de saída observado, $e(t)$ é ruído branco com propriedades de segunda ordem de média nula e variância σ^2 e q^{-1} é o operador deslocamento em atraso. $G(q^{-1})$ é denominada função de transferência e é uma função racional em q^{-1} , significando simplesmente que os pólos serão os zeros do polinômio denominador. $G(q^{-1})$ é assumida estável. $H(q^{-1})$ é referido como o modelo de ruído assumido como um filtro mônico² e estável. O termo $H(q^{-1})e(t) = v(t)$ representa um distúrbio aditivo na planta.

Em lugar de se calcular (4.2), prefere-se usar uma descrição do modelo que possua um número finito de valores ou coeficientes denominados parâmetros, que podem ser determinados usando um conjunto finito N de medições das entradas e saídas $Z^N = \{u(1), y(1), \dots, u(N), y(N)\}$, sobre um intervalo de tempo $1 \leq t \leq N$:

$$\mathcal{M}(\theta) \quad : \quad \hat{y}(t/\theta) = G(q^{-1}, \theta)u(t) + H(q^{-1}, \theta)e(t) \quad (4.3)$$

² $H(q^{-1})$ é mônico se o coeficiente zero é 1: $H(q^{-1}) = \sum_{k=0}^{\infty} h(k)q^{-k}$, $h(0) = 1$.

onde os parâmetros a serem determinados são denotados pelo vetor θ , que dão origem ao conjunto de modelos parametrizados $\mathcal{M}(\theta) = \{\mathcal{M}(\theta) : \theta \in D_{\mathcal{M}} \subset \mathbb{R}^d\}$. Isto significa que o vetor θ pertence ao conjunto de valores admissíveis $D_{\mathcal{M}}$, sobre o conjunto \mathbb{R}^d , onde d é a dimensão de θ . Assume-se que o sistema \mathcal{S} está dentro do conjunto modelo \mathcal{M} e que a estrutura do conjunto modelo \mathcal{M} é uniformemente estável (Ljung, 1999).

Se o modelo (4.3) é utilizado e considerando o aspecto de desempenho da predição do modelo, pode-se enunciar uma medida deste desempenho, denominado erro de predição, definido por:

$$\varepsilon(t, \theta) = y(t) - \hat{y}(t/\theta) \quad (4.4)$$

Por outro lado, filtrar o erro de predição é uma importante opção e equivale a filtrar primeiro os dados entrada-saída (pré-filtragem) para depois aplicá-los nos preditores. Mediante um filtro L , efeitos de distúrbios de alta frequência não essenciais no problema de modelagem podem ser removidos, ou também ressaltar determinadas propriedades dos modelos para torná-los de interesse para controle.

Se o preditor é linear e invariante no tempo com y e u escalares, então a filtragem do erro de predição $\varepsilon_F(t, \theta)$ será dada por:

$$\begin{aligned} \varepsilon_F(t, \theta) &= L(q^{-1})\varepsilon(t, \theta) = \frac{L(q^{-1})}{H(q^{-1}, \theta)}(y(t) - G(q^{-1}, \theta)u(t)) \\ &= \frac{1}{H(q^{-1}, \theta)}(L(q^{-1})y(t) - G(q^{-1}, \theta)L(q^{-1})u(t)) \end{aligned} \quad (4.5)$$

o que equivale a pré-filtrar através de $L(q^{-1})$ os sinais observados da saída e da entrada:

$$\begin{aligned} y_F(t) &= L(q^{-1})y(t) \\ u_F(t) &= L(q^{-1})u(t) \end{aligned} \quad (4.6)$$

A aplicação do método do erro de predição por Mínimos Quadrados, nestes N dados de entrada-saída, fornece a estimativa $\hat{\theta}_N$ de θ :

$$\hat{\theta}_N = \hat{\theta}_N(Z^N) = \arg \min_{\theta \in D_{\mathcal{M}}} V_N(\theta) \quad (4.7)$$

$$V_N(\theta) = \frac{1}{2N} \sum_{t=1}^N \varepsilon_F^2(t, \theta) \quad (4.8)$$

Aqui o erro de predição um-passo-à frente para o modelo $\mathcal{M}(\theta)$ é $\varepsilon(t, \theta) = y(t) - \hat{y}(t/t-1)$, enquanto $L(q^{-1})$ é o filtro que atua como um ponderador de frequência. A estimativa $\hat{\theta}_N$ resulta:

$$\begin{aligned} \hat{\theta}_N &= \arg \min_{\theta \in D_{\mathcal{M}}} \frac{1}{2N} \sum_{t=1}^N \frac{L(q^{-1})}{H(q^{-1}, \theta)}(y(t) - G(q^{-1}, \theta)u(t)) \\ &= \arg \min_{\theta \in D_{\mathcal{M}}} \frac{1}{2N} \sum_{t=1}^N \frac{L(q^{-1})}{H(q^{-1}, \theta)}(G_0(q^{-1}) - G(q^{-1}, \theta)u(t)) + H_0(q^{-1})e_0(t) \end{aligned} \quad (4.9)$$

G_0 e H_0 definem a estrutura do "sistema verdadeiro". Como notado em Ljung (1999), isto pode ser escrito em forma equivalente no domínio da frequência, usando a relação de Parseval em $V_N(\theta)$ e analisando as propriedades de convergência de $\hat{\theta}_N$ quando N tende ao infinito ($N \rightarrow \infty$):

$$\lim_{N \rightarrow \infty} V_N(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} (|G_0(e^{jw}) - G(e^{jw}, \theta)|^2 \phi_u + \phi_v) \frac{|L(e^{jw})|^2}{|H(e^{jw}, \theta)|^2} dw \quad (4.10)$$

onde

$$\phi_v = |H_0(e^{jw})|^2 \phi_{e_0}$$

Na equação (4.10), ϕ_u e ϕ_v representam o espectro de potência para a entrada e o distúrbio, respectivamente. u e e_0 serão independentes se o sistema opera em malha aberta, assim como $H(q^{-1}, \theta)$ será um filtro fixo se o modelo de ruído é fixo e θ -independente. Portanto, pode-se obter o seguinte modelo limite:

$$\theta^* \approx \arg \min_{\theta \in D_{\mathcal{M}}} \int_{-\pi}^{\pi} |G_0(e^{jw}) - G(e^{jw}, \theta)|^2 \frac{|L(e^{jw})|^2}{|H(e^{jw})|^2} \phi_u dw \quad (4.11)$$

onde $|L(e^{jw})|^2 \phi_u / |H(e^{jw})|^2$ atua ponderando a distribuição de polarização de G . Usando este conhecimento é possível controlar em que faixas de frequência o modelo seria "melhor", mediante a seleção apropriada do espectro do sinal de entrada $\phi_u(w)$, da estrutura de $H(q^{-1}, \theta)$ e do pré-filtro $L(q^{-1})$. A relação (4.11) consolida a importância do papel que o projeto de sinais de excitação ocupa dentro da identificação de modelos relevantes para controle. Pode-se ver agora como as realizações dos sinais pseudo-aleatórios multi-nível (através de ϕ_u), consideradas no Capítulo 3, contribuem no ajuste destes modelos.

Um controlador pode ter um bom desempenho apesar de estar projetado sobre um modelo medíocre, sempre que este tenha o desempenho desejado em certas frequências, isto é, na largura de faixa do sistema em malha fechada. A dependência entre o modelo e a relação (4.11), tem originado pesquisas em identificação para controle, que envolve procedimento de projeto tanto em malha aberta como em malha fechada, como pode ser visto nos trabalhos de Rivera et al. (1992), Van den Hof e Schrama (1995), Hjalmarsson et al. (1996), Gevers (1997), Shook et al. (1991), Shook et al. (1992), Gopaluni et al. (2002), Fardin (2000), Schön (2001) e Gopaluni et al. (2003).

4.3 Identificação para Controle Preditivo

Não é difícil entender a tendência inicial na identificação de sistemas da busca de modelos que fossem interface entre os modelos matemáticos e o mundo real, modelos nos quais se procurava a aproximação com o mundo real (modelo "verdadeiro"). Sendo assim, aparenta ser contraditório que um modelo para controle não seja necessariamente uma descrição precisa do modelo verdadeiro para que este apresente um bom desempenho no projeto do controlador baseado em modelo. Como visto

anteriormente, a característica importante é que o modelo para controle descreva com precisão as características dinâmicas que sejam essenciais para o projeto do controlador.

No contexto do controle preditivo baseado em modelo, denominado pelas siglas MBPC (*Model Based Predictive Controller*) ou simplesmente por MPC, um modelo explícito e medições obtidas do processo são utilizadas para prever o comportamento futuro do processo. A ação de controle é calculada com o objetivo de minimizar a diferença entre a resposta predita do processo e alguma trajetória desejada. Note-se então uma característica que um modelo para controle preditivo deve possuir: precisão nas previsões. Isto é, a minimização acima usando previsões erradas pode levar a trajetória de controle ótima a uma solução sub-ótima ou provavelmente nem sequer a uma solução aceitável (Rossiter e Kouvaritakis, 2001).

Em muitas aplicações a informação necessária para o desenvolvimento de modelos derivados da aplicação de balanços de massa, de energia e momentos (modelos fenomenológicos, fundamentais ou de primeiros princípios) é insuficiente, ou o modelo é muito complexo para ser usado no projeto do controlador (como no caso do modelo IAWQ No. 1 da seção 1.3). Nestes casos, a modelagem empírica a partir de informação dinâmica da planta, obtida das variáveis de entrada-saída, constitui uma alternativa atrativa.

Geralmente controladores MPC são projetados baseados em modelos empíricos. Como consequência, a identificação de modelos é uma etapa chave na implementação de controladores preditivos (Seborg, 1999). Ao mesmo tempo, a identificação do modelo é a etapa mais crítica, assim como também consumidora de tempo e recursos no projeto de um controlador industrial (Rivera et al., 1992; Seborg, 1999; Morari e Lee, 1999; Zhu, 1998). Por exemplo, Zhu (1998) estima em duas ou três semanas a duração da etapa de testes de identificação em uma torre atmosférica de uma unidade de petróleo cru, com a supervisão constante do engenheiro de controle. Quin e Badgwell (2003) calcularam que a etapa do protocolo de testes pode demandar de 5 a 15 dias, enquanto que Morari e Lee (1999) estimam que os esforços de modelagem podem constituir até 90% do tempo e custos necessários no projeto do controlador MPC.

Existe a percepção, no campo de estratégias de controladores preditivos, que o desenvolvimento da identificação deva ser orientada para a obtenção de estimadores que atendam às especificações do algoritmo de controle, conceito também denominado Modelagem sobre Demanda -*Model-on-Demand* (Braun, 2001). Como consequência, o problema de identificação para controladores MPC tem começado a atrair a atenção de círculos acadêmicos e industriais (Cott, 1995; Zhu, 1998; Rossiter e Kouvaritakis, 2001).

A seguir, nesta seção, é feita uma revisão de conceitos fundamentais do controlador preditivo, incluindo uma breve revisão de textos e tutoriais que proporcionem ao leitor interessado um subsídio de pesquisa sobre a família de controladores preditivos. Estes conceitos são parte dos fundamentos teóricos para o desenvolvimento do problema de identificação para controle preditivo.

4.3.1 Conceitos Gerais e Formulação do MPC

Existe hoje um consenso geral tanto em círculos acadêmicos como em industriais de que o controle preditivo é a técnica de controle avançado com maior impacto na engenharia de controle industrial. O sucesso desta técnica deve-se principalmente a que é uma tecnologia de controle genérica que é capaz de incluir em sua formulação restrições impostas por equipamentos ou por limites de segurança (Maciejowsky, 2002). O controle preditivo, que na verdade não é uma estratégia de

controle específica, mas uma classe de algoritmos de controle desenvolvida em torno de certas idéias comuns; foi concebido originalmente para aplicações nas indústrias de refino e petroquímicas. Hoje em dia, o sucesso dos algoritmos MPC é comprovado pela grande variedade de aplicações em processos industriais, que incluem indústrias como a química, processamento de alimentos, automotora, aeroespacial, metalurgia, celulose e papel, entre outras. Quin e Badgwell (1996) contabilizaram aproximadamente 2200 aplicações em processos industriais no mercado americano, enquanto em outro *survey* encontraram mais de 4500 aplicações em diferentes áreas, considerando somente informação fornecida por fabricantes de tecnologias MPC até meados de 1999 (Quin e Badgwell, 2003).

Breve revisão da literatura MPC

A literatura relacionada à teoria e prática do MPC é vasta. Têm sido publicados diversos textos, entre eles o texto de Maciejowsky (2002), que apresenta um desenvolvimento do conceito do controle preditivo e de como este está relacionado com tópicos da teoria de controle (por exemplo estimação de estados, modelagem de distúrbios e análise de resposta em frequência), assim como detalhes sobre a solução do problema de otimização. Um outro livro onde são expostos conceitos fundamentais de controle preditivo generalizado, denominado GPC, é o texto de Camacho e Bordons (1995). Sánchez e Rodellar (1996) apresentam um bom ponto de partida para o estudo da técnica MPC, dentro do contexto do MPC adaptativo do qual o GPC faz parte. Por outro lado Mosca (1995) analisa o controle preditivo dentro do contexto da teoria de controle linear quadrático e do controle adaptativo, introduzindo o Controle Adaptativo Preditivo Multipasso, denominado MUSMAR (*Multistep Predictive Adaptive Control*).

O sucesso do controle preditivo linear, denominado LMPC, motivou a extensão desta metodologia a problemas de controle não linear, abordagem referida como controle preditivo não linear, NLMPC (*Nonlinear Predictive Control*). Meadows e Rawling (1997), fazem uma excelente revisão de tópicos importantes para a síntese do NLMPC, tais como modelagem, programação não linear, seleção dos parâmetros de ajuste, estabilidade e desempenho.

Um excelente tutorial sobre a teoria de controle preditivo linear e não linear é apresentado em Rawlings (1999). *Surveys* de aplicações industriais, como os de Quin e Badgwell (1996), Quin e Badgwell (2000) e Quin e Badgwell (2003), são especialmente recomendados por apresentarem uma introdução tanto da teoria como da prática da técnica de controle preditivo. Uma visão crítica do estado de arte, dos requisitos e necessidades do controle preditivo são encontrados em Henson (1998), Morari e Lee (1999), Lee (2000) e Nikolaou (2001). Mayne et al. (2000) fazem uma revisão do controle preditivo linear e não linear com restrições, enfatizando a análise de estabilidade e otimalidade, que de acordo com esses autores, já atingiu maturidade suficiente no LMPC e devem atrair agora o interesse de pesquisadores em controle não linear. Também é reconhecido que ainda existem muitas outras questões por resolver, como as relacionadas à robustez, adaptabilidade, tratamento de não linearidades, construção do modelo, monitoramento do desempenho, cômputo e implementação (Nikolaou, 2001).

Definição geral do controlador MPC

Para o propósito do presente trabalho, o controle preditivo é definido como o conjunto de esquemas de controle que utiliza um modelo interno do processo para prever a resposta futura da planta

sobre um intervalo de tempo finito, denominado *horizonte de predição*, levando em consideração informação passada (medições) disponível até um instante atual t . Um passo essencial no controle preditivo é a predição da saída do processo que é função do estado atual da planta e das variáveis de entrada futuras, variáveis que são calculadas como a solução de um problema de otimização *on-line*, (em tempo real) em malha aberta, para um instante de amostragem sobre um horizonte de tempo finito, denominado *horizonte de controle*. Finalmente, o primeiro elemento da seqüência de controle ótimo, $u(t/t)$, é injetado na planta e o problema de otimização é resolvido novamente para o próximo intervalo de tempo, deslocando a seqüência em uma unidade e atualizando as variáveis de entrada e saída da planta. Por esta razão esta etapa é referida como controle de horizonte retrocedente (*receding horizon control*, ou RHC), ou também como controle de horizonte móvel (*moving horizon control*).

Lei de controle

Os principais passos que descrevem a lei de controle preditiva são:

- Registro das entradas e saídas passadas do processo $u(t-1), u(t-2), \dots, u(t-j)$ e $y(t-1), y(t-2), \dots, y(t-j)$.
- Especificação da trajetória de referência, definindo a trajetória desejada para a saída do processo, $r(t+k)$, $k = N_1, \dots, N_2$. N_1 e N_2 são o horizonte de predição mínimo e máximo respectivamente.
- Predição da saída do processo, usando um modelo apropriado, calculado sobre o horizonte de predição, $N_p : [N_1, N_2]$. A saída predita deve estimar os valores de $y(t)$ nos instantes $t+k$, a partir das medições das entradas e saídas até o instante t e dos sinais de controle futuros.
- Cálculo da seqüência de controle, usando o modelo do processo, com a finalidade de satisfazer alguma função objetivo:
 - que minimize o desvio da saída predita do processo com relação à trajetória de referência (erro de predição), ao longo do horizonte de predição; e
 - que minimize o esforço de controle para que a saída predita do processo atinja a referência. A seqüência de controle está sujeita a restrições de operação pré-especificadas.

Pode ser encontrada uma solução analítica, se um critério quadrático é usado, para um modelo linear e sem restrições.

- O primeiro elemento da seqüência de controle calculada, $u(t/t)$, é injetado no processo, enquanto os elementos seguintes $u(t+1/t), u(t+2/t), \dots, u(t+i-1/t)$ são rejeitados. Note que $i = 1, \dots, N_c$, com N_c o horizonte de controle, assumindo que depois deste horizonte os esforços de controle são nulos: $u(t+i-1/t) = 0$ para $i > N_c$.

A seqüência de controle é novamente calculada no próximo intervalo de amostragem, sendo atualizada com as novas medições do processo, usando o princípio do horizonte retrocedente.

Índice de Desempenho Geral

Como visto anteriormente, o controlador preditivo calcula uma seqüência futura da variável de entrada, que conduz o processo a um ponto de operação desejado, sem violar restrições, processo que geralmente é descrito como em Rawlings (1999), usando a seguinte função objetivo:

$$J = \sum_{k=N_1}^{N_2} (\hat{y}(t+k/t) - r(t+k))^T Q(k) (\hat{y}(t+k/t) - r(t+k)) + \sum_{i=1}^{N_c} \Delta u(t+i-1/t)^T R(i) \Delta u(t+i-1/t) \quad (4.12)$$

- $\hat{y}(t+k/t)$: saída predita calculada no instante t .
- $r(t+k)$: trajetória de referência.
- $\Delta u(t+i-1/t)$: variação da variável de entrada.
- Q, R : penalização sobre o erro de rastreamento da predição e sobre o desvio da variável de entrada.
- N_1 : horizonte de predição mínimo.
- N_2 : horizonte de predição máximo.
- N_c : horizonte de controle.

A função objetivo (4.12) é calculada utilizando o princípio do horizonte retrocedente, onde somente a primeira ação de controle calculada sobre o horizonte N_c é utilizada. Note que operador $\Delta = 1 - q^{-1}$ é utilizado sobre $u(t)$ para obter a forma variacional da entrada:

$$\Delta u(t) = u(t) - u(t-1) \quad (4.13)$$

A razão de usar a forma variacional em (4.12), em lugar de $u(t)$, é permitir erros estáticos nulos entre a saída e o set-point em presença de distúrbios constantes na carga do tipo degrau. Esta é uma situação comum em processos industriais, em que $r(t)$ é constante durante longos períodos de tempo ou em que apresenta mudanças lentas. Outro benefício obtido mediante a introdução do integrador digital dentro da malha é conseguir transferências da ação de controle "suaves", desestimulando variações desnecessárias do sinal de entrada.

Um gráfico que ilustra a definição e a lei de controle preditivo é mostrado na Figura 4.1. Note que a predição da saída ao longo do horizonte de predição é função da variável de controle ótima, calculada sobre o horizonte de controle, depois do qual é mantida constante, isto é $\Delta u(t+k-1/t) = 0$ para $k > N_c$, sempre que a saída predita seja próxima à referência futura.

MPC com restrições

Como mencionado anteriormente, para o caso do MPC com um modelo linear invariante no tempo e em ausência de restrições, uma solução analítica pode ser encontrada. Na versão restrita, os cálculos de controle em tempo real são realizados resolvendo um problema de programação quadrática ou de programação linear a cada instante de amostragem, quando novas medições estejam disponíveis, antecipando futuras violações nas restrições. Estas restrições podem ser impostas nas

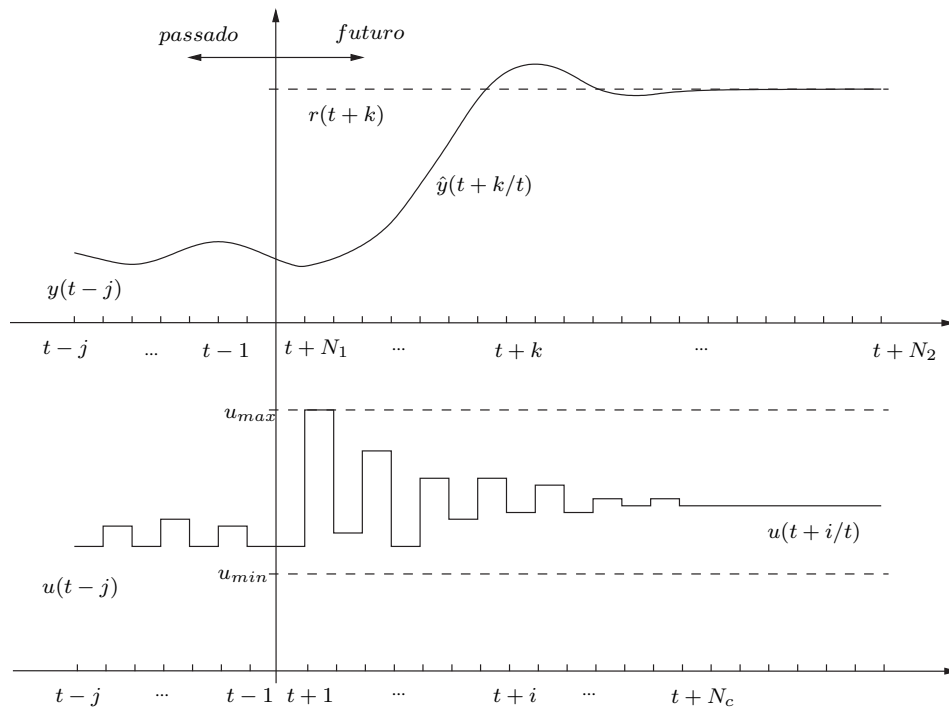


Fig. 4.1: Estratégia do controlador preditivo.

variáveis de entrada (devido a limitações no atuador, tais como saturação ou restrições na rapidez de cambio) e nas variáveis de saída (usualmente associadas a limitações operacionais, tais como considerações de segurança, qualidade final do produto ou especificações do equipamento), as que podem ser de três tipos:

- Restrição fortemente ativa (*hard constraint*): indica uma região que não pode ser violada nunca.
- Restrição fracamente ativa (*soft constraint*): permite a violação das mesmas em certos momentos, o que pode ser aceitável durante a operação da planta.
- Aproximação do *set-point*: é uma restrição fracamente ativa imposta pelo set-point, resultando em uma função objetivo penalizada em ambos os lados da restrição.

É importante notar que restrições nas entradas são geralmente restrições fortemente ativas, no sentido que devem ser sempre satisfeitas, enquanto restrições nas saídas são geralmente fracamente ativas, pois é permitida a violação para tornar o problema de otimização factível.

Predição da saída

A predição da saída, $\hat{y}(t+k/t)$, é obtida a partir de um modelo que pode adotar as duas formas seguintes (Haber et al., 2003)

- Transformação da equação do modelo do processo em uma equação de predição, resolvendo a equação de Diophantine, como usualmente abordado na lei do controlador preditivo generalizado;
- Sucessivas iterações vários passos-à frente, usando um modelo do processo.

Em ambas as formas, a saída predita é uma transformação do modelo do processo, baseada geralmente em modelos empíricos cujo mapeamento parametrizado dos dados passados, Z^{t-1} , no espaço de saída possui a seguinte estrutura geral (Ljung, 1999):

$$\hat{y}(t/\theta) = g(t, Z^{t-1}; \theta) \quad (4.14)$$

Introduz-se $Z^{t-1} = (y^{t-1}, u^{t-1}) = (y(1), u(1), \dots, y(t-1), u(t-1))$, para denotar as medições das entradas e saídas disponíveis até o instante $t - 1$. θ é um vetor finito unidimensional usado também para parametrizar o mapeamento. $\hat{y}(t/\theta)$ representará a saída $y(t)$ a partir dos dados passados, assim o modelo que define somente a função do preditor é denominado *modelo preditor* ou *equação do preditor*.

Como mencionado anteriormente na subseção 4.2.1, o critério usual que considera a capacidade preditiva dos diferentes modelos candidatos é o erro de predição, cujo objetivo é ajustar o modelo de predição, parametrizado pelo vetor θ , aos dados de entrada-saída registrados:

$$\begin{aligned} \varepsilon(t, \theta) &= y(t) - \hat{y}(t/\theta) \\ &= y(t) - g(t, Z^{t-1}; \theta) \end{aligned} \quad (4.15)$$

Métodos que utilizam o critério em (4.15) são denominados Métodos do Erro de Predição Mínimo, PEM (*Prediction Error Methods*), porque o parâmetro θ estimado pelo método será aquele que minimiza o erro de predição. Geralmente o problema PEM utiliza a norma quadrática para avaliar o erro:

$$\begin{aligned} V_N(\theta, Z^N) &= \frac{1}{2N} \sum_{t=1}^N \varepsilon^2(t, \theta) \\ &= \frac{1}{2N} \sum_{t=1}^N [y(t) - g(t, Z^{t-1}; \theta)]^T [y(t) - g(t, Z^{t-1}; \theta)] \end{aligned} \quad (4.16)$$

para um conjunto de N dados entrada e saída. O vetor de parâmetros estimados, $\hat{\theta}_N$, é calculado segundo:

$$\hat{\theta}_N = \hat{\theta}_N(Z^N) = \arg \min_{\theta \in D_{\mathcal{M}}} V_N(\theta, Z^N) \quad (4.17)$$

Portanto, a predição da saída para o modelo geral (4.14) é definido recursivamente como:

$$\hat{y}_1(t/\hat{\theta}_N) = \hat{y}_p(t/\hat{\theta}_N) = g(t, Z_p^{t-1}; \hat{\theta}_N) \quad (4.18)$$

sendo:

$$Z_p^{t-1} = \{\hat{y}_p(t-1/\hat{\theta}_N), u(t-1), y(t-2), u(t-2), \dots, y(1), u(1)\}$$

Note que a predição de saída do sistema depende dos dados de entrada-saída até o instante $t-1$ e do vetor de parâmetros $\hat{\theta}_N$, razão pela qual o preditor da equação (4.18) é denominado *preditor um-passo-à frente*, denotado por $\hat{y}_1(t/\hat{\theta}_N)$ ou equivalentemente por $\hat{y}_p(t/\hat{\theta}_N)$.

Da mesma forma, predições de vários passos-à frente são obtidas mediante sucessivas iterações até o instante $t-k$. Neste caso, as predições são função dos dados passados e das k predições, como mostrado em:

$$\hat{y}_k(t/\hat{\theta}_N) = g(t, Z_k^{t-1}; \hat{\theta}_N) \quad (4.19)$$

onde o vetor de entradas e saídas utilizado aqui resulta:

$$Z_k^{t-1} = \{\hat{y}_k(t-1/\hat{\theta}_N), u(t-1), \hat{y}_k(t-2), u(t-2), \dots, \hat{y}_k(t-k), u(t-k), \dots, y(1), u(1)\}$$

Portanto, o preditor da equação (4.19) é denominado *preditor k-passos-à frente*, quando é usado o modelo geral (4.14).

4.3.2 Limitações do preditor um-passo-à frente

Como visto, a maioria dos modelos usados no contexto do controle preditivo são identificados usando algoritmos de mínimos quadrados que minimizam o erro de predição um-passo-à frente. Devido a isto, o preditor associado também tem a forma um-passo-à frente. Estes modelos não necessariamente dão boas predições de horizonte estendido.

Por exemplo, se t denota o instante atual de amostragem, um preditor 10-passos-à frente realizará a predição do sistema até o instante $t+10$. Partindo do instante atual t , propaga-se repetidamente a predição um-passo-à frente padrão 10 vezes para obter a predição da saída no instante $t+10$, mas isto às custas de um erro acumulado, que será maior conforme o horizonte de predição aumente. Sabe-se que na presença de incertezas estruturais não modeladas, o modelo de mínimos quadrados não apresenta boas predições para um horizonte estendido longo (Shook et al., 1992). Este problema está presente não somente no preditor k -passos-à frente padrão, mas também em observadores k -passos-à frente (Lim e Phan, 1997).

Um exemplo que ilustra a limitação do preditor um-passo-à frente é mostrado a seguir.

Exemplo 4.3.1. Considere o processo auto-regressivo de primeira ordem: (Wahlberg e Ljung, 1986)

$$y(t) - a_0 y(t-1) = e(t) \quad (4.20)$$

A estimação do parâmetro $\theta = a_0$, usando o método do erro de predição um-passo-à frente, resulta na conhecida estimação de mínimos quadrados:

$$\hat{a}_N^{(1)} = \arg \min_a \sum_{t=1}^N (y(t) - a y(t-1))^2 \quad (4.21)$$

com variância assintótica calculada como em Ljung (1999):

$$\text{var } \hat{a}_N^{(1)} \approx \frac{1}{N} \frac{1}{\text{var } y} = \frac{1}{N} (1 - a_0^2) \quad (4.22)$$

a predição dois-passos-à frente para este processo resulta:

$$y(t + 2/a) = a^2 y(t) \quad (4.23)$$

da mesma forma, a estimação baseada na predição dois-passos-à frente é obtida usando a predição um-passo-à frente conhecida:

$$\hat{a}_N^{(2)} = \arg \min_a \sum_{t=2}^N (y(t) - a^2 y(t-2))^2 \quad (4.24)$$

e a variância assintótica para esta estimação resulta:

$$\text{var } \hat{a}_N^{(2)} \approx \frac{1}{N} \frac{1 + 3a_0^2}{4a_0^2} \frac{1}{\text{var } y} = \frac{1}{N} \frac{1 + 3a_0^2}{4a_0^2} (1 - a_0^2) \quad (4.25)$$

que resulta sendo consideravelmente maior que $\text{var } \hat{a}_N^{(1)}$. □

A primeira abordagem de solução deste problema encontra-se no algoritmo MUSMAR (Greco et al., 1984), onde adota-se um modelo de mínimos quadrados para cada horizonte de predição. Apesar da relevância, esta abordagem não foi percebida pela comunidade de pesquisadores em controle preditivo, porque foi desenvolvida para resolver problemas de convergência e de erros de modelagem no contexto do controle adaptativo.

Mais tarde o problema também foi abordado de forma mais direta em Shook et al. (1991) e Shook et al. (1992). Os autores mostraram que as predições de horizonte estendido do GPC, derivadas de um modelo identificado pelo algoritmo de mínimos quadrados recursivos, trazem inerentes os problemas de convergência e robustez desta classe de algoritmos perante inevitáveis incertezas estruturais entre planta e modelo. Em presença de incertezas o algoritmo de mínimos quadrados recursivos não fornecerá as melhores predições de horizonte estendido, reduzindo assim a efetividade do GPC adaptativo.

Uma outra tendência na identificação de modelos, que resulta em melhores predições de horizonte estendido, utiliza a mesma idéia do controlador MUSMAR, que emprega diferentes modelos para cada horizonte de predição. No contexto do controle preditivo esta abordagem é denominada identificação de preditores múltiplos-passos-à frente, MSPI (*Multi-Step-Ahead Prediction Identification*) Haber et al. (2003). Na verdade, a primeira vez que esta idéia foi utilizada dentro do contexto MPC foi em Liu et al. (1999). Neste trabalho, múltiplos modelos de predição foram construídos simultânea e paralelamente, um para cada passo de predição. Pouco depois Rossiter (2000), identifica modelos de predição para cada horizonte de predição, mostrando depois em Rossiter et al. (2000) a aplicação desta técnica no projeto de um controlador preditivo multimodelo numa unidade de geração de energia.

Na seguinte seção serão mostradas as estruturas do preditor por iteração sucessiva padrão, do preditor de identificação preditiva de horizonte estendido (LRPI) e do preditor de múltiplos-passos-à frente (MSPI). O objetivo é mostrar as vantagens e desvantagens de cada preditor na melhoria do

desempenho do controlador MPC, quando utilizado um modelo de ordem reduzida que traz consigo inevitáveis incertezas de modelagem.

4.4 Modelagem de Preditores k -passos-à frente

Para descrever os métodos de identificação de preditores considere novamente a descrição de sistema mostrada na equação (4.3):

$$\mathcal{M}(\theta) \quad : \quad \hat{y}(t/\theta) = G(q^{-1}, \theta)u(t) + H(q^{-1}, \theta)e(t)$$

onde o vetor θ é o vetor de parâmetros que pertence ao subconjunto \mathcal{M} de \mathbb{R}^d . Esta relação define um conjunto de funções racionais candidatas:

$$\begin{aligned} G &= \{G(q^{-1}, \theta) / \theta \in \mathcal{M}\} \\ H &= \{H(q^{-1}, \theta) / \theta \in \mathcal{M}\} \end{aligned} \quad (4.26)$$

A seqüência $e(t)$ é assumida como uma seqüência de variáveis aleatórias independentes, com média nula e variância λ . Com base no conjunto de modelos (4.3) é que são formulados os métodos de modelagem de preditores k -passos-à frente.

4.4.1 Modelo preditor por iteração sucessiva padrão

Este modelo preditor está baseado na descrição entrada-saída dada pela seguinte equação a diferenças (Wahlberg e Ljung, 1986):

$$y(t) + a_1y(t-1) + \dots + a_{n_a}y(t-n_a) = b_1u(t-1) + \dots + b_{n_b}u(t-n_b) + e(t) \quad (4.27)$$

Como o termo de ruído branco $e(t)$ é introduzido como um erro de modelagem, a equação (4.27) é com freqüência denominada modelo "equação de erro". Neste modelo é assumido um intervalo de amostragem igual a uma unidade de tempo. Para facilitar a notação ele pode ser re-escrito como:

$$A(q^{-1})y(t) = B(q^{-1})u(t) + e(t) \quad (4.28)$$

com os polinômios A e B descritos por:

$$\begin{aligned} A(q^{-1}) &= 1 + a_1q^{-1} + \dots + a_{n_a}q^{-n_a} \\ B(q^{-1}) &= b_1q^{-1} + \dots + b_{n_b}q^{-n_b} \end{aligned}$$

que definem ao sistema (4.3), quando:

$$G(q, \theta) = \frac{B(q^{-1})}{A(q^{-1})}, \quad H(q, \theta) = \frac{1}{A(q^{-1})} \quad (4.29)$$

A realização do preditor é obtida calculando o valor seguinte da saída dadas as observações pre-

vias:

$$y(t) = -a_1 y(t-1) - \dots - a_{n_a} y(t-n_a) + b_1 u(t-1) + \dots + b_{n_b} u(t-n_b) \quad (4.30)$$

Para uma notação mais compacta é definido o seguinte vetor:

$$\varphi(t) = [-y(t-1) \dots -y(t-n_a) \ u(t-1) \dots u(t-n_b)]^T \quad (4.31)$$

que denota ao vetor de regressores, em que a denominação "regressor" é usada para enfatizar que o cálculo de $y(t)$ é efetuado a partir de dados passados. Da mesma forma, o vetor de parâmetros é definido como:

$$\theta = [a_1 \ a_2 \ \dots \ a_{n_a} \ b_1 \ b_2 \ \dots \ b_{n_b}]^T \quad (4.32)$$

Como $y(t)$ depende dos parâmetros de θ e dos valores passados, pode-se denominar ao valor calculado $y(t)$ por $\hat{y}(t/\theta)$, de modo que a equação (4.30) possa ser escrita como:

$$\hat{y}(t/\theta) = \theta^T \varphi(t) + e(t) = \varphi(t)^T \theta + e(t) \quad (4.33)$$

que é conhecida como uma regressão linear para a estimação de θ , onde o método de mínimos quadrados (LS) é uma ferramenta padrão. Por esta razão, abusando da terminologia, o modelo (4.27) é muitas vezes referido como "modelo de mínimos quadrados" ou modelo LS.

O modelo preditor por iteração sucessiva é baseado no modelo (4.33), denominado também como modelo preditor LS. O preditor k -passos-à frente realizará a predição da saída do sistema no instante atual t , substituindo iterativamente o valor predito atual até o horizonte de predição k . Por exemplo, para o sistema de primeira ordem:

$$\hat{y}(t/\theta) = a_1 y(t-1) + b_1 u(t-1) \quad (4.34)$$

a equação de predição k -passos-à frente resulta:

$$\begin{aligned} \hat{y}(t+1/t) &= a_1 y(t) + b_1 u(t) \\ \hat{y}(t+2/t) &= a_1 \hat{y}(t+1) + b_1 u(t+1) \\ &= a_1^2 y(t) + a_1 b_1 u(t) + b_1 u(t+1) \\ &\vdots = \vdots \\ \hat{y}(t+k/t) &= a_1^k y(t) + b_1 \sum_{i=1}^k a_1^{k-i} u(t+i-1) \end{aligned} \quad (4.35)$$

Como somente os sinais antigos são armazenados ou medidos, o sinal de saída atual poderá ser calculado a partir dos valores medidos antigos:

$$\hat{y}(t/t-k) = a_1^k y(t-k) + b_1 \sum_{i=1}^k a_1^{k-i} u(t-k+i-1) \quad (4.36)$$

Modelos preditores de ordens maiores podem ser conseguidos também mediante a solução recursiva da seguinte equação de Diophantine:

$$1 = E_k(q^{-1})A(q^{-1}) + q^{-k}F_k(q^{-1}) \quad (4.37)$$

realizada para o horizonte de predição k , tal que os polinômios E_{k+1} e F_{k+1} sejam obtidos quando conhecidos os valores de E_k , F_k e os polinômios $A(q^{-1})$ e $B(q^{-1})$ da planta (4.28). Esta predição k -passos-à frente é obtida a partir de:

$$\hat{y}(t+k/t) = \frac{B(q^{-1})}{A(q^{-1})}u(t+k) + \frac{1}{A(q^{-1})}e(t+k) \quad (4.38)$$

substituindo (4.37) em (4.38) tem-se:

$$\hat{y}(t+k/t) = \frac{B(q^{-1})}{A(q^{-1})}u(t+k) + E_k(q^{-1})e(t+k) + \frac{F_k(q^{-1})}{A(q^{-1})}e(t) \quad (4.39)$$

colocando em evidência $e(t)$ de (4.28) e substituindo em (4.39), resultará a seguinte predição:

$$\hat{y}(t+k/t) = q^k \frac{B(q^{-1})}{A(q^{-1})}u(t) + E_k(q^{-1})e(t+k) + F_k(q^{-1})y(t) - \frac{F_k(q^{-1})B(q^{-1})}{A(q^{-1})}u(t) \quad (4.40)$$

fatorando $u(t)$:

$$\hat{y}(t+k/t) = F_k(q^{-1})y(t) + E_k(q^{-1})e(t+k) + (q^k - F_k(q^{-1}))\frac{B(q^{-1})}{A(q^{-1})}u(t) \quad (4.41)$$

e usando novamente a identidade de Diophantine (4.37), resulta:

$$\hat{y}(t+k/t) = F_k(q^{-1})y(t) + E_k(q^{-1})B(q^{-1})u(t+k) + E_k(q^{-1})e(t+k) \quad (4.42)$$

cujos últimos termos contêm informações independentes dos sinais medidos até o instante t , porém a predição de variância mínima de $y(t+k)$ é descrita mediante o preditor k -passos-à frente para qualquer ordem dos polinômios A e B :

$$\hat{y}(t+k/t) = F_k(q^{-1})y(t) + E_k(q^{-1})B(q^{-1})u(t+k) \quad (4.43)$$

4.4.2 Modelo preditor de identificação preditiva de horizonte estendido

O modelo preditor de identificação preditiva de horizonte estendido (LRPI) foi desenvolvido por Shook et al. (1991), dentro do contexto do controle preditivo generalizado. O preditor LRPI é baseado em uma extensão do modelo geral da seção 4.4, o denominado modelo CARIMA (*Controller Auto-Regressive Integrated Moving-Average*):

$$\hat{A}(q^{-1})y(t) = \hat{B}(q^{-1})u(t-1) + e(t) \quad (4.44)$$

com os polinômios \hat{A} e \hat{B} descritos por:

$$\begin{aligned}\hat{A}(q^{-1}) &= 1 + \hat{a}_1 q^{-1} + \dots + \hat{a}_{n_a} q^{-n_a} \\ \hat{B}(q^{-1}) &= \hat{b}_0 + \hat{b}_1 q^{-1} + \dots + \hat{b}_{n_b} q^{-n_b}\end{aligned}\quad (4.45)$$

e o termo $e(t)$ denotando o erro de predição, considerado como a seguinte realização estocástica:

$$e(t) = \frac{T(q^{-1})}{D(q^{-1})} \xi(t) \quad (4.46)$$

em que T e D são polinômios em q^{-1} . D é o operador $\Delta = 1 - q^{-1}$. $T(q^{-1})\xi(t)/\Delta$ descreve ao erro de predição como uma média móvel (MA) de uma seqüência aleatória decorrelacionada de média nula e variância σ^2 , $\xi(t)$. O papel do operador Δ é introduzir ação integral ao modelo tal que rejeite distúrbios de carga modelados como degraus aleatórios para tempos aleatórios, no caso determinístico, ou do tipo movimento Browniano, no caso estocástico.

A determinação do polinômio T é um processo complexo e resulta em um problema de estimação não linear. No lugar, pode-se considerar como parâmetro do projeto, assumido como sendo mônico e estável:

$$T(q^{-1}) = 1 + t_1 q^{-1} + \dots + t_{n_t} q^{-n_t} \quad (4.47)$$

Os sinais y e u são então filtrados mediante T/Δ antes do procedimento de estimação. Esta representação do modelo de ruído, como um sinal média-móvel-integrado, resulta mais apropriada para muitas aplicações de controle de processos químicos em que um dos objetivos de controle seja a atenuação de distúrbios (Clarke et al., 1987).

A formulação do princípio de identificação preditiva de horizonte estendido, para um número N de medições e para horizontes de predição mínimo, N_1 e máximo, N_2 , é expressada considerando a seguinte função objetivo em malha aberta:

$$J_{LRPI} = \frac{1}{N} \sum_{t=1}^N \frac{1}{N_2 - N_1 + 1} \sum_{j=N_1}^{N_2} [y(t) - \hat{y}(t + k/t)]^2 \quad (4.48)$$

que constitui na média da soma dos quadrados dos erros de predição k -passos-à frente. No entanto, a função objetivo (4.48) é não linear, dado que as predições de horizonte estendido da saída são não lineares nos parâmetros (Clarke et al., 1987):

$$\hat{y}(t + k/t) = \frac{F_k}{T} y(t) + \frac{E_k \hat{B}}{T} \Delta u(t + k - 1) \quad (4.49)$$

Estas predições são obtidas seguindo o procedimento da seção 4.4.1, obviando o argumento q^{-1} , com a seguinte identidade de Diophantine:

$$T = E_k \hat{A} \Delta + q^{-k} F_k \quad (4.50)$$

note que os polinômios E_k e F_k são definidos unicamente por \hat{A} no intervalo de predição k . Clara-

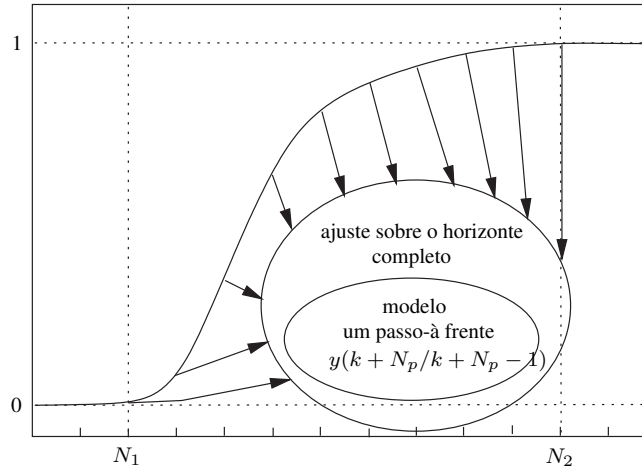


Fig. 4.2: Princípio do preditor de identificação preditiva de horizonte estendido (Haber et al., 2003).

mente, a solução da equação (4.48) resulta num problema de mínimos quadrado não linear.

Shook et al. (1991) mostraram, no domínio da frequência, que a função custo (4.48) pode ser substituída por outra na qual é introduzido um pré-filtro, L_{lrpi} , na etapa do estimador de mínimos quadrados, que está associado à predição um-passo-à frente. No caso em que a mesma entrada, mesmo modelo e o mesmo processo atuam nas funções custo, então a função custo de mínimos quadrados resulta equivalente à função custo LRPI, que está associada ao preditor k -passos-à frente, quando é satisfeita a seguinte condição:

$$|L_{lrpi}(e^{jw})|^2 = \frac{1}{N_p} \sum_{j=N_1}^{N_2} |E_k(e^{jw})|^2 \quad (4.51)$$

em que $N_p = N_2 - N_1 + 1$ corresponde ao horizonte de predição considerado. O filtro L_{lrpi} é fortemente influenciado pelo horizonte de predição e em menor medida pelo polinômio \hat{A} estimado. Na Figura 4.2 é ilustrado o princípio da identificação preditiva de horizonte estendido. Note que a estimação do modelo preditor LRPI é conseguida ao longo do horizonte de predição finito N_p . É importante ressaltar que os polinômios \hat{A} e \hat{B} estimados são resultados do procedimento de minimização LRPI:

$$\{\hat{A}(q^{-1}), \hat{B}(q^{-1})\} = \arg \min J_{LRPI} \quad (4.52)$$

os que descrevem o modelo preditor LRPI mantendo sua estrutura ao longo do horizonte de predição e que são diferentes aos polinômios da planta "real":

$$A(q^{-1})y(t) = B(q^{-1})u(t-1) \quad (4.53)$$

Desta forma obtém-se duas contribuições. A primeira é que a função custo (4.48) pode ser abordada como um problema linear e a segunda é que o filtro L_{lrpi} pode ser projetado para controle preditivo, dado que ele é fundamentalmente função dos horizontes de predição N_1 e N_2 .

Importância de N_2 na distribuição do erro de polarização

A função objetivo LRPI, no domínio da frequência, é aproximada assintoticamente ($N \rightarrow \infty$) usando a relação de Parseval para um modelo de ruído $H(q^{-1})$:

$$\lim_{N \rightarrow \infty} J_{LRPI} = \frac{1}{2\pi} \int_{-\pi}^{\pi} |G_0(e^{jw}) - G(e^{jw}, \theta)|^2 \frac{|L(e^{jw})|^2}{|H(e^{jw}, \theta)|^2} \phi_u dw \quad (4.54)$$

A distribuição do erro de polarização pode ser escrita em função do filtro (4.51):

$$\lim_{N \rightarrow \infty} J_{LRPI} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{N_p} \sum_{k=N_1}^{N_2} |E_k(e^{jw})|^2 \left| \frac{B(e^{jw})}{A(e^{jw})} - \frac{\hat{B}(e^{jw}, \theta)}{\hat{A}(e^{jw}, \theta)} \right|^2 \frac{1}{|H(e^{jw}, \theta)|^2} \phi_u(w) dw \quad (4.55)$$

Por simplicidade assume-se que $N_1 = 1$, que implica em $N_p = N_2$ e assume-se também que o espectro de entrada ϕ_u é uniforme para a largura de banda de análise. Se $N_2 = 1$, então $L_{lrpi,1} = 1$ o que implica que a estratégia LRPI será equivalente ao método de mínimos quadrados. Isto pode ser deduzido comparando (4.55) com a distribuição de polarização na equação (4.11).

A medida que N_2 aumenta a distribuição do erro de polarização se tornará mais uniforme na faixa de frequência analisada, pois quando $N_2 \rightarrow \infty$ tem-se:

$$\frac{1}{N_p} \sum_{j=N_1}^{N_2} |E_k(e^{jw})|^2 \rightarrow \frac{1}{|\hat{A}(e^{jw}, \theta)|^2} \quad \text{e} \quad |H(e^{jw})|^2 \rightarrow \frac{1}{|\hat{A}(e^{jw}, \theta)|^2} \quad (4.56)$$

Se o espectro de entrada ϕ_u é uniforme, então a distribuição do erro de polarização será também uniforme ao longo de toda a faixa de frequência, sem enfatizar as frequências mais altas. Esta ênfase ocorre em modelos de ordem reduzida, pois justamente nas frequências mais altas é que se evidenciam os efeitos de dinâmicas não modeladas. Uma discussão mais rigorosa pode ser estudada em Fardin (2000), que aborda o problema LRPI e analisa a influência do filtro na distribuição de polarização. A influência do espectro de potência da entrada não é considerada. No entanto, em Gopaluni et al. (2002) utiliza-se inicialmente a identificação LRPI para calcular espectros de potência de entrada ótimos para sistemas em malha aberta, influenciando desta maneira a distribuição da polarização do erro. Em ambos os casos os experimentos de identificação são realizados com modelos lineares de baixa ordem, utilizando sinais de excitação aleatórios.

Implementação do algoritmo

A implementação tipo *batch* do algoritmo de identificação do modelo preditor LRPI obedece os seguintes passos:

- i. Iniciar para $L_{lrpi,1} = 1$.
- ii. Filtrar os dados de entrada-saída com $L_{lrpi,1} \Delta/T$, onde Δ é escolhido para ser $(1 - q^{-1})$ e T pode ser especificado pelo usuário com a finalidade de conseguir regulação assintótica.
- iii. Usar os mínimos quadrados sobre os dados filtrados para obter um $\hat{\theta}[i]$.

iv Usar alguma rotina de fatoração espectral para resolver:

$$L_{lrpi,1}(q^{-1})L_{lrpi,1}(q) = \left(\sum_{j=N_1}^{N_2} E_k(q^{-1})E_k(q) \right) \quad (4.57)$$

v. Voltar ao passo ii até a convergência dos parâmetros.

A forma recursiva é quase idêntica ao esquema por batch, com a diferença de que a cada instante de amostragem, o filtro $L_{lrpi,1}$, baseado no polinômio \hat{A} estimado no instante anterior, é usado. Desta forma, os dados são filtrados antes que o procedimento de estimação recursiva seja efetuado.

4.4.3 Modelo preditor de múltiplos-passos-à frente

Uma alternativa à abordagem do problema de identificação preditiva de horizonte estendido, é a estimação dos modelos de predição de forma separada, um para cada horizonte de predição. Este procedimento corresponde ao modelo preditor de múltiplos-passos-à frente (MSPI). Na Figura (4.3) é ilustrado o princípio MSPI. Note a diferença com o princípio de modelagem do preditor LRPI da Figura (4.2).

A justificativa desta abordagem reside em que a função custo LRPI, por ser uma média dos erros de predição ao longo do horizonte de predição, pode não apresentar as mesmas magnitudes de erro para horizontes de predição curtos e estendidos (Schön, 2001). Portanto, o MSPI removeria o problema da média imposto nos modelos estimados que usam a função custo LRPI.

Para obter uma formulação geral deste modelo preditor, a estrutura geral (4.14) deve ser re-escrita da seguinte forma:

$$y(t) = f_1(y(t-1), y(t-2), \dots, y(t-n_a), u(t-1), u(t-2), \dots, u(t-n_b), e(t)) \quad (4.58)$$

onde interessa conhecer a estrutura do modelo de predição das saídas para cada k -ésimo passo de predição. Partindo do caso especial em que $n_a = n_b = 1$ na equação (4.27). Aqui a equação (4.58) é reduzida ao sistema de primeira ordem:

$$y(t) = f_1(y(t-1), u(t-1), e(t)) \quad (4.59)$$

resultando nas seguintes predições de $y(t)$, conhecidos os dados até o instante t :

$$\hat{y}(t+1/t) = f_1(y(t), u(t), e(t+1)) \quad (4.60)$$

$$\hat{y}(t+2/t) = f_1(y(t+1), u(t+1), e(t+2)) \quad (4.61)$$

Substituindo (4.60) em (4.61), temos:

$$\begin{aligned} \hat{y}(t+2/t) &= f_1(f_1(y(t), u(t), e(t+1)), u(t+1), e(t+2)) \\ &= f_2(y(t), u(t+1), u(t), e(t+2), e(t+1)) \end{aligned} \quad (4.62)$$

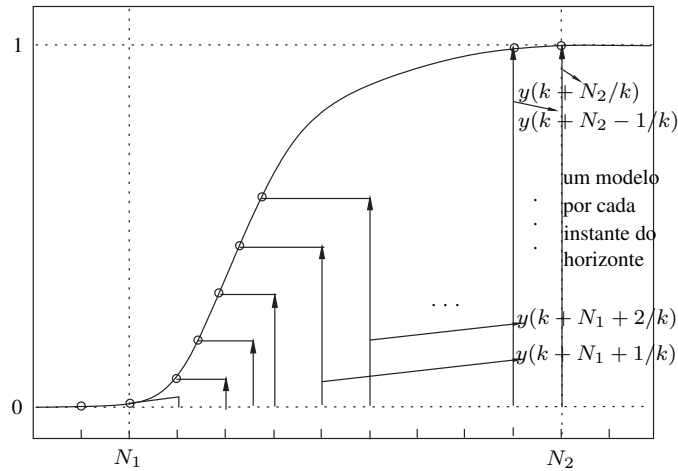


Fig. 4.3: Princípio do preditor de múltiplos-passos-à frente (Haber et al., 2003).

da mesma forma, de (4.62) tem-se o seguinte passo de predição:

$$\hat{y}(t + 3/t) = f_2(y(t + 1), u(t + 2), u(t + 1), e(t + 3), e(t + 2)) \quad (4.63)$$

Substituindo (4.60) em (4.63), resulta:

$$\begin{aligned} \hat{y}(t + 3/t) &= f_2(f_1(y(t), u(t), e(t + 1)), u(t + 2), u(t + 1), e(t + 3), e(t + 2)) \\ &= f_3(y(t), u(t), u(t + 1), u(t + 2), e(t + 1), e(t + 2), e(t + 3)) \end{aligned} \quad (4.64)$$

Desta forma podem ser obtidos k modelos de predição gerais. Cada uma destas equações de predição podem ser expandidas em modelos linearizados localmente através do Teorema de Taylor:

$$\hat{y}(t + 1/t) = \alpha_0^{(1)} y(t) + \beta_0^{(1)} u(t) + \omega(t + 1) \quad (4.65)$$

onde o superíndice (i) denota o i -ésimo passo de predição e ω denota um pequeno resíduo. Os coeficientes α e β são tais que (4.65) seja o modelo linearizado da planta não linear (4.59). O modelo (4.65) resulta no seguinte modelo preditor um-passo-à frente (assumindo que o valor esperado de $\omega(t + 1)$ é zero):

$$\hat{y}(t + 1/t) = \alpha_0^{(1)} y(t) + \beta_0^{(1)} u(t) \quad (4.66)$$

Obviamente que para os seguintes passos de predição não poderia ser usada a equação (4.65) repetidamente, dado que a planta é geralmente não linear. Uma forma alternativa é obter modelos linearizados localmente aplicando o Teorema de Taylor para cada equação de predição:

$$\hat{y}(t + 2/t) = \alpha_0^{(2)} y(t) + \beta_0^{(2)} u(t) + \beta_1^{(2)} u(t + 1) \quad (4.67)$$

$$\hat{y}(t + 3/t) = \alpha_0^{(3)} y(t) + \beta_0^{(3)} u(t) + \beta_1^{(3)} u(t + 1) + \beta_2^{(3)} u(t + 2) \quad (4.68)$$

onde assume-se também que o valor esperado de $\omega(t+1)$ e $\omega(t+2)$ é zero.

É importante notar que os coeficientes do preditor de múltiplos-passos-à frente são distintos aos coeficientes do preditor por iteração sucessiva padrão. Sabe-se que a equação de predição para um sistema de primeira ordem por iteração sucessiva resulta ser:

$$\hat{y}(t+k/t) = a_1^k y(t) + b_1 \sum_{i=1}^k a_1^{k-i} u(t+i-1)$$

Por exemplo, se for considerado um sistema de primeiro ordem ($n_a = 1$ e $n_b = 1$), as seguintes relações serão atendidas:

$$\alpha_0^{(k)} \neq a_1^k; \quad \beta_{i-1}^{(k)} \neq b_1 a_1^{k-i}; \quad \forall i = 1, 2, \dots, k$$

pois o método separa cada k equação de predição e estima os coeficientes $\alpha^{(k)}$ e $\beta^{(k)}$ diretamente com um algoritmo de estimação, por exemplo, o algoritmo de mínimos quadrados.

Modelo preditor múltiplos-passos-à frente: formulação geral

Para obter a formulação geral, recorre-se novamente à estrutura geral da equação (4.58):

$$y(t) = f_1(y(t-1), y(t-2), \dots, y(t-n_a), u(t-1), u(t-2), \dots, u(t-n_b), e(t))$$

De forma similar ao sistema de primeira ordem, as saídas previstas para cada horizonte de predição, conhecidos os dados até o instante t , podem ser obtidas como segue:

$$\begin{aligned} \hat{y}(t+1/t) &= f_1(y(t), y(t-1), \dots, y(t-n_a+1), u(t), u(t-1), \dots, u(t-n_b+1), e(t+1)) \\ \hat{y}(t+2/t) &= f_1(\hat{y}(t+1), y(t), \dots, y(t-n_a+2), u(t+1), u(t), \dots, u(t-n_b+2), e(t+2)) \\ &= f_2(y(t), y(t-1), \dots, y(t-n_a+2), y(t-n_a+1), u(t+1), u(t), u(t-1) \dots, \\ &\quad u(t-n_b+2), e(t+1), e(t+2)) \\ &\vdots \\ &\vdots \end{aligned} \tag{4.69}$$

Repetindo o procedimento até o horizonte $k \geq 0$, pode-se obter a estrutura geral para o preditor múltiplos-passos-à frente:

$$\hat{y}(t+k/t) = f_k(y(t), y(t-1), \dots, y(t-n_a+2), y(t-n_a+1), u(t+k-1), \dots, u(t-n_b+2), u(t-n_b+1), e(t+k), \dots, e(t+1)) \tag{4.70}$$

Da mesma forma, $\hat{y}(t+k/t)$ pode ser expandido usando o Teorema de Taylor para obter um sistema linearizado localmente que corresponderá ao *modelo preditor múltiplos-passos-à frente*, dados os

sinais medidos até o instante t :

$$\begin{aligned} \hat{y}(t+k/t) = & \beta_{k-1}^{(k)}u(t+k-1) + \beta_{k-2}^{(k)}u(t+k-2) + \cdots + \beta_0^{(k)}u(t) + \\ & \alpha_0^{(k)}y(t) + \alpha_1^{(k)}y(t-1) + \cdots + \alpha_{n_a-1}^{(k)}y(t-n_a+1) + \\ & \gamma_1^{(k)}u(t-1) + \gamma_2^{(k)}u(t-2) + \cdots + \gamma_{n_b-1}^{(k)}u(t-n_b+1) \end{aligned} \quad (4.71)$$

Estimação do modelo preditor múltiplos-passos-à frente

Para obter a estimação do modelo preditor múltiplos-passos-à frente, podem ser usadas rotinas de estimação de mínimos quadrados. Obviamente, qualquer outro método de estimação pode ser usado. Utilizam-se rotinas do *System Identification Toolbox* (Mathworks, Inc.) para implementar o algoritmo de estimação aproximada (Schön, 2001). A aproximação é feita fixando a estrutura do modelo de predição, antes que a identificação seja realizada.

Para fins de estimação, o sinal de saída atual deve ser predito a partir de valores medidos antigos, que foram previamente armazenados. Assim sendo, o modelo preditor (4.71) é reescrito como segue:

$$\begin{aligned} \hat{y}(t/t-k) = & \beta_{k-1}^{(k)}u(t-1) + \beta_{k-2}^{(k)}u(t-2) + \cdots + \beta_0^{(k)}u(t-k) + \\ & \alpha_0^{(k)}y(t-k) + \alpha_1^{(k)}y(t-k-1) + \cdots + \alpha_{n_a-1}^{(k)}y(t-k-n_a+1) + \\ & \gamma_1^{(k)}u(t-k-1) + \gamma_2^{(k)}u(t-k-2) + \cdots + \gamma_{n_b-1}^{(k)}u(t-k-n_b+1) \end{aligned} \quad (4.72)$$

ou, ainda, no seguinte modelo preditor:

$$\hat{y}(t/t-k, \theta) = \tilde{A}(q^{-1}, \theta)y(t) + \tilde{B}(q^{-1}, \theta)u(t) \quad (4.73)$$

com os polinômios \tilde{A} e \tilde{B} correspondentes a:

$$\begin{aligned} \tilde{A}(q^{-1}, \theta) = & -\alpha_0^{(k)}q^{-k} - \cdots - \alpha_{n_a-1}^{(k)}q^{-(n_a+k-1)} \\ \tilde{B}(q^{-1}, \theta) = & \beta_{k-1}^{(k)}q^{-1} + \cdots + \beta_0^{(k)}q^{-k} + \gamma_1^{(k)}q^{-(k+1)} + \cdots + \gamma_{n_b-1}^{(k)}q^{-(n_b+k-1)} \end{aligned} \quad (4.74)$$

Como pode ser visto em (4.73), o preditor múltiplos-passos-à frente pode ser considerado como um sistema MISO³ para fins de estimação, com duas entradas, $u(t)$ e $y(t)$ e uma saída $y(t/t-k)$ de ordem (n_a+k-1) no polinômio \tilde{A} e (n_b+k-1) no polinômio \tilde{B} .

Se comparada com a estrutura ARX:

$$\begin{aligned} y(t) = & -a_1y(t-1) - a_2y(t-2) \dots - a_{n_a}y(t-n_a) + \\ & b_1u(t-d) + b_2u(t-d-1) + \dots + b_{n_b}u(t-d-n_b+1) \end{aligned} \quad (4.75)$$

que é utilizada pela função `arx` do *System Identification Toolbox*, pode-se notar da relação (4.72) que a diferença reside nos primeiros termos da saída y , onde não aparecem os primeiros $(k-1)$ termos. Portanto, pode-se utilizar a função `arx` do *Toolbox* para realizar a aproximação, de forma que os $(k-1)$ parâmetros líderes sejam considerados parâmetros nulos e não sejam estimados.

³Síglas do inglês *Multiple Input Single Output*, que denotam a um sistema de múltiplas entradas e uma saída

Na literatura de identificação de sistemas, estes modelos são freqüentemente denominados modelos *caixa-cinza*, dado que alguns dos parâmetros são predefinidos. Assim, é possível identificar cada equação de predição independentemente, de modo que seja minimizado um erro de predição, para cada passo de predição, usando a seguinte função custo:

$$V_N(\theta, k) = \frac{1}{N} \sum_{t=1}^N (y(t) - \hat{y}(t/t - k, \theta))^2 \quad (4.76)$$

Note que é minimizada a diferença entre o valor atual do sinal de saída $y(t)$ e o valor predito do mesmo sinal $\hat{y}(t/t - k, \theta)$, com a informação disponível até o instante $(t - k)$. É esta diferença no tempo que é introduzida para *aproximar* o modelo de predição.

4.4.4 Erros de Polarização

Sistemas reais geralmente resultam sendo mais complexos que as estruturas admissíveis dos modelos estimados e portanto nunca poderão ser descritos completamente pelas estruturas dos modelos estimados. Como consequência, existirão erros de polarização nos modelos estimados. Nesta seção se discute o erro de polarização e como este é influenciado pelo passo de predição N_p .

Para analisar os erros de polarização pode-se assumir que o número de medições N tende ao infinito, para que seja possível separá-los dos erros de variância (Ljung e Glad, 1994). Se o erro de predição, $\varepsilon(t, \theta)$, é uma seqüência independente de variáveis estocásticas (ruído branco), então a lei dos grandes números na forma mais simples implicará:

$$\frac{1}{N} \sum_{t=1}^N \varepsilon^2(t, \theta) \rightarrow E[\varepsilon^2(t, \theta)] = \bar{V}(\theta), \quad \text{quando } N \rightarrow \infty \quad (4.77)$$

onde $\bar{V}(\theta)$ denota a variância de ε . A convergência será também uniforme no parâmetros θ . Isto implica que:

$$\hat{\theta}_N \rightarrow \theta^* = \arg \min_{\theta} \bar{V}(\theta), \quad \text{quando } N \rightarrow \infty$$

Em outras palavras, a estimativa $\hat{\theta}_N$ converge ao valor que minimiza $E[\varepsilon^2(t, \theta)]$. Este resultado contém a informação do erro de polarização. Dado que não se pode obter um modelo exato, pelo menos pode-se obter a melhor aproximação disponível dentro do conjunto de modelos tal que seja minimizada a variância do erro de predição.

Este resultado também pode ser interpretado no domínio da freqüência, tal como realizado na seção (4.2.1). Pode-se demonstrar que o espectro de $\varepsilon(t, \theta)$ para $L \equiv 1$ corresponde a (se u e v são independentes):

$$\phi_{\varepsilon}(w, \theta) = |G_0(e^{jw}) - G(e^{jw}, \theta)|^2 \frac{\phi_u(w)}{|H(e^{iw}, \theta)|^2} + \frac{\phi_v(w)}{|H(e^{iw}, \theta)|^2} \quad (4.78)$$

onde $H(q, \theta)$ é o modelo do termo de ruído individualmente parametrizado de $G(q, \theta)$. Mediante a

fórmula de Parseval em (4.77):

$$E [\varepsilon^2(t, \theta)] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi_{\varepsilon}(w, \theta) dw$$

é obtida a distribuição de polarização de um-passo-à frente:

$$\theta^* = \lim_{N \rightarrow \infty} \hat{\theta}_N = \arg \min_{\theta} \int_{-\pi}^{\pi} |G_0(e^{jw}) - G(e^{jw}, \theta)|^2 \frac{\phi_u(w)}{|H(e^{jw}, \theta)|^2} dw \quad (4.79)$$

pois a contribuição do último termo de (4.78) pode ser considerada nula dado que u e v foram assumidas independentes. O estimador converge para θ^* , fazendo com que a função frequência do modelo $G(e^{jw}, \theta^*)$ seja tão próxima quanto possível do "verdadeiro" $G_0(e^{jw})$ medida na norma quadrática da frequência com função peso:

$$\frac{\phi_u(w)}{|H(e^{jw}, \theta)|^2} \quad (4.80)$$

que também pode ser aplicável para dados a partir de qualquer outro sistemas arbitrário, não linear ou variante no tempo (Ljung, 1999).

Este resultado pode ser generalizado para qualquer k arbitrário, mediante a distribuição da polarização k -passos-à frente (Wahlberg e Ljung, 1986):

$$\theta^* = \lim_{N \rightarrow \infty} \hat{\theta}_N = \arg \min_{\theta} \int_{-\pi}^{\pi} |G_0(e^{jw}) - G(e^{jw}, \theta)|^2 \frac{|\bar{H}_k(e^{jw}, \theta)|^2}{|H(e^{jw}, \theta)|^2} \phi_u(w) dw \quad (4.81)$$

cuja função peso associada corresponde a:

$$\frac{|\bar{H}_k(e^{jw}, \theta)|^2}{|H(e^{jw}, \theta)|^2} \phi_u(w) \quad (4.82)$$

onde $\bar{H}_k(q^{-1}, \theta)$ é definido por:

$$\bar{H}_k(q^{-1}, \theta) = 1 + \sum_{r=1}^{k-1} h(r, \theta) q^{-r} \quad (4.83)$$

em que $h(r, \theta)$ representa os coeficientes do filtro \bar{H}_k .

Wahlberg e Ljung (1986) ilustram como o incremento de k consegue uma penalização uniforme para todas as frequências, quando $1 < k < \infty$. Analisa-se inicialmente o comportamento em frequên-

cia quando $k = 1$ em (4.82):

$$\frac{\bar{H}_k(q^{-1}, \theta)}{H(q^{-1}, \theta)} = \frac{1}{H(q^{-1}, \theta)} \quad (4.84)$$

Isto corresponde ao modelo de ruído do preditor um-passo-à frente, $H(q^{-1}, \theta) = 1/A(q^{-1}, \theta)$. A função peso (4.84) é tipicamente um filtro passa-alta e se $u(t)$ é banda larga (por exemplo, ruído branco), então esta função irá enfatizar as altas frequências. É por isto que se o preditor um-passo-à frente (modelo LS) é usado associado a um modelo de ordem reduzida, que apresente dinâmicas não modeladas nas altas frequências, então haverá o ajuste entre o modelo e o sistema “real” precisamente na região de alta frequência com ausência de informação sobre as dinâmicas do sistema real, originando estimações polarizadas (Wahlberg e Ljung, 1986).

Quando $k = \infty$ a função peso apresentará a seguinte equivalência:

$$\frac{\bar{H}_k(q^{-1}, \theta)}{H(q^{-1}, \theta)} \equiv 1 \quad (4.85)$$

e se o sinal de excitação for também do tipo ruído branco ($\phi_u(w) = 1$), então a função de distribuição de polarização em (4.81) tenderá a ser uniforme em toda a faixa de frequência, dado que a penalização tende a 1. Portanto, o preditor k -passos-à frente tenderá a amenizar a ênfase nas altas frequências conforme k aumente. Esta é uma propriedade interessante para modelos a serem usados em controladores preditivos, implicando que previsões de horizontes longos seriam menos polarizadas se modelos de múltiplos-passos-à frente fossem usados.

4.4.5 Relação erro de predição

Uma forma de validar a capacidade preditiva dos modelos preditores é utilizar a relação que compara o desempenho do preditor k -passos-à frente com o desempenho do preditor um-passo-à frente, isto é, calculando os erros de predição dos diferentes modelos com os dados medidos até o instante t :

$$R_k = \frac{\sum_{t=1}^N (\hat{y}_k(t/t - k) - y(t))^2}{\sum_{t=1}^N (\hat{y}_1(t/t - 1) - y(t))^2} \quad (4.86)$$

No caso de k modelos de predição de k -passos-à frente, tal como na abordagem de Rossiter (2000), o índice k nas previsões $\hat{y}(t/t - k)$ do numerador indica qual preditor esta sendo usado para estimar o modelo, para um horizonte de predição dado. Note que quando $R_k < 1$ o erro de predição k -passos-à frente será menor que o erro de predição um-passo-à frente, implicando que o modelo preditor possui melhor precisão e que este pode ser considerado como um preditor candidato.

Este índice foi utilizado por Rossiter e Kouvaritakis (2001) para validação de multi-modelos identificados em malha aberta (um modelo para cada horizonte de predição), para sistemas lineares e não lineares. Com o mesmo intuito, este índice é usado no presente trabalho para validar os diferentes modelos preditores dando uma medida do potencial do modelo preditor para aplicações em controle

preditivo.

4.4.6 Comparação dos diferentes tipos de modelos preditores: estudo de caso

Para ilustrar o desempenho dos modelos preditores apresentados, é utilizado nesta seção o seguinte modelo, considerado como planta ou modelo real:

$$G(s) = \frac{2 \times 229}{(s + 1)(s^2 + 30s + 229)} \quad (4.87)$$

usado inicialmente por Rohrs et al. (1984) na síntese de controladores adaptativos. Posteriormente este modelo foi usado por Shook et al. (1991) e mais tarde por Fardin (2000), para analisar os problemas que dinâmicas não modeladas, de modelos de baixa ordem, causam em preditores relevantes para controle preditivo.

O modelo (4.87) em forma discreta, para intervalos de amostragem de 0,05s, resulta:

$$G(q^{-1}) = \frac{0,006559q^{-1} + 0,01803q^{-2} + 0,003021q^{-3}}{1 - 1,891q^{-1} + 1,117q^{-2} - 0,2122q^{-3}} \quad (4.88)$$

Para este estudo identifica-se um modelo de ordem reduzida. O sinal de excitação $u(t)$ no experimento de identificação é um sinal pseudo-aleatório binário (PRBS) de amplitude unitária e de 6 registros de deslocamento, projetado de acordo com o procedimento descrito na seção (3.6.3) para um número de níveis $m = 2$. O tamanho do passo de integração numérica é de 0,1s. O procedimento adotado é o seguinte. Simula-se o modelo em malha aberta para o sinal de excitação projetado adicionando à saída um distúrbio aleatório de distribuição normal com variância 0,01. Os dados entrada-saída obtidos são usados para identificar um modelo de primeiro ordem. Baseado neste modelo é determinado o modelo preditor por iteração sucessiva padrão baseado no modelo LS (preditor LS), enquanto o modelo preditor de identificação preditiva de horizonte estendido (preditor LRPI) e o modelo preditor múltiplos-passos-à frente (preditor MSPI) são identificados para o passo de predição k , segundo os procedimentos explicados nas sub-seções 4.4.2 e 4.4.3, respectivamente.

A seguir são mostrados, para fins de ilustração, os procedimentos de determinação dos modelos preditores para 2-passos-à frente. Em seguida é realizada a análise do desempenho no domínio do tempo e da frequência para passos de predição maiores.

Predição por iteração sucessiva padrão

O modelo reduzido de primeira ordem no qual é baseado o preditor por iteração sucessiva padrão, de uma entrada e uma saída, corresponde a:

$$y(t) = \frac{\hat{b}_1 q^{-1}}{1 + \hat{a}_1 q^{-1}} u(t) \quad (4.89)$$

Depois de realizar o experimento de identificação, o modelo de mínimos quadrados (modelo LS)

em função dos parâmetros θ e dos valores passados é:

$$\hat{y}(t/\theta) = 0,9862y(t-1) + 0,002296u(t-1) \quad (4.90)$$

As previsões por iteração sucessiva padrão k -passos-à frente são determinadas usando a relação (4.35) ou (4.36). Portanto, a previsão para $k = 1$, dado que são conhecidas as medições prévias, resulta:

$$\hat{y}(t/t-1) = 0,9862y(t-1) + 0,002296u(t-1) \quad (4.91)$$

Da forma similar é determinada a previsão por iteração sucessiva padrão 2-passos-à frente:

$$\hat{y}(t/t-2) = 0,97259y(t-2) + 0,00223u(t-2) + 0,00229u(t-1) \quad (4.92)$$

Predição por identificação preditiva de horizonte estendido

O modelo CARMA de primeira ordem, em que o preditor LRPI é baseado, é obtido obviando o operador Δ da equação (4.44). O preditor para k -passos de previsão é estimado usando o filtro $T = 1 - 0.8q^{-1}$. Quando $k = 1$ a previsão LRPI é equivalente à previsão por iteração sucessiva padrão, que resulta sendo óbvio, pois quando $N_2 = N_1 = 1$ o filtro L_{lrpi} é 1 e a estratégia LRPI resulta ser equivalente ao preditor um-passo-à frente, como visto na sub-seção 4.4.2. Usando a mesma estratégia, é estimada a previsão LRPI 2-passos-à frente, que resulta:

$$\hat{y}(t/t-2) = 0,9865y(t-1) + 0,0032u(t-1) \quad (4.93)$$

Note que a estrutura da equação de previsão (4.93) continua sendo a de um sistema de primeira ordem.

Predição múltiplos-passos-à frente

As equações de previsão múltiplos-passos-à frente são determinadas segundo o procedimento de estimação visto na sub-seção 4.4.3. Nota-se também que a equação de previsão MSPI com horizonte de previsão $k = 1$ é equivalente à previsão por iteração sucessiva padrão, pois no primeiro passo é resolvido o mesmo problema de Mínimos Quadrados. A previsão MSPI 2-passos-à frente é estimada usando a função `arx` de estrutura fixa como mostrado a seguir:

$$M2 = \text{arx}(\text{data}, [2 \ 2 \ 1], 'fixedparameter', [1])$$

onde o primeiro termo do polinômio \tilde{A} é fixado como termo nulo. Note que se não existe tempo morto, então tem-se $d = 1$ porque as medições são realizadas a intervalos discretos de amostragem. A equação de previsão correspondente é:

$$\hat{y}(t/t-2) = 0,9516y(t-2) + 0,00013u(t-2) + 0,0045u(t-1) \quad (4.94)$$

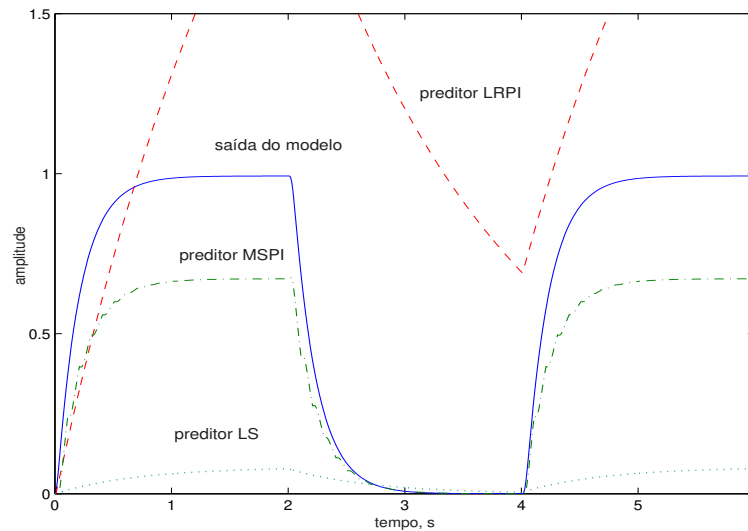


Fig. 4.4: Resposta do modelo real (—) e dos preditores LS (...), LRPI (- -) e MSPI (-.-) 10-passos-à frente.

Análise comparativo

Como pode ser visto, para $k > 1$ existe diferenças entre a equação de predição por iteração sucessiva padrão (4.92), a equação de predição de identificação preditiva de horizonte estendido (4.93) e a equação de predição múltiplos-passos-à frente (4.94). O que é difícil de se perceber é qual modelo exibirá as melhores predições de horizonte estendido. Portanto, para mostrar o desempenho dos modelos no domínio do tempo e da frequência, são calculados os modelos preditores para um passo de predição arbitrário $k = 10$. A análise é complementada com os resultados fornecidos pela relação de erro de predição (4.86).

O resultado gráfico da análise no domínio do tempo é mostrado na Figura 4.4, onde pode ser verificado que as predições do preditor LS são as de menor qualidade. Como explicado na seção 4.3.2, a variância do erro aumenta com o incremento do horizonte de predição. Se o modelo reduzido é estimado com métodos PEM um-passo-à frente, como é o método de mínimos quadrados, então dinâmicas não modeladas nas frequências altas provocaram estimações polarizadas. Como consequência, o preditor 10-passos-à frente associado traz inerente o efeito destas estimações polarizadas.

Note também que o preditor LRPI apresenta maior erro de predição se comparado com as predições MSPI. O efeito do aumento do horizonte de predição no preditor LRPI pode ser visto na Figura 4.5. Nota-se um melhor desempenho nas predições para $k = 40$, quando comparadas com as predições LRPI para $k = 10$. Isto pode ser entendido no domínio da frequência. Conforme o horizonte de predição aumenta, mais uniformemente distribuída em frequência se torna a distribuição do erro de polarização da relação (4.55), resultando em modelos melhores, com parâmetros estimados menos polarizados. Este resultado concorda com os obtidos por Fardin (2000) ao analisar o efeito do aumento do horizonte de predição na distribuição do erro de polarização da estratégia LRPI.

Como os modelos preditores estimados serão utilizados para predições de horizonte estendido no controle preditivo, é desejável que estas predições apresentem um comportamento de baixa frequência

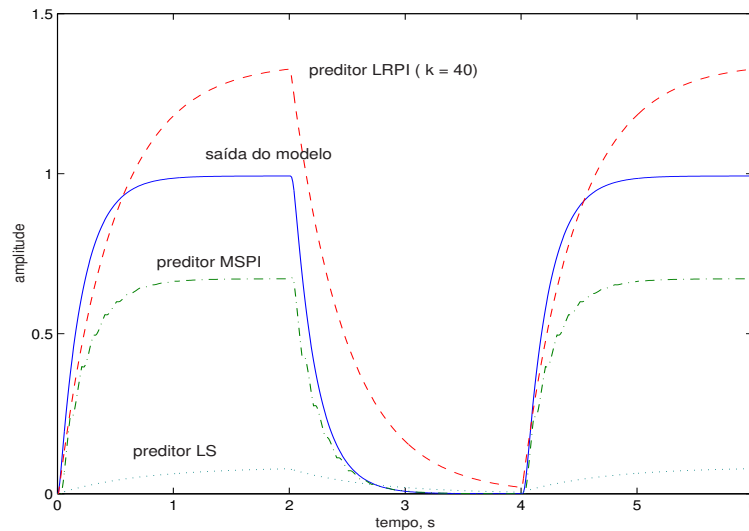


Fig. 4.5: Resposta do modelo real (—) e dos preditores LS (...) e MSPI (- -) 10-passos-à frente e do preditor LRPI (-.-) 40-passos-à frente.

preciso. Portanto, na Figura 4.6 são mostrados os resultados da análise no domínio da frequência. Os diagramas de Bode para os três modelos preditores 10-passos-à frente são comparados com a resposta em frequência do modelo real. Se conclui a partir da Figura 4.6, que a predição MSPI 10-passos-à frente descreve melhor o comportamento em baixa frequência do sistema real, quando comparado com os apresentados pelos outros preditores.

Uma forma de avaliar o potencial dos modelos preditores é usar a relação erro de predição, R_k , descrita na sub-seção 4.4.5, que compara a evolução do erro de predição de um modelo preditor em função do horizonte de predição. O objetivo é complementar a análise no domínio do tempo e da frequência mediante a inspeção do comportamento das predições para horizontes mais longos.

É comparado o desempenho do erro de predição LRPI e MSPI para 20-passos-à frente, resultado mostrado na Figura 4.7. Note que a queda do erro relativo do preditor MSPI é mais rápida que a do preditor LRPI, mas verifica-se que o valor final para o preditor LRPI resulta sendo próximo do valor final para o preditor MSPI. Para verificar esta tendência é realizada a estimação dos modelos preditores LRPI e MSPI para um horizonte de predição maior. Sabe-se que em controle preditivo são típicos horizontes de predição próximos do tempo de acomodação (Clarke et al., 1987; Seborg et al., 1989). No caso em questão, o tempo de acomodação é de 3,5 s, que corresponde a um horizonte de predição de 35 intervalos de amostragem. Portanto, na Figura 4.8 é mostrado o desempenho do erro de predição dos preditores LRPI e MSPI para 40-passos-à frente. Verifica-se então que as predições dos modelos preditores LRPI e MSPI para horizontes longos de predição (acima de 10), apresentam valores do erro de predição menores e próximos, à medida que aumenta o horizonte de predição. A diferença reside no grau de complexidade do preditor resultante, pois predições MSPI de longos horizontes demandarão ordens maiores dos polinômios estimados \tilde{A} e \tilde{B} , como pode ser verificado a partir da relação (4.74).

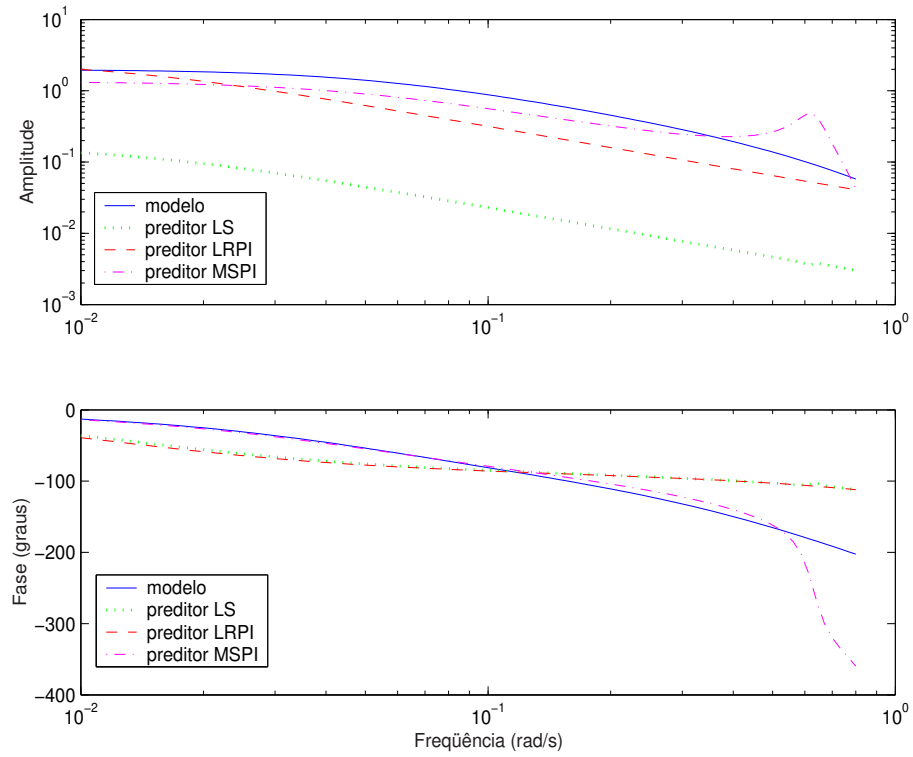


Fig. 4.6: Diagramas de Bode dos preditores LS, LRPI e MSPI 10-passos-à frente.

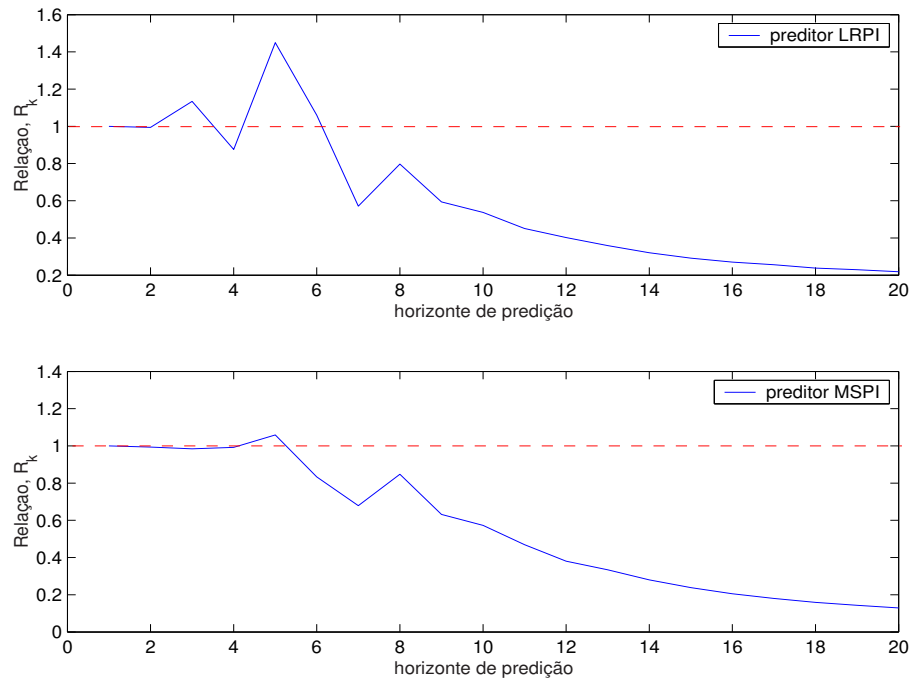


Fig. 4.7: Relação de erro de predição entre os preditores LRPI e MSPI 20-passos-à frente.

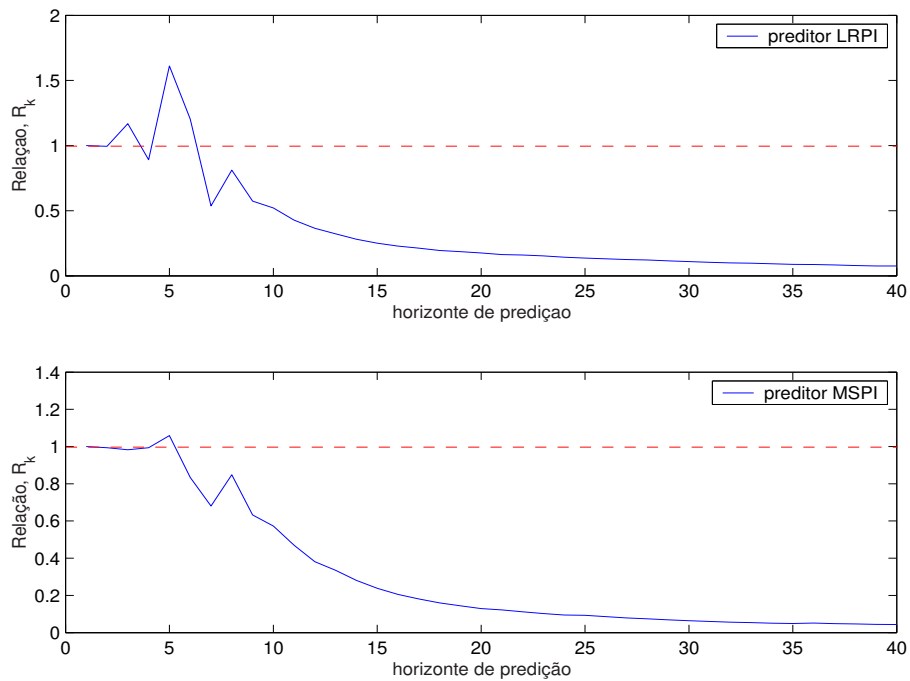


Fig. 4.8: Relação de erro de previsão entre os preditores LRPI e MSPI 40-passos-à frente.

4.5 Conclusão

Neste capítulo os antecedentes e a relação entre as linhas de pesquisa da identificação de sistemas e o controle automático foram expostos como motivação para o estudo da identificação orientada ao controle. Como exemplo do potencial que traz consigo a identificação orientada ao controle se mostra a influência que pré-filtros e o projeto do sinal de excitação exercem na especificação de modelos para controle. É com esta finalidade que no capítulo anterior se estudaram guias de projeto para sinais de excitação relevantes para controle.

Conceitos gerais para a classe de algoritmos de controle preditivo foram expostos como marco teórico para o estudo da identificação para controle preditivo. A limitação do modelo preditor por iteração sucessiva padrão baseado em modelos de ordem reduzida, usada geralmente na literatura do controle preditivo, é consequência destes modelos serem identificados com algoritmos que minimizam o erro de previsão um-passo-à frente. Se os modelos reduzidos apresentam dinâmicas não modeladas nas altas frequências, o preditor associado irá enfatizar justamente esta região com ausência de informação de dinâmicas em altas frequências do sistema "verdadeiro".

Para abordar o problema da modelagem de preditores de horizonte estendido foram apresentadas as estruturas do modelo preditor de horizonte estendido (LRPI), introduzido por Shook et al. (1991) e do modelo preditor de múltiplos-passos-à frente (MSPI), ambos implementados no presente estudo. Na validação do desempenho dos preditores por simulação verificou-se a limitação do preditor por iteração sucessiva padrão, comparando-o com o desempenho dos preditores LRPI e MSPI para um conjunto de sinais de entrada diferentes dos da identificação. A análise no domínio do tempo e da frequência mostram um melhor rastreamento do preditor MSPI em relação ao sistema "verdadeiro",

especialmente nas baixas frequências onde é interessante para o projeto do controle de processos. A validação por meio do índice Relação Erro de Predição (R_k) mostra que para horizontes de predição mais longos o erro de predição final do MSPI é ligeiramente menor ao do preditor LRPI, não obstante a diminuição de erro do LRPI resulta próximo ao observado pelo preditor MSPI quando o horizonte de predição aumenta.

Finalmente, é importante ressaltar que para a escolha do modelo preditor tem-se que validar seu desempenho em malha fechada tendo presente o custo computacional extra do algoritmo de identificação de múltiplos-passos-à frente na aplicação pretendida. Por esta razão, que no seguinte capítulo são avaliados os modelos preditores dentro do contexto do controle preditivo, tal que seja possível realizar a escolha da estratégia de identificação do preditor relevante para o controle preditivo, tendo presente as características e dificuldades no controle do processo de todos ativados.

Capítulo 5

Controle Preditivo Baseado em Modelos de Predição Múltiplos

Como visto anteriormente, o controle preditivo baseado em modelo não é uma estratégia de controle específica e sim uma classe de métodos de controle desenvolvida em torno de certas idéias comuns. O método de controle preditivo generalizado, proposto por Clarke et al. (1987), é um dos representantes mais populares desta família tanto na indústria como também em círculos acadêmicos. Isto é demonstrado pelo grande número de aplicações (Clarke, 1988; Camacho e Bordons, 1995; Bordons e Camacho, 1998). Será adotado no presente trabalho o método de Clarke *et al.* por duas razões. Primeiro, por abordar processos com não linearidades moderadas (Camacho e Bordons, 1995), unificando em sua formulação as vantagens do controle adaptativo e do controle de horizonte estendido com relação à falta de robustez do algoritmo recursivo de mínimos quadrados frente a dinâmicas não modeladas. Vale lembrar que os processos de remoção por lodos ativados são processos variantes no tempo com características não lineares. Segundo, por ter sido a referência no desenvolvimento dos modelos preditores analisados no capítulo anterior, podendo assim servir de comparação no desempenho dos controladores preditivos baseados nestes preditores.

Neste capítulo, a formulação padrão do controlador preditivo generalizado é revista com o intuito de esclarecer a modificação proposta. Em seguida, é mostrada a formulação modificada do controlador preditivo generalizado baseado no modelo preditor múltiplos-passos-à frente. É introduzida nesta seção a versão modificada do algoritmo recursivo de mínimos quadrados usando a atualização de Bierman. Na última seção é feita uma análise comparativa do desempenho dos controladores baseados nos modelos preditores por iteração sucessiva, por identificação de horizonte estendido e por múltiplos-passos-à frente, utilizando uma vez mais o modelo de Rohrs como planta.

5.1 Controle Preditivo Generalizado (GPC)

O princípio básico do GPC é calcular uma seqüência de sinais futuros de controle de tal forma que seja minimizada uma função custo definida sobre um horizonte de predição. O índice a ser otimizado é a esperança matemática de uma função quadrática que mede a distância entre a saída predita do sistema com alguma seqüência de referência sobre este horizonte de predição adicionando a essa função quadrática uma medida do esforço de controle. A solução do problema de otimização a

cada instante é implementada usando também o princípio do horizonte retrocedente. O GPC dá uma solução explícita na ausência de restrições, podendo abordar plantas instáveis e de fase não mínima (Camacho e Bordons, 1995).

A seguir apresenta-se a formulação original do GPC (Clarke et al., 1987), que utiliza como modelo preditor o preditor por iteração sucessiva padrão estudado na seção 4.4.1. A formulação é descrita inicialmente como um método de controle não adaptativo de uma variável de entrada e uma variável de saída (processo SISO), com o objetivo de facilitar a apresentação e também para simplificar a notação.

5.1.1 Formulação padrão do controlador GPC

O controlador preditivo generalizado é baseado no modelo linear (ou linearizado) de uma planta na forma CARIMA (*Controlled Auto-Regressive Integrated Moving-Average*) (Clarke et al., 1987):

$$A(q^{-1})y(t) = B(q^{-1})u(t-1) + \frac{C(q^{-1})}{\Delta}\xi(t) \quad (5.1)$$

No modelo CARIMA $u(t)$, $y(t)$ e $\xi(t)$ representam o sinal de entrada, de saída e o distúrbio no instante t . A , B e C são polinômios no operador atraso unitário q^{-1} , com A e C mônicos:

$$\begin{aligned} A(q^{-1}) &= 1 + a_1q^{-1} + \dots + a_mq^{-1} \\ B(q^{-1}) &= b_0 + b_1q^{-1} + \dots + b_mq^{-1} \\ C(q^{-1}) &= 1 + c_1q^{-1} + \dots + c_mq^{-1} \end{aligned} \quad (5.2)$$

Se a planta tiver tempo morto não nulo, então o elemento líder do polinômio B é considerado zero. A denominação media móvel integrada:

$$x(t) = \frac{C(q^{-1})}{\Delta}\xi(t) \quad (5.3)$$

considera $\xi(t)$ como uma seqüência aleatória descorrelacionada. Esta é a representação mais adequada para muitas aplicações industriais, pois permite modelar distúrbios do tipo degrau ou do movimento browniano que estão presentes em plantas industriais (Clarke et al., 1987).

A formulação do controlador GPC considera o modelo de distúrbio C/Δ como parâmetro de projeto do controlador, assumido como T/Δ . Isto é, em lugar de considerar o polinômio C como resultado de algum procedimento de identificação. Sendo assim, o polinômio T pode ser utilizado para melhorar a rejeição ao distúrbio, adicionando robustez ao controlador frente a dinâmicas não modeladas. As propriedades do polinômio T e seu papel na melhoria da robustez do GPC são discutidas com mais detalhe em Clarke e Mohtadi (1989). Neste sentido, Oliveira et al. (1996) dão diretrizes para a seleção da ordem do polinômio T considerando a ordem do polinômio estimado \hat{A} :

$$\begin{aligned} \text{grau}(T) &\geq \text{grau}(\hat{A}) - 1 && \text{se } \lambda_u \neq 0 \\ \text{grau}(T) &\geq \text{grau}(\hat{A}) - 1 + N_c && \text{se } \lambda_u = 0 \end{aligned} \quad (5.4)$$

sendo usada a notação $\text{grau}(X)$ para denotar o grau do polinômio X , N_c o horizonte de controle,

λ_u a ponderação sobre o controle e assumindo que $\text{grau}(\hat{A}) \leq \text{grau}(\hat{B}) + 1$, com \hat{B} a estimação do polinômio B . Em relação as raízes do polinômio T , recomenda-se $T = (1 - \beta q^{-1})^n$, sendo β uma constante próxima do valor da constante tempo dominante e n igual a ordem de $\text{grau}(\hat{A}) + 1$ (Mohtadi, 1988).

O modelo descrito em (5.1) pode ser equivalentemente representado pela seguinte equação:

$$A(q^{-1})\Delta y(t) = B(q^{-1})\Delta u(t-1) + T(q^{-1})\xi(t) \quad (5.5)$$

onde a ação do operador $\Delta = 1 - q^{-1}$ é assegurar uma ação integral no controlador e cancelar assim efeitos de distúrbios do tipo degrau na saída, como por exemplo perturbações na carga, comuns no controle de processos industriais e especialmente em processos de lodo ativado.

É no modelo (5.5) que o critério de controle quadrático ótimo parcialmente restrito é enunciado, em função da saída e da entrada incremental:

$$J(u, t) = E\left\{ \sum_{k=N_1}^{N_2} [\hat{y}(t+k/t) - r(t+k)]^2 + \lambda_u \sum_{k=1}^{N_c} [\Delta u(t+k-1)]^2 \right\} \quad (5.6)$$

$$s.a. : \Delta u(t+k) = 0 \quad , k = N_c, \dots, N_2$$

onde N_1 e N_2 são os horizontes de custo mínimo e máximo respectivamente. N_c é o horizonte de controle e $r(t)$ é o sinal de referência ou *set-point*. A constante positiva λ_u pondera a importância relativa entre o erro de rastreamento da predição com a ‘energia’ de controle. A esperança matemática em (5.6) é usada para indicar que os valores de controle escolhidos são calculados condicionados aos dados disponíveis até o instante t e assumindo o modelo de distúrbio aleatório. Desta forma, o cálculo da ação de controle do GPC envolve a solução de um problema de controle ótimo de horizonte finito padrão.

Predições ótimas e solução GPC

Para resolver o problema enunciado mediante a minimização de (5.6), tem-se que calcular um conjunto de predições k -passos-à frente da saída: $\hat{y}(t+k)$, para $k = N_1, \dots, N_2$ com a informação coletada até o instante t e nos valores futuros dos incrementos de controle. Estas predições são obtidas com o modelo de planta CARIMA, a partir da seguinte equação de Diophantine:

$$T(q^{-1}) = E_k(q^{-1})A(q^{-1})\Delta + q^{-k}F_k(q^{-1}) \quad (5.7)$$

onde E_k e F_k são polinomiais com graus $k-1$ e n_a , respectivamente. Usando (5.1) e (5.7) pode-se obter:

$$y(t+k) = \frac{B(q^{-1})}{A(q^{-1})}u(t+k-1) + E_k(q^{-1})\xi(t+k) + \frac{F_k(q^{-1})}{A(q^{-1})\Delta} \quad (5.8)$$

Ao colocar em evidência $\xi(t)$ da equação (5.5) e substituí-lo em (5.8), consegue-se:

$$y(t+k) = \frac{F_k(q^{-1})}{T(q^{-1})}y(t) + \frac{E_k(q^{-1})B(q^{-1})}{T(q^{-1})}\Delta u(t+k-1) + E_k(q^{-1})\xi(t+k) \quad (5.9)$$

Sabe-se que E_k tem grau $k - 1$, portanto os componentes da seqüência $\xi(t)$ estarão todos no futuro podendo ser considerados nulos no preditor ótimo, dados os sinais medidos até o instante t e qualquer $u(t + i)$ sendo $i > 1$, dando como resultado a predição de variância mínima da saída:

$$\hat{y}(t + k/t) = \frac{F_k(q^{-1})}{T(q^{-1})}y(t) + \frac{E_k(q^{-1})B(q^{-1})}{T(q^{-1})}\Delta u(t + k - 1) \quad (5.10)$$

Nesta expressão, $\hat{y}(t + k/t)$ é função dos valores conhecidos até o instante t e das entradas de controle futuras, ainda a serem calculadas. Portanto, para distinguir entre os valores de controle passados e futuros usa-se a seguinte equação de Diophantine:

$$E_k(q^{-1})B(q^{-1}) = G_k(q^{-1})T(q^{-1}) + q^{-k}\Gamma_k(q^{-1}) \quad (5.11)$$

que substituída na expressão (5.10), resulta na seguinte predição:

$$\hat{y}(t + k/t) = G_k(q^{-1})\Delta u(t + k - 1) + \Gamma_k(q^{-1})u^f(t - 1) + F_k(q^{-1})y^f(t) \quad (5.12)$$

ou equivalentemente:

$$\hat{y}(t + k/t) = G_k(q^{-1})\Delta u(t + k - 1) + \hat{y}_{rl}(t + k/t) \quad (5.13)$$

onde o primeiro termo do lado direito corresponde à resposta forçada (predita), devida aos futuros incrementos de controle $\Delta u(t + k - 1)$, $k \geq 0$. O segundo termo, $\hat{y}_{rl}(t + k/t)$, é a predição da resposta livre (predita) de $y(t + k)$, assumindo que os futuros incrementos de controle depois de tempo $t + 1$ são nulos.

Na expressão (5.12), $u^f(t - 1)$ e $y^f(t)$, são versões filtradas de $\Delta u(t)$ e $y(t)$:

$$\begin{aligned} u^f(t) &= T^{-1}(q^{-1})\Delta u(t) \\ y^f(t) &= T^{-1}(q^{-1})y(t) \end{aligned} \quad (5.14)$$

Assim, a resposta livre predita é dada por

$$\hat{y}_{rl}(t + k/t) = \Gamma_k(q^{-1})u^f(t - 1) + F_k(q^{-1})y^f(t) \quad (5.15)$$

As equações de Diophantine podem ser resolvidas recursivamente para obter as predições ótimas $\hat{y}(t + k)$, em lugar de realizar cálculos individuais, procedimento que reduz consideravelmente o tempo de cômputo das predições, como mostrado em Camacho e Bordons (1995).

Pode-se definir agora um vetor \mathbf{f} , composto das predições da resposta livre de $\hat{y}(t + k/t)$, definidas pela equação (5.15):

$$\mathbf{f} = [\hat{y}_{rl}(t + 1/t), \hat{y}_{rl}(t + 2/t), \dots, \hat{y}_{rl}(t + N_2/t)]^T \quad (5.16)$$

e também, o vetor de futuros incrementos de controle $\tilde{\mathbf{u}}$:

$$\tilde{\mathbf{u}} = [\Delta u(t), \Delta u(t + 1), \dots, \Delta u(t + N_c - 1), \dots, \Delta u(t + N_2 - 1)]^T \quad (5.17)$$

e o vetor de saídas controladas e previstas:

$$\hat{y} = [\hat{y}(t + 1/t), \hat{y}(t + 2/t), \dots, \hat{y}(t + N_2/t)]^T \quad (5.18)$$

Portanto, a partir das equações de predição mostradas em (5.13), podem-se escrever as predições de horizonte estendido:

$$\begin{bmatrix} \hat{y}(t + 1/t) \\ \hat{y}(t + 2/t) \\ \vdots \\ \hat{y}(t + N_c/t) \\ \vdots \\ \hat{y}(t + N_2/t) \end{bmatrix} = \begin{bmatrix} g_0 & \cdots & \cdots & 0 \\ g_1 & g_0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ g_{N_c-1} & g_{N_c-2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ g_{N_2-1} & g_{N_2-2} & \cdots & g_0 \end{bmatrix} \begin{bmatrix} \Delta u(t) \\ \Delta u(t + 1) \\ \vdots \\ \Delta u(t + N_c - 1) \\ \vdots \\ \Delta u(t + N_2 - 1) \end{bmatrix} + \begin{bmatrix} \hat{y}_{rl}(t + 1/t) \\ \hat{y}_{rl}(t + 2/t) \\ \vdots \\ \hat{y}_{rl}(t + N_c/t) \\ \vdots \\ \hat{y}_{rl}(t + N_2/t) \end{bmatrix} \quad (5.19)$$

onde as equações da resposta livre predita correspondem a:

$$\begin{bmatrix} \hat{y}_{rl}(t + 1/t) \\ \hat{y}_{rl}(t + 2/t) \\ \vdots \\ \hat{y}_{rl}(t + N_c/t) \\ \vdots \\ \hat{y}_{rl}(t + N_2/t) \end{bmatrix} = \begin{bmatrix} \Gamma_1(q^{-1})u^f(t - 1) + F_1(q^{-1})y^f(t) \\ \Gamma_2(q^{-1})u^f(t - 1) + F_2(q^{-1})y^f(t) \\ \vdots \\ \Gamma_{N_c}(q^{-1})u^f(t - 1) + F_{N_c}(q^{-1})y^f(t) \\ \vdots \\ \Gamma_{N_2}(q^{-1})u^f(t - 1) + F_{N_2}(q^{-1})y^f(t) \end{bmatrix}$$

A matriz composta dos parâmetros g_i correspondentes à resposta ao degrau do modelo de planta CARIMA, mostrada em (5.19), é denotada por \mathbf{G} . Note que a matriz \mathbf{G} é uma matriz triangular inferior de dimensão $N_2 \times N_2$. Uma outra forma de expressar a matriz \mathbf{G} pode ser obtida usando uma restrição no critério de controle quadrático GPC da equação (5.6), isto é que $\Delta u(t + k) = 0$, para $N_c \leq k \leq N_2$

$$\bar{u} = [\Delta u(t), \Delta u(t + 1), \dots, \Delta u(t + N_c - 1), 0, 0, \dots]^T \quad (5.20)$$

sendo a matriz \mathbf{G} re-escrita como

$$\mathbf{G} = \begin{bmatrix} g_0 & 0 & \cdots & 0 \\ g_1 & g_0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ g_{N_c-1} & g_{N_c-2} & \cdots & g_0 \\ \vdots & \vdots & & \vdots \\ g_{N_2-1} & g_{N_2-2} & \cdots & g_{N_2-N_c} \end{bmatrix}$$

onde as dimensões da matriz são $N_2 \times N_c$. A introdução desta restrição traz consigo uma favorável

redução computacional no cálculo das predições da saída. O efeito de alterar N_1 é simplesmente apagar filas superiores de \mathbf{G} .

As relações de predição entrada-saída da planta mostradas em (5.19) podem ser agora escritas em forma compacta como uma equação vetorial:

$$\hat{\mathbf{y}} = \mathbf{G}\bar{\mathbf{u}} + \mathbf{f} \quad (5.21)$$

A solução do vetor de mínimo incremento de controle, $\bar{\mathbf{u}}$, resulta da minimização da seguinte função quadrática irrestrita, re-escrita a partir do funcional (5.6):

$$\begin{aligned} \mathbf{J}(\bar{\mathbf{u}}) &= (\hat{\mathbf{y}} - \mathbf{r})^T (\hat{\mathbf{y}} - \mathbf{r}) + \lambda \bar{\mathbf{u}}^T \bar{\mathbf{u}} \\ &= \bar{\mathbf{u}}^T \mathbf{G}^T \mathbf{G} \bar{\mathbf{u}} + 2\mathbf{f}^T \mathbf{G} \bar{\mathbf{u}} + \mathbf{f}^T \mathbf{f} - 2\bar{\mathbf{u}}^T \mathbf{G}^T \mathbf{r} - 2\mathbf{f}^T \mathbf{r} + \lambda \bar{\mathbf{u}}^T \bar{\mathbf{u}} \end{aligned} \quad (5.22)$$

Para minimizar este funcional o conjunto de gradientes é igualado a 0:

$$\nabla_{\bar{\mathbf{u}}} = 0 \quad : \quad 2\mathbf{G}^T \mathbf{G} \bar{\mathbf{u}} + 2\mathbf{f} \mathbf{G}^T - 2\mathbf{G}^T \mathbf{r} + 2\lambda \bar{\mathbf{u}} = 0$$

resultando

$$\bar{\mathbf{u}} = (\mathbf{G}^T \mathbf{G} + \lambda_u \mathbf{I})^{-1} \mathbf{G}^T (\mathbf{r} - \mathbf{f}) \quad (5.23)$$

onde \mathbf{r} é o vetor contendo a seqüência do sinal de referência:

$$\mathbf{r} = [r(t+1), r(t+2), \dots, r(t+N_2)]^T \quad (5.24)$$

O mecanismo utilizado para fechar a malha e forçar um controle de realimentação no GPC é implementar somente o primeiro elemento de $\bar{\mathbf{u}}$, isto é, $\Delta u(t)$. Depois a estratégia do horizonte retrocedente é aplicada para calcular a solução para o problema de controle ótimo para o passo seguinte, usando os dados disponíveis até o instante $t+1$. Dado que somente é tomado o primeiro elemento de $\Delta u(t)$, a ação de controle incremental resultante é escrita como:

$$\Delta u(t) = \bar{g}^T (\mathbf{r} - \mathbf{f}) \quad (5.25)$$

onde \bar{g}^T é a primeira fila de $(\mathbf{G}^T \mathbf{G} + \lambda_u \mathbf{I})^{-1} \mathbf{G}^T$, resultando na ação de controle do GPC:

$$u(t) = u(t-1) + \bar{g}^T (\mathbf{r} - \mathbf{f}) \quad (5.26)$$

Assim, a estratégia de controle em malha fechada é incluída na etapa do projeto, incorporando o cálculo da matriz de ganho do controlador, $K_{gpc} = (\mathbf{G}^T \mathbf{G} + \lambda_u \mathbf{I})^{-1} \mathbf{G}^T$ e os filtros, junto com a implementação envolvendo a construção do sinal \mathbf{f} e do cálculo de controle, $\bar{\mathbf{u}}$, via a equação (5.23).

Considerações em Malha Fechada

Nesta seção é formulado o projeto do controlador preditivo generalizado em malha fechada, incorporando uma técnica de controle linear não adaptativo, como é o projeto do controlador por

alocação de pólos para uma entrada de controle $u(t)$, seqüência de referência $w(t)$ e saída controlada $y(t)$:

$$R(q^{-1})u(t) = H(q^{-1})w(t) + S(q^{-1})y(t) \quad (5.27)$$

Esta formulação será demonstrada ser um equivalente do GPC em malha fechada e será depois unificada na seção 5.1.3 com o esquema de estimação de parâmetros recursivos, tal como abordado em Wellstead e Zarrop (1991). Isto é baseado no Princípio de Equivalência-Certeza (*certainty equivalence principle*) (Goodwin e Sin, 1984), onde uma lei de controle existente pode ser combinada com uma técnica de estimação de parâmetros.

Para obter uma formulação em malha fechada equivalente à de um controlador linear realimentado, foi assumido que os coeficientes da primeira fila da matriz (5.23) são os números reais α_i , descritos pela seguinte relação para $i = 1, \dots, N_2$, com $N_1 = 1$:

$$(\alpha_1, \alpha_2, \dots, \alpha_{N_2})^T = (1, 0, \dots, 0)^T (\mathbf{G}^T \mathbf{G} + \lambda \mathbf{I})^{-1} \mathbf{G}^T \quad (5.28)$$

Porém o primeiro elemento pode ser re-escrito segundo:

$$\begin{aligned} \Delta u(t) &= (1, 0, \dots, 0)^T (\mathbf{G}^T \mathbf{G} + \lambda_u \mathbf{I})^{-1} \mathbf{G}^T (\mathbf{r} - \mathbf{f}) \\ &= (\alpha_1, \alpha_2, \dots, \alpha_{N_2})^T (\mathbf{r} - \mathbf{f}) \end{aligned} \quad (5.29)$$

Utilizando as relações (5.24) e (5.15) consegue-se a seguinte equivalência:

$$\Delta u(t) = \sum_{i=1}^{N_2} \alpha_i r(t+i) - \sum_{i=1}^{N_2} \alpha_i \frac{\Gamma_i(q^{-1})}{T(q^{-1})} \Delta u(t-1) - \sum_{i=1}^{N_2} \alpha_i \frac{F_i(q^{-1})}{T(q^{-1})} y(t) \quad (5.30)$$

que re-arranjado:

$$(T(q^{-1}) + \sum_{i=1}^{N_2} \alpha_i q^{-1} \Gamma_i(q^{-1})) \Delta u(t) = T(q^{-1}) \sum_{i=1}^{N_2} \alpha_i r(t+i) - \sum_{i=1}^{N_2} \alpha_i F_i(q^{-1}) y(t) \quad (5.31)$$

resulta na seguinte equação equivalente do controlador GPC em malha fechada, tal como usado em Clarke (1996):

$$R(q^{-1}) \Delta u(t) = \alpha T(q^{-1}) w(t) - S(q^{-1}) y(t) \quad (5.32)$$

de onde podem ser obtidas as equivalências para os polinômios R e S , assumindo $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_{N_2})^T$:

$$\begin{aligned} R(q^{-1}) &= T(q^{-1}) + \sum_{i=1}^{N_2} \alpha_i q^{-1} \Gamma_i(q^{-1}), \\ S(q^{-1}) &= \sum_{i=1}^{N_2} \alpha_i F_i(q^{-1}) \end{aligned} \quad (5.33)$$

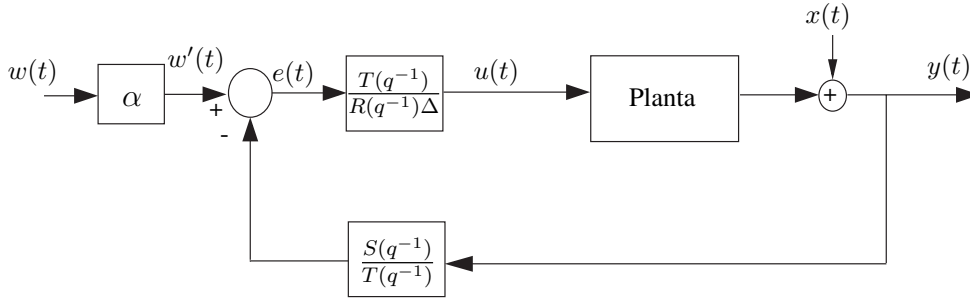


Fig. 5.1: Sistema de controle em malha fechada com os polinômios R e S .

É assumido também que a seqüência de referência $r(t+i) = w(t)$, sendo $w(t)$ o *set-point* para os instantes futuros $i = 1, \dots, N_2$. Esta suposição é válida em controle de processos. Em aplicações tipo batelada, especialmente em robótica, *set-points* programados podem também ser facilmente introduzidos no GPC (Clarke, 1988).

A equação (5.32) mostra o GPC como uma forma de síntese de controlador linear por realimentação mas que incorpora um critério de otimização. Pode ser demonstrado que o GPC equivalente em malha fechada é sintetizado em função dos polinômios R e S como mostrado na Figura 5.1, onde estes polinômios são projetados para que a saída do sistema $y(t)$ rastreie as mudanças do sinal de referência $r(t)$ em presença de distúrbio aleatório. Esta forma é especialmente útil na implementação do Controlador Preditivo Generalizado em Simulink[®], na sua versão auto-ajustável.

Os parâmetros de projeto do controlador N_1 , N_2 , N_c e λ_u influenciam o desempenho do controlador em malha fechada. Isto pode ser demonstrado usando as definições de R , S e α , obtidas em (5.32). Para isto, combina-se o modelo da planta (5.5) com a equação do controlador (5.32):

$$(A(q^{-1})\Delta R(q^{-1}) + q^{-1}B(q^{-1})S(q^{-1}))y(t) = B(q^{-1})T(q^{-1})w(t) + T(q^{-1})R(q^{-1})\xi(t) \quad (5.34)$$

O lado esquerdo pode ser escrito em função das definições de R e S :

$$\begin{aligned} & A(q^{-1})\Delta R(q^{-1}) + q^{-1}B(q^{-1})S(q^{-1}) \\ &= A(q^{-1})\Delta(T(q^{-1}) + \sum_{i=1}^{N_2} \alpha_i \Gamma_i(q^{-1})q^{-1}) + q^{-1}B(q^{-1}) \sum_{i=1}^{N_2} \alpha_i F_i(q^{-1}) \\ &= A(q^{-1})\Delta T(q^{-1}) + \sum_{i=1}^{N_2} \alpha_i (A(q^{-1})\Delta \Gamma_i(q^{-1}) + B(q^{-1})F_i(q^{-1}))q^{-1} \end{aligned} \quad (5.35)$$

que por sua vez é re-escrita usando as equações de Diophantine: (5.7) e (5.11)

$$\begin{aligned}
A(q^{-1})\Delta R(q^{-1}) + q^{-1}B(q^{-1})S(q^{-1}) \\
&= A(q^{-1})\Delta(T(q^{-1}) + T(q^{-1})\sum_{i=1}^{N_2}\alpha_i(B(q^{-1}) - A(q^{-1})\Delta G_i(q^{-1}))q^{i-1}) \\
&= T(q^{-1})(A(q^{-1})\Delta + \sum_{i=1}^{N_2}\alpha_i(B(q^{-1}) - A(q^{-1})\Delta G_i(q^{-1}))q^{i-1}) \\
&= T(q^{-1})A_c
\end{aligned} \tag{5.36}$$

Da equação (5.34) pode-se obter:

$$y(t) = \frac{B(q^{-1})}{A_c}w(t) + \frac{R(q^{-1})}{A_c}\xi(t) \tag{5.37}$$

Pode-se notar que o polinômio observador $T(q^{-1})$ é cancelado na função de transferência em malha fechada e que a estabilidade e desempenho do sistema em malha fechada estarão governados pelos pólos do polinômio A_c . Da observação do polinômio A_c em (5.36), evidencia-se que este é uma função complexa do conjunto de parâmetros do controlador GPC padrão: N_1, N_2, N_c, λ_u . É importante ressaltar que apesar de a dinâmica do observador não aparecer na função de transferência em malha fechada desde $r(t)$ a $y(t)$, esta dinâmica continuará a aparecer na função de transferência desde o ruído à saída. Portanto, o papel do polinômio observador T será basicamente afetar a rejeição de distúrbios quando existam dinâmicas não modeladas (Bitmead et al., 1990; Robinson e Clarke, 1991).

5.1.2 Formulação Modificada: Controle Preditivo Generalizado usando Modelos de Predição Múltiplos

O modelo preditor com estrutura ARX:

$$\hat{y}(t/t - k, \theta) = \tilde{A}(q^{-1}, \theta)y(t) + \tilde{B}(q^{-1}, \theta)u(t) \tag{5.38}$$

com polinômios de estrutura aproximada \tilde{A} e \tilde{B} :

$$\begin{aligned}
\tilde{A}(q^{-1}, \theta) &= -\alpha_0^{(k)}q^{-k} - \dots - \alpha_{n_a-1}^{(k)}q^{-(n_a+k-1)} \\
\tilde{B}(q^{-1}, \theta) &= \beta_{k-1}^{(k)}q^{-1} + \dots + \beta_0^{(k)}q^{-k} + \gamma_1^{(k)}q^{-(k+1)} + \dots + \gamma_{n_b-1}^{(k)}q^{-(n_b+k-1)}
\end{aligned} \tag{5.39}$$

são o ponto de partida para derivar a formulação modificada do GPC usando modelos de predição múltiplos. Como descrito anteriormente, na seção 4.4.3, os parâmetros α_i, β_i e γ_i são os parâmetros estimados para cada horizonte de predição de forma independente. Usando estes parâmetros, a predição da saída múltiplos-passos-à frente, $\hat{y}(t + k/t)$, obtido pelo desenvolvimento em série de Taylor,

corresponde a:

$$\begin{aligned} \hat{y}(t+k/t) = & \beta_{k-1}^{(k)}u(t+k-1) + \beta_{k-2}^{(k)}u(t+k-2) + \dots + \beta_0^{(k)}u(t) + \\ & \alpha_0^{(k)}y(t) + \alpha_1^{(k)}y(t-1) + \dots + \alpha_{n_a-1}^{(k)}y(t-n_a+1) + \\ & \gamma_1^{(k)}u(t-1) + \gamma_2^{(k)}u(t-2) + \dots + \gamma_{n_b-1}^{(k)}u(t-n_b+1) \end{aligned} \quad (5.40)$$

Note que as duas últimas somas de (5.40) dependem dos valores passados da saída do processo e dos valores passados das variáveis de entrada. Portanto elas podem ser consideradas como a resposta livre (predita), se o sinal de controle é mantido constante. A primeira soma depende somente dos valores futuros do sinal de controle e pode ser interpretada como a resposta forçada (predita), assumindo $y(t-k) = 0$ e $u(t-k) = 0$, para $k = 0, 1, \dots$, ou seja, condições iniciais zero.

As equações de predição para o horizonte de predição mínimo e máximo N_1 e N_2 e para um horizonte de controle N_c são definidas para k abrangendo os valores $k = 1, 2, \dots, N_2$, levando em conta a restrição $u(t+k) = 0$, para $N_c \leq k \leq N_2$:

$$\begin{bmatrix} y(t+1/t) \\ y(t+2/t) \\ \vdots \\ y(t+N_c/t) \\ \vdots \\ y(t+N_2/t) \end{bmatrix} = \begin{bmatrix} \beta_0^{(1)} & 0 & \dots & 0 \\ \beta_0^{(2)} & \beta_1^{(2)} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ \beta_0^{(N_c)} & \beta_1^{(N_c)} & \dots & \beta_{N_c-1}^{(N_c)} \\ \vdots & \vdots & & \vdots \\ \beta_0^{(N_2)} & \beta_1^{(N_2)} & \dots & \beta_{N_2-N_c}^{(N_2)} \end{bmatrix} \begin{bmatrix} u(t) \\ u(t+1) \\ \vdots \\ u(t+N_c-1) \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} \hat{y}_{rl}^{(1)}(t+1/t) \\ \hat{y}_{rl}^{(2)}(t+2/t) \\ \vdots \\ \hat{y}_{rl}^{(N_c)}(t+N_c/t) \\ \vdots \\ \hat{y}_{rl}^{(N_2)}(t+N_2/t) \end{bmatrix} \quad (5.41)$$

Note que as relações de predição entrada-saída mostradas em sua forma estendida em (5.41) apresentam a estrutura vista na formulação padrão do GPC. Estas relações de predição podem ser escritas em forma compacta através da equação vetorial:

$$\hat{\mathbf{y}} = \mathcal{B}\mathbf{u} + \tilde{\mathbf{f}} \quad (5.42)$$

definida para o vetor de saídas controladas e preditas, $\hat{\mathbf{y}}$, para o vetor de futuras ações de controle, \mathbf{u} e para o vetor das predições da resposta livre, $\tilde{\mathbf{f}}$, dados por:

$$\begin{aligned} \hat{\mathbf{y}} &= [\hat{y}(t+1/t), \hat{y}(t+2/t), \dots, \hat{y}(t+N_2/t)]^T \\ \mathbf{u} &= [u(t), u(t+1), \dots, u(t+N_c), 0, \dots, 0]^T \\ \tilde{\mathbf{f}} &= [\hat{y}_{rl}^{(1)}(t+1/t), \hat{y}_{rl}^{(2)}(t+2/t), \dots, \hat{y}_{rl}^{(N_2)}(t+N_2/t)]^T \end{aligned}$$

As predições da resposta livre para cada passo de predição, $\hat{y}_{rl}^{(k)}(t+k/t)$, são definidas como:

$$\hat{y}_{rl}^{(k)}(t+k/t) = \mathcal{G}_k(q^{-1})u(t-1) + \mathcal{A}_k(q^{-1})y(t) \quad (5.43)$$

onde os polinômios \mathcal{A}_k e \mathcal{G}_k são definidos para cada passo de predição $k = 1, 2, \dots, N_2$ e para as

ordens do modelo identificado n_a e n_b :

$$\begin{aligned}\mathcal{A}_k(q^{-1}) &= \alpha_0^{(k)} + \alpha_1^{(k)}q^{-1} + \dots + \alpha_{n_a-1}^{(k)}q^{-(n_a-1)} \\ \mathcal{G}_k(q^{-1}) &= \gamma_1^{(k)} + \gamma_2^{(k)}q^{-1} + \dots + \gamma_{n_b-1}^{(k)}q^{-(n_b-2)}\end{aligned}\quad (5.44)$$

O modelo preditor múltiplos-passos-à frente serve então para definir a função objetivo cuja minimização em relação a \mathbf{u} dá a lei de controle composta pelo vetor de entradas futuras de controle. A lei de controle é obtida mediante a minimização do desvio entre a saída controlada e o sinal de referência junto com a entrada de controle, tal como mostrado no funcional quadrático do GPC visto em (5.6). O problema de otimização irrestrito minimiza o funcional:

$$\begin{aligned}\mathbf{J}(\mathbf{u}) &= (\hat{\mathbf{y}} - \mathbf{r})^T (\hat{\mathbf{y}} - \mathbf{r}) + \lambda \mathbf{u}^T \mathbf{u} \\ &= \mathbf{u}^T \mathcal{B}^T \mathcal{B} \mathbf{u} + 2\tilde{\mathbf{f}}^T \mathcal{B} \mathbf{u} + \tilde{\mathbf{f}}^T \tilde{\mathbf{f}} - 2\mathbf{u}^T \mathcal{B}^T \mathbf{r} - 2\tilde{\mathbf{f}}^T \mathbf{r} + \lambda \mathbf{u}^T \mathbf{u}\end{aligned}\quad (5.45)$$

A solução é obtida igualando o conjunto de gradientes a 0:

$$\nabla_{\mathbf{u}} = 0 \quad : \quad 2\mathcal{B}^T \mathcal{B} \mathbf{u} + 2\tilde{\mathbf{f}} \mathcal{B}^T - 2\mathcal{B}^T \mathbf{r} + 2\lambda \mathbf{u} = 0$$

resultando a seqüência de entradas futuras ótimas dadas por:

$$\mathbf{u} = (\mathcal{B}^T \mathcal{B} + \lambda_u \mathbf{I})^{-1} \mathcal{B}^T (\mathbf{r} - \tilde{\mathbf{f}}) \quad (5.46)$$

onde \mathbf{r} é a seqüência do sinal de referência, visto em (5.24). Tal como na formulação padrão, o GPC baseado em modelos de predição múltiplos, denotadas doravante pelas siglas GPC-MSPI, implementa somente o primeiro elemento de \mathbf{u} , isto é $u(t)$. O princípio de horizonte retrocedente é aplicado para calcular uma nova entrada de controle ótima, usando os dados disponíveis no instante $t + 1$. Finalmente a ação de controle do GPC-MSPI resulta sendo:

$$u(t) = \bar{\beta}^T (\mathbf{r} - \tilde{\mathbf{f}}) \quad (5.47)$$

onde $\bar{\beta}^T$ é a primeira fila de $(\mathcal{B}^T \mathcal{B} + \lambda_u \mathbf{I})^{-1} \mathcal{B}^T$.

Assumindo que os coeficientes da primeira fila de (5.46) são os números reais α_i para $i = 1, 2, \dots, N_2$ e $N_1 = 1$, pode-se demonstrar seguindo o procedimento descrito na seção 5.1.1, que a formulação equivalente em malha fechada do controlador corresponde à seguinte relação:

$$\begin{aligned}(1 + \sum_{i=1}^{N_2} \alpha_i q^{-1} \mathcal{G}_i(q^{-1}))u(t) &= \sum_{i=1}^{N_2} \alpha_i r(t+i) - \sum_{i=1}^{N_2} \alpha_i \mathcal{A}(q^{-1})y(t) \\ &= \sum_{i=1}^{N_2} \alpha_i r(t+N_2)q^{-(N_2-i)} - \sum_{i=1}^{N_2} \alpha_i \mathcal{A}(q^{-1})y(t)\end{aligned}\quad (5.48)$$

Assume-se que a seqüência de referência é conhecida ou pre-determinada para os instantes futuros, portanto $r(t+N_2) = w(t)$, suposição que resulta sendo válida em aplicações de controle de

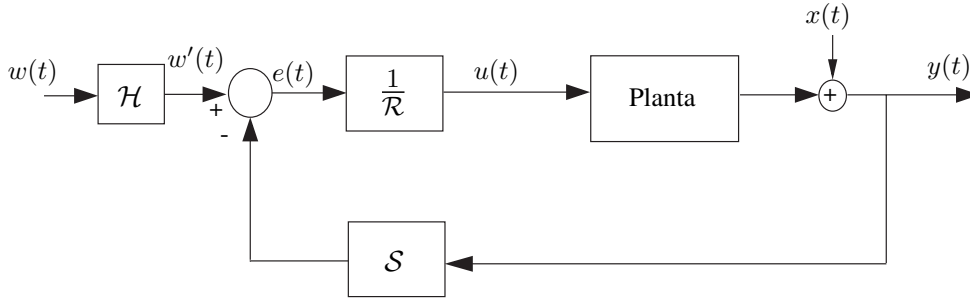


Fig. 5.2: Sistema de controle em malha fechada do controlador GPC-MSPI com os polinômios \mathcal{R} , \mathcal{S} e \mathcal{H} .

processos. Logo, a formulação em malha fechada do controlador GPC-MSPI resulta sendo:

$$\mathcal{R}(q^{-1})u(t) = \mathcal{H}(q^{-1})w(t) - \mathcal{S}(q^{-1})y(t) \quad (5.49)$$

onde os polinômios \mathcal{R} , \mathcal{S} e \mathcal{H} são dados por:

$$\begin{aligned} \mathcal{R}(q^{-1}) &= \left(1 + \sum_{i=1}^{N_2} \alpha_i q^{-i} \mathcal{G}_i(q^{-1})\right) \\ \mathcal{S}(q^{-1}) &= \sum_{i=1}^{N_2} \alpha_i \mathcal{A}(q^{-1}) \\ \mathcal{H}(q^{-1}) &= \sum_{i=1}^{N_2} \alpha_i q^{-(N_2-i)} \end{aligned}$$

De onde é obtido o GPC baseado em identificação de preditores múltiplos-passos-à frente (GPC-MSPI) equivalente em malha fechada, sintetizado em função dos polinômios \mathcal{R} , \mathcal{S} e \mathcal{H} , como mostrado no esquema da Figura 5.2. Esta estrutura em malha fechada do controlador em malha fechada do GPC-MSPI foi implementado em Simulink[®], de forma similar ao GPC padrão.

5.1.3 GPC auto-ajustável

Utilizando o Princípio de Equivalência-Certeza é possível unificar o projeto do controlador GPC não adaptativo com o do estimador de parâmetros, como mostrado no esquema da Figura 5.3. O controlador GPC auto-ajustável pode ser visto como duas malhas. Uma malha interna que consiste do controlador linear por realimentação sintetizado nos polinômios \mathcal{R} , \mathcal{S} e α da formulação padrão (ou \mathcal{R} , \mathcal{S} e \mathcal{H} da formulação do GPC baseado no modelo preditor múltiplos-passos-à frente). Os polinômios do controlador são então ajustados pela malha externa, que é composta do estimador recursivo de mínimos quadrados e do projeto do controlador GPC. O estimador recursivo atualizará os parâmetros do modelo identificado, usando as novas medidas dos dados entrada-saída, para cada instante t . Controladores que apresentam esta configuração são denominados auto-ajustáveis (GPC-STR).

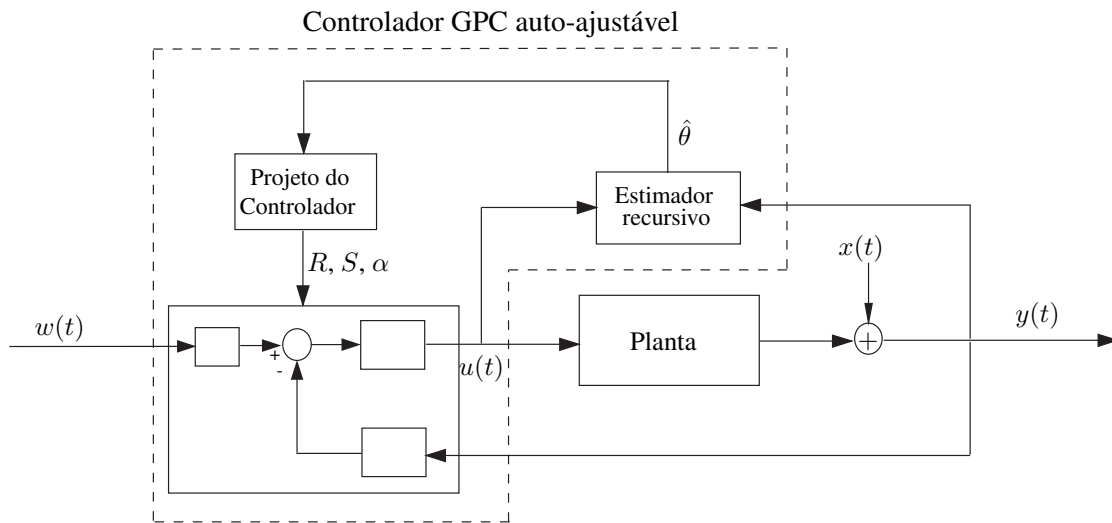


Fig. 5.3: Esquema do controlador GPC auto-ajustável baseado nos modelos preditores.

Estimador Recursivo

No estimador recursivo do controlador GPC se utilizam algoritmos recursivos de mínimos quadrados, que como explicado, atualizam o valor dos parâmetros do modelo reduzido ou identificado. Estes algoritmos utilizam o modelo de regressão linear:

$$\hat{y}(t/\theta) = \theta^T \varphi(t) + e(t) \quad (5.50)$$

descrito na seção 4.4.1. Esta regressão é utilizada para atualizar os θ -parâmetros do modelo identificado. O procedimento é descrito brevemente a seguir. O erro de predição um-passo-à frente:

$$\epsilon(t) = y(t) - \hat{y}(t/t-1) \quad (5.51)$$

que define o desvio da saída medida no instante t em relação à predição da saída um-passo-à frente, é utilizado pelo algoritmo de mínimos quadrados recursivos para selecionar um $\hat{\theta}$, que minimize o critério $\sum \epsilon^2$ com base nos dados entrada-saída registrados ou medidos. A solução deste problema resulta na equação normal:

$$\hat{\theta}(t) = \left[\sum_{i=1}^t \varphi^T(i) \varphi(i) \right]^{-1} \sum_{i=1}^t \varphi^T(i) y(i) \quad (5.52)$$

utilizada para atualizar os parâmetros segundo:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \mathbf{P}(t) \varphi(t) \epsilon(t) \quad (5.53)$$

onde a matriz de covariância $\mathbf{P}(t) = \left[\sum_{i=1}^t \varphi^T(i)\varphi(i) \right]^{-1}$ é atualizada segundo:

$$\begin{aligned} \mathbf{K}(t) &= \frac{\mathbf{P}(t-1)\varphi(t)}{1 + \varphi^T\mathbf{P}(t-1)\varphi(t)} \\ \mathbf{P}(t) &= [\mathbf{I} - \mathbf{K}(t)\varphi^T(t)]\mathbf{P}(t-1) \end{aligned} \quad (5.54)$$

Sabe-se que o grande inconveniente do algoritmo é que $\mathbf{P}(t)$ pode ser mal condicionado. Portanto, na literatura do controle auto-ajustável se prefere utilizar versões numericamente mais robustas do algoritmo recursivo de mínimos quadrados, como são os algoritmos que fatoram a matriz de covariância. O algoritmo de fatoração U-D de Bierman é introduzido então como uma modificação no algoritmo de mínimos quadrados recursivos dentro do projeto do controlador GPC como aconselhado em Clarke (1996).

Basicamente o algoritmo U-D assume $\mathbf{P}(t)$ como fatorado por:

$$\mathbf{P}(t) = \mathbf{U}(t)\mathbf{D}(t)\mathbf{U}^T(t) \quad (5.55)$$

onde $\mathbf{U}(t)$ é uma matriz triangular superior com diagonal unitária e $\mathbf{D}(t)$ é uma matriz diagonal. O algoritmo calcula as equações do estimador de mínimos quadrados obtendo recursivamente $\mathbf{U}(t)$ e $\mathbf{D}(t)$ a partir de $\mathbf{U}(t-1)$ e $\mathbf{D}(t-1)$. Isto resulta no cálculo do ganho do estimador, $\mathbf{L}(t)$, que é então utilizado para atualizar os parâmetros estimados $\theta(t)$ sendo conhecido o valor anterior $\theta(t-1)$, segundo:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \mathbf{L}(t)\epsilon(t) \quad (5.56)$$

onde o ganho $\mathbf{L}(t)$ é definido pelo vetor $\mathbf{P}(t)\varphi(t)$, como assumido usualmente na formulação original do algoritmo recursivo de mínimos quadrados.

5.2 Desempenho do controlador preditivo usando os diferentes modelos de predição

A seguir são mostrados nesta seção os resultados da implementação do controlador GPC adaptativo baseados nos modelos preditores por iteração sucessiva padrão (modelo LS), de identificação preditiva de horizonte estendido (modelo LRPI) e de múltiplos-passos-à frente (modelo MSPI).

Como planta é considerada o modelo de Rohrs visto na seção 4.4.6. O intervalo de amostragem é 0,01s. Para todos os casos é avaliado o desempenho do controlador para mudanças no *set-point* ou na referência, em seqüências de mudanças do tipo degrau a cada 75s. A matriz de covariância para os algoritmos recursivos é $10I$ em todos os casos. O parâmetro inicial $\hat{b}_1(0)$ é fixado em 1, enquanto que os outros são considerados nulos no instante inicial. Os parâmetros de ajuste dos controladores GPC são $N_1 = N_c = 3$, $N_2 = 10$, $\lambda_u = 10$. São simulados dois cenários, no primeiro deles os controladores estão livres de ruído na saída, enquanto que no segundo é adicionado ruído gaussiano de média nula e variância 0,01.

Na Figura 5.4 pode-se ver a comparação dos desempenhos do controlador GPC adaptativo ba-

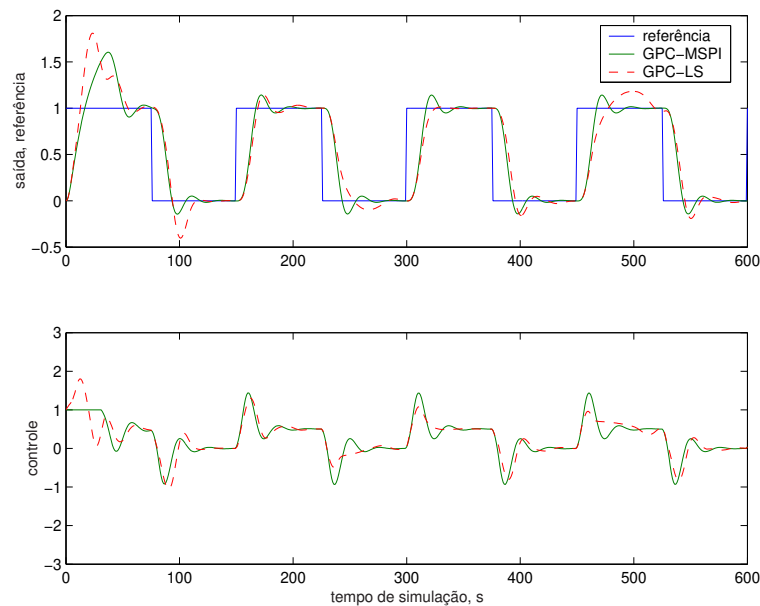


Fig. 5.4: Desempenhos do controlador GPC com modelo MSPI e modelo LS, sem ruído aditivo na saída.

seado em previsões do modelo LS (GPC-LS) e do modelo MSPI (GPC-MSPI). O desempenho é avaliado para mudanças na referência mencionadas no parágrafo anterior. Note que o GPC-LS apresenta instabilidade em malha fechada, com sobre-picos maiores que os do controlador GPC-MSPI. Note também que o GPC-LS requer intervalos de tempo maiores para atingir o estado estacionário.

Na Figura 5.5 é comparado o desempenho entre o GPC-MSPI e o controlador GPC baseado em identificação preditiva de horizonte estendido (GPC-LRPI). O transitório observado nos intervalos iniciais da simulação são próprios de estimadores recursivos com parâmetros iniciais diferentes dos de estado estacionário. Neste caso o controlador GPC-MSPI é ligeiramente superior, pois apresenta sobre-picos menores. As pequenas oscilações observadas são próprias de controladores adaptativos de ordem reduzida.

A vantagem na utilização de previsões de múltiplos-passos-à frente são evidentes quando é adicionado ruído na medição, suposição que é realista para a maioria de casos práticos, especialmente no controle de processos industriais. Os parâmetros do controlador são os mesmos considerados no caso de ausência de ruído para o modelo de ordem reduzida. Na Figura (5.6) se mostra o desempenho do GPC-LS. Note o comportamento oscilatório em malha fechada, em que a perda de convergência do controlador em atingir a referência é evidente, especialmente no intervalo de 125 a 225 s. Obviamente isto geraria a saturação de atuadores reais, fenômeno indesejável em plantas de tratamento.

A resposta em malha fechada do controlador GPC-LRPI é mostrado na Figura 5.7, onde pode ser visto que o GPC-LRPI mostra-se mais robusto ante uma perturbação na medição quando comparado com o desempenho do controlador GPC-LS, o que coincide com as observações feitas no trabalho de Fardin (2000). No entanto, uma melhor adaptabilidade e robustez pode ser observada na resposta em malha fechada do controlador GPC-MSPI, como visto na Figura (5.8) onde tem-se amplitudes de sobre-picos menores.

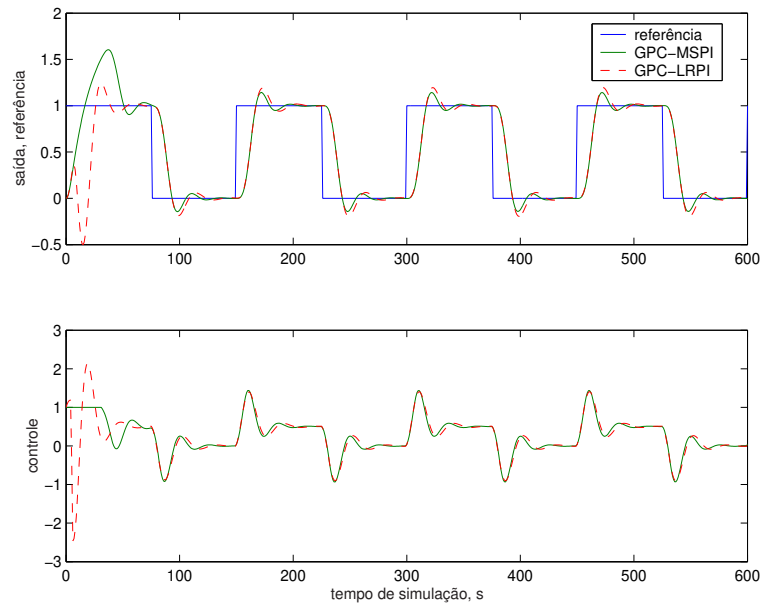


Fig. 5.5: Desempenhos do controlador GPC com modelo MSPI e com modelo LRPI, sem ruído aditivo na saída.

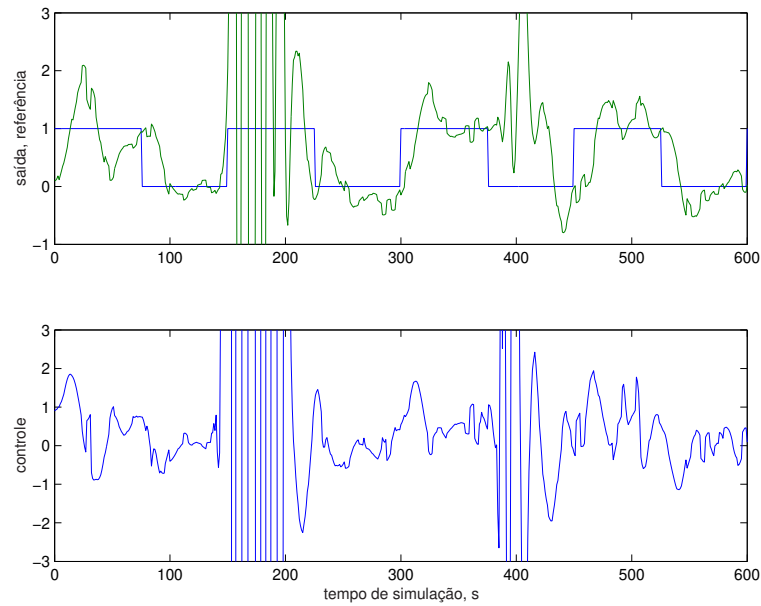


Fig. 5.6: Desempenho do controlador GPC com modelo LS, com ruído aditivo na saída.

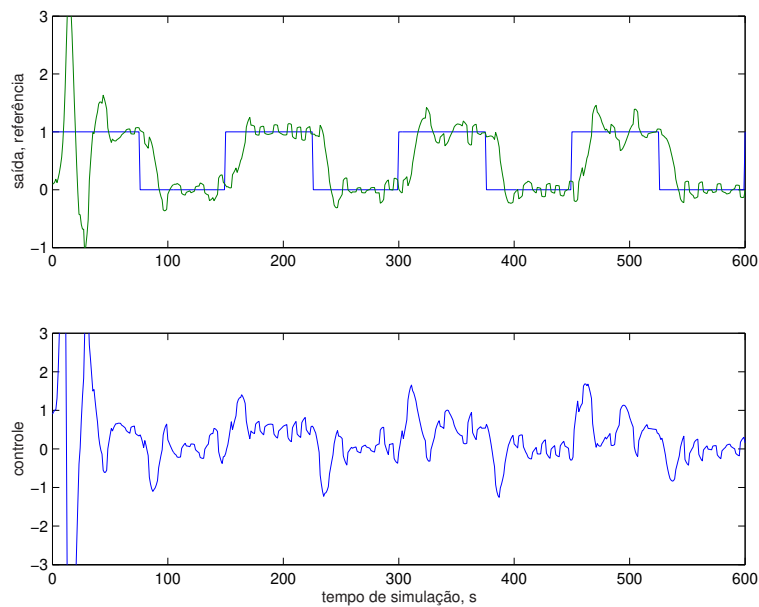


Fig. 5.7: Desempenho do controlador GPC com modelo LRPI, com ruído aditivo na saída.

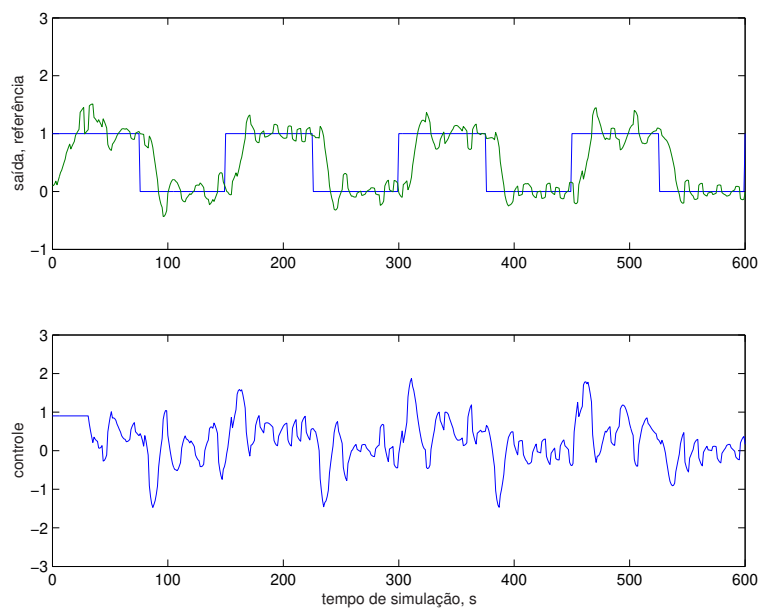


Fig. 5.8: Desempenho do controlador GPC com modelo MSPI, com ruído aditivo na saída.

Tab. 5.1: Índices de erro do desempenho em malha fechada dos controladores GPC sem ruído na saída.

$\sigma_y^2 = 0,0$	N_2	N_1	λ_u	J_p	J_u
GPC-LS	10	3	10	95,948	21,639
GPC-LRPI	10	3	10	99,638	32,637
GPC-MSPI	10	3	10	84,888	23,878

Tab. 5.2: Índices de erro do desempenho em malha fechada dos controladores GPC com ruído na saída.

$\sigma_y^2 = 0,01$	N_2	N_1	λ_u	J_p	J_u
GPC-LS	10	3	10	$9,71 \times 10^3$	$3,86 \times 10^3$
GPC-LRPI	10	3	10	145,173	98,713
GPC-MSPI	10	3	10	94,166	76,407

Para ilustrar quantitativamente o desempenho dos modelos de predição no projeto dos controladores GPC, introduzem-se dois índices de erro, o índice de erro quadrático de rastreamento da referência, J_p e o índice da magnitude absoluta do incremento da ação de controle, J_u , definidas a seguir:

$$J_p = \sum_{i=1}^{N_{sim}} ((r(i) - y_m(i))^2); \quad (5.57)$$

$$J_u = \sum_{i=1}^{N_{sim}} |u(i) - u(i-1)| \quad (5.58)$$

em que $r(i)$, $y_m(i)$ e $u(i)$ representam os valores da referência, da saída e da entrada da ação de controle no instante de medição i . Apresentam-se os resultados destes índices para o caso de ausência de ruído na saída, na Tabela 5.1 e com ruído de medição, na Tabela 5.2.

Através da inspeção dos resultados pode-se notar que o controlador GPC-MSPI apresenta um menor índice de erro de rastreamento da referência, a pesar de precisar de um maior esforço de controle se comparado com o mostrado pelo controlador GPC-LS. Os resultados mostrados na Tabela 5.2, o caso com ruído na saída, confirma a eficácia do modelo de predição de múltiplos-passos-à frente na melhoria de desempenho em malha fechada. O controlador GPC-MSPI mostra ser robusto a perturbações na saída para mudanças na referência. Os valores dos índices de erro para o controlador GPC-LRPI mostram ser os maiores, embora cabe ressaltar que o tempo de processamento, se comparado com o controlador GPC-MSPI, é bem menor. Isto é explicado em base à complexidade dos

modelos preditores, pois como é sabido o GPC-MSPI estima um modelo para cada passo de predição, cuja ordem aumenta conforme aumenta o passo de predição. Assim, para o modelo estimado de primeira ordem tem-se que no passo de predição $k = 10$, os polinômios MSPI \tilde{A} e \tilde{B} serão de décima ordem. Vale lembrar que a ordem do modelo LRPI é conservada ao longo processo de estimação.

5.3 Conclusão

Neste capítulo mostrou-se por simulação o desempenho em malha fechada do controlador GPC baseado nos três tipos de modelos preditores: o preditor por iteração sucessiva, o preditor por identificação preditiva de horizonte estendido e o preditor de múltiplos-passos-à frente. Os controladores preditivos na identificação de preditores MSPI e LRPI resultam sendo uma alternativa atrativa no caso de existir dinâmicas não modeladas que produzam variâncias do erro de predição cada vez maiores quando o modelo reduzido é estimado usando o algoritmo um-passo-à frente.

Verificou-se nas simulações um melhor comportamento em malha fechada do controlador GPC com predições de múltiplos-passos-à frente, especialmente no caso de perturbações na saída, que degradam o desempenho dos controladores com predições LS e LRPI. Para efeitos de comparação um horizonte de predição máximo igual a 10 foi usado. No entanto, é importante considerar o custo computacional extra presente na estimação do modelo preditor MSPI. Nos casos de que não for problemática a grandeza do modelo MSPI, que aumenta conforme aumenta o horizonte de predição, o uso do controlador GPC-MSPI será recomendado. Caso contrário, será preferido o controlador GPC-LRPI visto que este estima um modelo conservando a ordem do modelo o que significa um menor custo computacional e erros de predição próximos ao do controlador GPC-MSPI se usados horizontes de predição maiores que 10.

Em base a análise realizada é escolhido o controlador GPC-LRPI para a abordagem de controle do processo de remoção biológica de nutrientes, a ser tratado no seguinte capítulo. Esta preferência é devido ao aumento da complexidade do modelo preditor MSPI conforme aumenta o passo de predição, porém aumentando também o tempo de processamento da estratégia de controle em malha fechada. A escolha por um algoritmo de controle mais eficiente é baseada também em problemas numéricos que podem se apresentar no modelo de simulação. É sabido que quando são simulados reatores biológicos, algumas concentrações podem mostrar valores negativos (que não são fisicamente possíveis), provocados em certas condições e distúrbios do sistema que atuam em combinação com a tolerância selecionada do algoritmo de integração numérica, algoritmo que porém precisará de um tempo de integração maior. Este problema é comum nos casos de concentração de oxigênio e nitrato nos reatores aeróbios e anóxicos (Copp, 2000), que correspondem aos processos que são o objetivo de controle, tal como será mostrado no seguinte capítulo.

Capítulo 6

O MPC Aplicado à Remoção Biológica de Nutrientes

A remoção biológica de nutrientes, tais como o nitrogênio e o fósforo, é um tópico que vem ganhando crescente atenção em plantas de lodos ativados nos últimos anos (Von Sperling, 2000; Weijers, 2000; Samuelsson e Carlsson, 2001; Lukasse, 2002). A remoção de nutrientes, além de ter importante repercussão sobre a qualidade do efluente e, conseqüentemente, da água do corpo receptor, também influi fortemente sobre o desempenho do sistema de tratamento. Para assegurar um controle efetivo da qualidade dos cursos de água superficiais, é extremamente importante o controle do lançamento destes compostos.

Em países mais desenvolvidos, a legislação existente para os padrões de lançamento, impõe limites não somente em relação à remoção da matéria orgânica carbonácea, mas também em relação aos nutrientes. Este fato e incentivos econômicos tem servido de estímulo na melhoria e desenvolvimento de estratégias de controle para processos de remoção biológica de nutrientes. A legislação pertinente de países em desenvolvimento, onde persistem ainda os problemas de remoção de matéria carbonácea, considera parcialmente os objetivos de remoção de nutrientes. Por exemplo, no Brasil a Resolução CONAMA No. 20 estabelece o nitrogênio na forma de amônia total em 5,0 mg/l como limite máximo de lançamento. Este não inclui um padrão para o nitrogênio total. Mesmo assim, é importante considerar a implantação de estratégias de controle nos sistemas de remoção de nutrientes nas plantas já existentes, ou nos novos projetos de plantas, remoção esta que pode ser vista como um segundo estágio natural no desenvolvimento dos processos de tratamento de esgotos.

Este Capítulo trata da aplicação do controlador preditivo baseado no modelo preditor por identificação preditiva de horizonte estendido (GPC-LRPI), ao processo de remoção biológica de nutrientes, essencialmente na remoção de nitrogênio. Apesar do processo ser inerentemente multivariável, as diferentes constantes de tempo do sistema de tratamento de esgotos permitem desacoplar o processo de remoção de nitrogênio em malhas separadas. Desta forma, o processo pode ser decomposto no controle do processo de nitrificação na zona aeróbia e no controle do processo de desnitrificação na zona anóxica (Olsson e Newell, 1999). Portanto, nas seções a seguir são discutidos aspectos operacionais e de projeto do controlador proposto para estes dois sub-processos, que são parcialmente descritos em Lara e Milani (2005a) e Lara e Milani (2005b).

6.1 A Remoção Biológica de Nitrogênio

O nitrogênio está presente nas águas residuais de várias formas, tais como: amônio (NH_3), amônio ionizado ou amônio (NH_4^+), nitrato (NO_3^-), nitrito (NO_2^-) e compostos orgânicos. É conhecido que altas concentrações de nitrogênio em águas despejadas nos corpos de água podem provocar diferentes problemas. Por exemplo, nessas condições, o nitrogênio estimula o crescimento de plantas aquáticas e algas, fazendo com que algas e bactérias se reproduzam muito rapidamente, incrementando dessa forma o consumo de oxigênio dissolvido no fenômeno conhecido como eutrofização. Para evitar estes e outros problemas provocados pelas altas concentrações no esgoto tratado, a quantidade de nitrogênio do esgoto afluente deve ser reduzida. Esta é basicamente a finalidade do processo de remoção biológica de nitrogênio.

6.1.1 O aspecto do acoplamento

Como visto na Seção 1.3, o processo de remoção biológica de nitrogênio em uma planta de tratamento de esgotos pode ser realizado através de um sistema de pré-desnitrificação em uma planta de lodos ativados. Quando um sistema de pré-desnitrificação é aplicado, a bactéria desnitrificante utiliza da melhor forma o material orgânico do esgoto afluente. Dado que a nitrificação se desenvolve na zona aerada, que é colocada depois da zona anóxica, água com alta concentração de nitrato do final da zona aeróbia, deve ser recirculada para ingresso da zona anóxica e assim assegurar a presença de nitratos. Para o objetivo de controle proposto, é possível caracterizar os efeitos dos principais acoplamentos entre a zona anóxica e aeróbia como sendo:

1. Redução da capacidade de desnitrificação no reator anóxico devido a um eventual transporte de oxigênio dissolvido (OD) via a recirculação interna do final da zona aeróbia para o início da zona anóxica.
2. Incremento da capacidade desnitrificante do bioreator na zona aeróbia como consequência da ocorrência simultânea dos processos de nitrificação e desnitrificação durante baixas concentrações de OD.
3. O processo de desnitrificação requer o nitrato proporcionado pelo processo de nitrificação. Se o processo de nitrificação é prejudicado, então o processo de desnitrificação também o será.

Os efeitos destes acoplamentos serão fracos se forem observadas as seguintes condições operacionais (Ingildsen, 2002):

- Assegurar baixa concentração de OD na recirculação interna de nitrato.
- Estabelecer como prioridade a remoção de amônio, em lugar de nitrogênio, na zona aeróbia, que pode ser conseguida com um controle do *set-point* de OD. Desta forma, o nível de OD pode ser suficientemente baixo, sem estimular o desenvolvimento do processo de desnitrificação.
- Garantir a formação de nitratos na remoção de amônio, para garantir uma vazão de recirculação rica em nitratos.

Note que implícito a esta análise qualitativa, aparece a forma de desacoplar o processo, segundo a escala de tempo das dinâmicas da concentração de amônio e de nitrato, em dois sub-processos com escalas de tempo similares. Análises quantitativas revelam que o processo de remoção de nutrientes pode ser decomposto, conforme a escala de tempo, em sub-processos de dinâmicas lenta, média e rápida (Steffens et al., 1997; Olsson e Newell, 1999; Ingildsen, 2002).

O problema de controle pode ser adicionalmente simplificado se as variáveis controladas de cada sub-processo não são diretamente as concentrações do efluente, e sim outras variáveis do processo que tenham efeito sobre essas concentrações, como é a concentração de nitrato na desnitrificação e a concentração de amônio na nitrificação. Em Vrečko et al. (2002) foram propostos controladores lineares PI para estas variáveis.

Assim sendo, neste trabalho é adotada a hipótese de que o problema de controle do processo de lodos ativados com pré-desnitrificação pode ser decomposto em um problema de controle do processo de nitrificação no último reator aeróbio, que é influenciado pelo valor de OD na escala de horas a dias (escala de tempo média) e no controle do processo de desnitrificação no reator anóxico influenciado pela vazão de recirculação interna, também na escala de horas a dias.

6.2 Controle da nitrificação

Em um reator nitrificante de uma planta de lodos ativados, oxigênio é consumido como resultado da remoção do carbono orgânico e de amônio. Para operar a planta de forma eficiente, a quantidade de oxigênio dissolvido na zona aeróbia precisa ser controlada. Vale lembrar que os custos do sistema de aeração são em geral os custos operacionais mais dispendiosos. Portanto, sua economia vem a ser um grande incentivo na melhoria do controle de OD.

Estratégias de controle de OD consideram normalmente um valor constante do OD, mediante a variação da vazão de ar injetada para garantir que uma quantidade suficiente de oxigênio seja fornecida ao processo. Uma forma alternativa a considerar um *set-point* constante de OD, é usar um *set-point* de OD variante no tempo, cuja variação possa ser determinada pela concentração de amônio ao final da última zona aeróbia. Esta estratégia de controle pode oferecer um melhor controle do amônio com uma gestão mais eficiente da vazão de injeção de ar (Lindberg, 1997). Esta estratégia, cujo esquema de controle é mostrado na Figura 6.1, é também denominada controle de amônio.

O principal objetivo nesta sub-seção é o controle da concentração de amônio no reator nitrificante através de um *set-point* variável de OD, usando medições *on-line* da concentração de amônio na saída do reator, em uma configuração de controle em cascata, aproveitando o fato de que a malha interna de controle de OD é mais rápida que a malha de controle de amônio (Olsson e Newell, 1999). O controlador preditivo generalizado baseado em identificação preditiva de horizonte estendido (GPC-LRPI) atua como controlador mestre, para manter uma concentração de amônio pré-especificada, mediante a variação do *set-point* de OD. A saída do controlador mestre é então o sinal de referência para o controlador PI de OD proporcionado no modelo *benchmark* ASWWTP-USP (Sotomayor et al., 2001), utilizado como controlador escravo na configuração em cascata (denominada também de controle supervisorio de OD). Assume-se que o OD pode ser controlado para qualquer trajetória do *set-point* na escala de tempo denominada pequena (Olsson e Newell, 1999). Finalmente, o desempenho da estratégia de controle sugerida é testado para verificar o rastreamento do *set-point* da concentração de amônio e a rejeição de distúrbios de carga do esgoto afluente. Em seguida, o mesmo

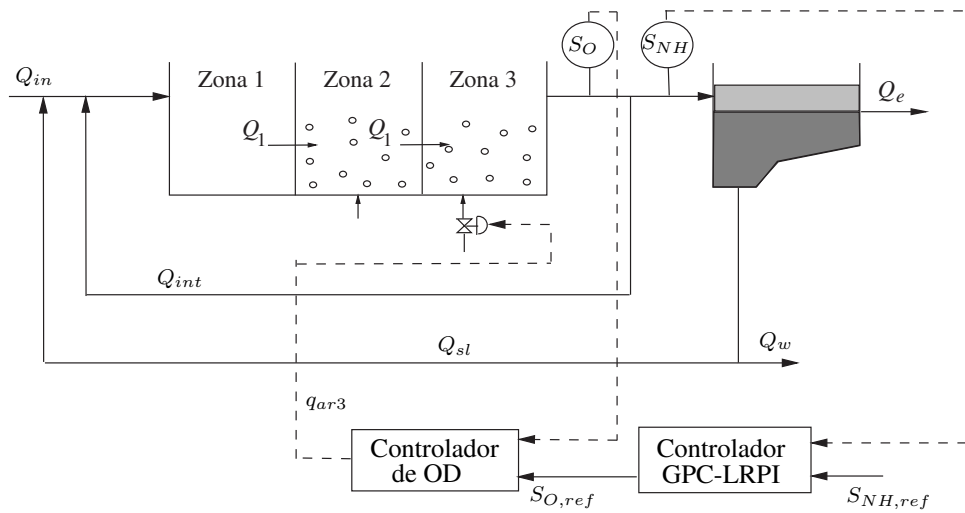


Fig. 6.1: Esquema de controle do set-point de OD para remoção de amônio.

é comparado com o desempenho obtido pelo controlador fixo do *set-point* de OD, mediante estudo de simulação usando o modelo *benchmark* ASWWTP-USP.

6.2.1 Identificação do modelo reduzido para a concentração de amônia

A identificação da dinâmica entre o *set-point* do OD e a concentração de amônio foi tratada em Lindberg (1997), que utiliza um modelo estimado com estrutura ARX de primeira ordem na entrada e na saída, para melhorar o ajuste de um controlador não linear. A diferença com o proposto neste trabalho é que o modelo de ordem reduzida é usado na síntese do controlador GPC-LRPI, modelo este obtido usando o algoritmo ortogonal de Gram-Schmidt Modificado (MGS) (Chen et al., 1989). A versão multivariável deste algoritmo, para diferentes ordens de linearidade, é resumida no Apêndice C. A facilidade do algoritmo para identificar modelos lineares de processos não lineares é estimulada usando sinais de excitação com característica de repetição inversa. Repete-se o procedimento para o experimento de identificação, mostrado na seção 3.8.1, onde foi usado o algoritmo com grau de não linearidade igual a 1 na identificação de um modelo linear parcimonioso para a concentração de oxigênio dissolvido no último reator biológico nitrificante.

Os sinais de excitação pseudo-aleatórios multi-níveis (MLPRS) são então projetados para satisfazer os requisitos de largura de banda relevante para controle, persistência da excitação e comprimento mínimo, baseados no conhecimento *a priori* do processo de remoção de amônio. Os parâmetros do sinal MLPRS são calculados mediante as diretrizes apresentadas no Capítulo 3. Como visto na subseção 3.4.2, estes sinais podem ser projetadas para exibir característica de repetição-inversa especialmente útil para identificar o comportamento linear de um sistema em presença de não linearidades, como mostrado nos estudos de aplicação de Srinivasan e Rengaswamy (1999).

Projeto do sinal de excitação

Para a especificação dos modos de operação do modelo, é conveniente verificar que o espaço do sinal do *set-point* de OD seja tal que excite a dinâmica da concentração de amônia, ao redor do valor desejado como *set-point* do controlador de amônio. O *set-point* para a concentração de amônio é adotado em 4 mg N/l, traduzindo a legislação vigente sobre lançamento de efluentes em corpos de água com objetivo de controle. A Resolução CONAMA No. 20 estabelece o limite de 5 mg N/l como padrão de lançamento para a concentração de amônio. Naturalmente que este é um valor de referência e que valores mais apropriados devem ser determinados estudando cada corpo de água.

O conhecimento *a priori* da dinâmica do processo é obtido através de sinais de excitação no *set-point* de OD, como sendo seqüências de degraus positivos e negativos de 40 h de duração. As amplitudes destes degraus são de 0,5 mg O₂/l e abrangem o intervalo entre 0,5 a 4,5 mg O₂/l. Isto gera um espaço de saída na concentração de amônio entre 3 a 7 mg N/l. A partir da análise da dinâmica do amônio, são determinadas a constante de tempo dominante mais curta, $\tau_{dom}^c = 3,65$ h e a constante de tempo dominante mais longa, $\tau_{dom}^l = 9,53$ h. Tendo em mente que a velocidade da dinâmica do sistema em malha fechada será mais rápida que a de malha aberta, é adotado um valor de $\alpha_s = 2$. Devido a este comportamento não linear da concentração de amônio, observam-se respostas com amplitudes diferentes. É assumido um fator relativo a um tempo de acomodação maior que o de 98%, que equivale a um valor de $\beta_s = 5$. Em conseqüência, a largura de banda do sinal de excitação resultante é $0,0210 \text{ rad/h} \leq w_s \leq 0,0548 \text{ rad/h}$. Devido ao comportamento não linear, o número de níveis do sinal de excitação deve ser $m \geq 3$. O tempo de comutação do sinal resulta $T_{sw} = 4$ h, originando um comprimento mínimo do sinal $N_s \geq 74,84$. O sinal pseudo-aleatório multi-nível (MLPRS) que satisfaz o requisito de baixa freqüência e de comprimento mínimo é aquele com 3 elementos no Campo de Galois, $GF(3)$, com registro de deslocamento de 4 estágios, $n_r = 4$ e com harmônicos múltiplos de 2 suprimidos. Isto corresponde a um número de harmônicos de potência uniforme igual a 18 dentro da largura de banda de interesse. A equação característica revisada a ser introduzida no Gerador MLPRS (ver Figura 3.6) é:

$$X = D^3 X + D^4 X \quad (6.1)$$

O tempo de duração de um ciclo do sinal MLPRS corresponde a $T_{cyc} = 320$ h. A taxa de amostragem inicial determinada é $T_{s,i} = 0,5$ h, como sugerido por Olsson e Newell (1999), aproximadamente $1/5 \cdot \tau_{dom}^c$ ou $1/10 \cdot \tau_{dom}^c$.

A resposta na concentração de amônio gerada pelo sinal de excitação do *set-point* de OD e o valor do OD obtido pelo sensor OD podem ser observados na Figura 6.2. O sensor é simulado como um sistema de primeira ordem (Suescon et al., 1998), com atraso na medição de 1,5 min e ruído branco aditivo na saída da medição de média nula e variância 0,01.

Na Figura 6.3(a) é mostrada a função de autocorrelação (FAC) do sinal de excitação MLPRS projetado. Esta FAC é característica de sinais com harmônicos múltiplos de 2 suprimidos, também denominados de sinais de repetição inversa. Na Figura 6.3(b) pode-se verificar a região de freqüência relevante para controle e o conteúdo de harmônicos para esta região, indicada pelo retângulo.

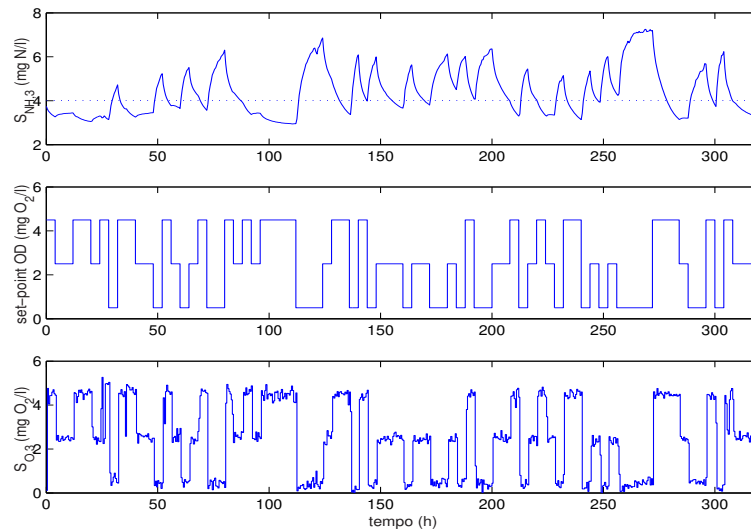
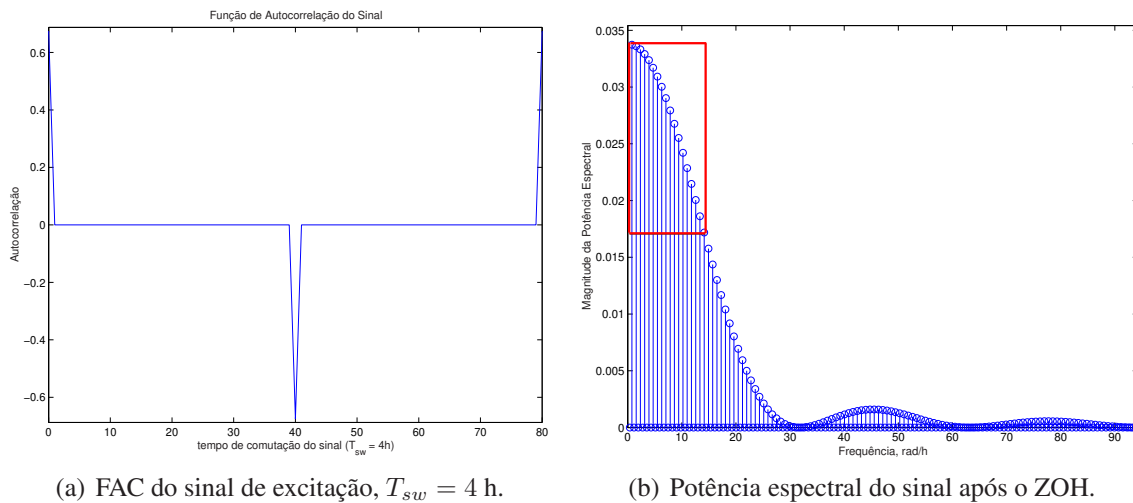


Fig. 6.2: Sinal de excitação do *set-point* de OD para gerar os dados de identificação da dinâmica entre o *set-point* de OD e a concentração de amônio. O gráfico superior corresponde à concentração de amônio, o intermediário é o *set-point* de OD e o inferior é o OD com ruído e atraso no sensor.



(a) FAC do sinal de excitação, $T_{sw} = 4$ h.

(b) Potência espectral do sinal após o ZOH.

Fig. 6.3: Característica de repetição inversa do sinal de excitação para o *set-point* de OD e a região de frequência relevante para controle deste sinal.

Seleção da estrutura e estimação dos parâmetros

Neste estágio do procedimento de identificação, é aplicado o algoritmo ortogonal de Gram-Schmidt Modificado (Chen et al., 1989), que associa a seleção da estrutura ótima com a estimação dos parâmetros do modelo NARX (ver Apêndice C). Os dados registrados do *set-point* de OD e da concentração de amônio são previamente escalonados e normalizados para estes serem usados pelo algoritmo de identificação. Para a aplicação do algoritmo não foi necessário alterar a taxa de amostragem inicial, pois depois de realizar a análise para a determinação da taxa de amostragem do algoritmo, esta resultou sendo similar à taxa de amostragem inicial, $T_s = 0,5$ h. Esta análise utiliza as funções de autocorrelação linear e não linear para a estimativa da taxa de amostragem do algoritmo e se necessário decimar os dados. Ver Aguirre (2000) para detalhes.

Como não se tem conhecimento *a priori* dos atrasos na estrutura do modelo NARX, são realizados ensaios para diferentes valores de atrasos máximos N_y, N_u, d . Para a seleção da estrutura do modelo parcimonioso são usados os critérios de Akaike (AIC4), de Bayes (BIC) e do Erro Máximo, segundo procedimento descrito no Capítulo 3 (resultados não incluídos aqui).

O modelo identificado selecionado da dinâmica do *set-point* de OD e a concentração de amônio corresponde a:

$$S_{NH,3}(t) - 1,610S_{NH,3}(t-1) + 0,630S_{NH,3}(t-2) = 0,012 - 0,004S_{O,sp}(t-4) + 0,012S_{O,sp}(t-7) \quad (6.2)$$

que foi determinado usando o algoritmo MGS para valores de atrasos máximos na saída, $N_y = 4$, na entrada, $N_u = 4$ e de tempo morto, $d = 3$.

Validação do modelo

Na Figura 6.4 é mostrada a validação do modelo identificado, com estrutura linear selecionada, com a saída da planta, onde pode-se notar a boa aderência das predições da concentração de amônio do modelo linear identificado. Note também que o *set-point* da concentração de amônio está contido dentro do espaço de saída desta variável, ressaltado pela linha no valor de 4 mg N/l. Naturalmente, a Figura 6.4 reflete o bom desempenho do modelo em reproduzir os dados para a identificação. Outro teste de validação é efetuado com um novo conjunto de dados independentes dos dados de identificação. São simulados distúrbios na carga do afluente, isto é, na concentração de substrato e na concentração de amônio. Na Figura 6.5 é verificada a boa aderência da predição para a concentração de amônio na zona aeróbia, para este novo conjunto de dados. Note que os distúrbios são mostrados normalizados, para facilitar a visualização.

6.2.2 Resultados da simulação em malha fechada

Na simulação em malha fechada do controlador de *set-point* de OD, é usado o modelo *benchmark* ASWWTP-USP para simular a planta de lodos ativados com pré-desnitrificação, para as condições operacionais seguintes. As vazões em estado estacionário são: afluente $Q_{in} = 4,17$ m³/h, recirculação interna de nitrato $Q_{int} = 2Q_{in}$, retorno de lodo $Q_{sl} = 0,5Q_{in}$ e descarte de lodo $Q_w = 25,8$ l/h. O valor total de carga orgânica carbonácea é 284 mg DQO/l e o nitrogênio total é 45,88 mg N/l.

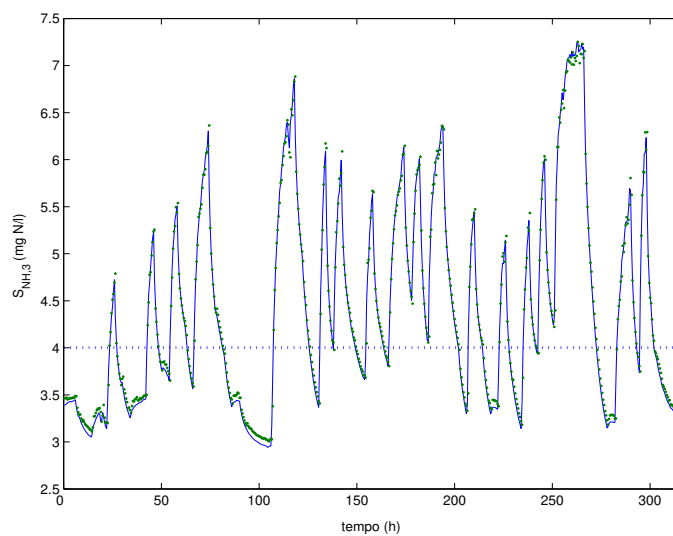


Fig. 6.4: Identificação do modelo da concentração de amônio: saída da planta (—) e saída do modelo identificado (··). *Set-point* da concentração de amônio em 4 mg N/l.

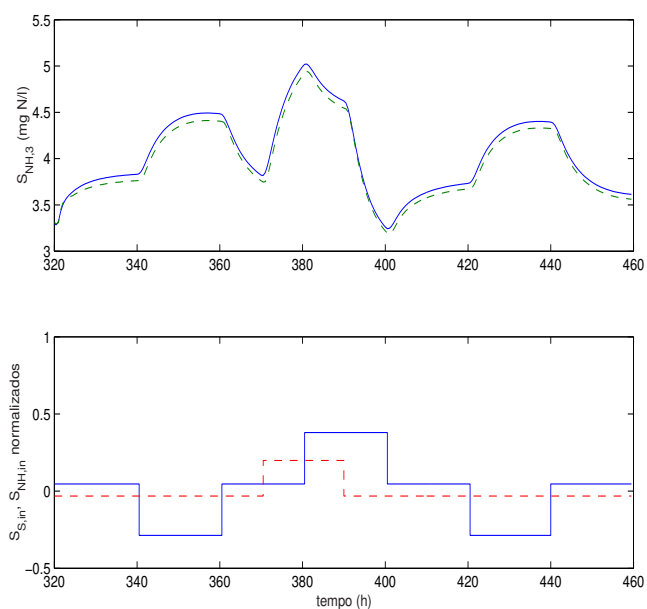


Fig. 6.5: Validação entre a saída da planta (—) e a predição do modelo (- -) frente a distúrbios no afluente, nas concentrações de substrato e amônio.

Os volumes são: zona anóxica 13 m^3 , primeira e segunda zonas aeróbias, 18 e 20 m^3 , respectivamente. As concentrações do esgoto afluyente para as variáveis que servirão de distúrbio externo são: substrato rapidamente biodegradável $S_{S,in} = 64 \text{ mg DQO/l}$ e amônio $S_{NH,in} = 12,5 \text{ mg N/l}$.

O modelo identificado da dinâmica entre o *set-point* de OD e a concentração de amônio mostrado na relação 6.2 é usado para projetar o controlador GPC-LRPI. O modelo reduzido tem duas principais contribuições no processo de síntese do controlador:

- Provê um modelo de estrutura ótima que é usado na estratégia de estimação LRPI, que permanece fixa ao longo do horizonte de predição do controlador.
- Provê os parâmetros iniciais para o algoritmo de estimação LRPI.

Parâmetros do controlador

Como visto na seção 5.1.1, o critério quadrático ótimo parcialmente restrito do GPC possui uma série de parâmetros que devem de ser selecionados para obter o desempenho em malha fechada desejado. Estes parâmetros incluem:

- O horizonte de predição mínimo, N_1 e máximo, N_2 .
- Horizonte de predição, $N_p = N_2 - N_1 + 1$.
- Horizonte de controle, N_c .
- Ponderação entre o erro de rastreamento da predição e o desvio de controle, λ_u .
- Taxa de amostragem do controlador, $T_{s,c}$.

Adicionalmente, para o ajuste do controlador GPC-LRPI, o polinômio ou filtro observador T , é adicionado e este conjunto de parâmetros. Diretrizes para a seleção destes parâmetros no ajuste do controlador preditivo DMC (*Dynamic Matrix Control*) são encontradas em Cutler e Ramaker (1980) e Seborg et al. (1989). No presente trabalho usa-se também como referência a estratégia de ajuste recomendada por Shridhar e Cooper (1997), para o DMC SISO sem restrições, aplicável a processos estáveis em malha aberta e que pode ser extrapolado ao ajuste do controlador GPC-LRPI. Isto é, baseado na semelhança entre o GPC e o DMC quando o polinômio que pondera a trajetória da saída predita é igual à unidade (McIntosh et al., 1991). Para este fim, a informação obtida da dinâmica do processo no projeto do sinal de excitação é utilizada para a seleção dos parâmetros do controlador.

A taxa de amostragem recomendada para o controlador deve ser $T_{s,c} \geq 0,1\tau_{dom}$ e $T_{s,c} \leq d_p$. Em base ao tempo morto observado do processo ($d_p = 0,5 \text{ h}$) e às constantes de tempo dominantes ($\tau_{dom}^l = 9,53 \text{ h}$ e $\tau_{dom}^c = 3,65 \text{ h}$), obtidos desde os ensaios dinâmicos, seleciona-se $T_{s,c} = 0,5 \text{ h}$. Esta escolha é também baseada no valor da taxa de amostragem usada para o experimento de identificação, $T_{s,i} = 0,5 \text{ h}$.

Sabe-se que aumentar o horizonte de predição, N_p resulta em ações de controle conservadoras o que tem efeito estabilizante no sistema de malha fechada. Basicamente, este parâmetro deve ser de 3 a 5 vezes a constante de tempo dominante do processo ou deve ser próximo do tempo de acomodação em amostras. Com isto se garante que seja predito o estado estacionário da saída. Devido à dinâmica

Tab. 6.1: Parâmetros do controlador GPC-LRPI.

N_1	N_p	N_c	λ_u	$T_{s,c}$
1	15	2	10	0,5 h

do tipo *stiff* exibida pelo processo, recomenda-se que o controlador seja capaz de reagir as dinâmicas lentas e rápidas. Portanto, garantindo a dinâmica mais rápida, o horizonte de predição de referência deve ser aproximadamente $3\tau_{dom}^c/T_{s,c}$. Levando em conta este valor e depois de diversos ensaios de tentativa e erro, selecionou-se $N_p = 15$, considerando $N_1 = 1$.

Como visto na seção 5.1.1, a principal utilização do horizonte de controle, N_c , é reduzir o esforço computacional mediante sua inclusão como restrição no critério de controle quadrático GPC da equação (5.6). No entanto, é demonstrado mediante simulações que o papel deste parâmetro vai além da simplificação computacional. Valores pequenos resultam em respostas conservadoras, enquanto valores grandes de N_c proporcionariam uma melhor ação de rastreamento (Wellstead e Zarrop, 1991). Em geral, recomenda-se um horizonte de controle relativo ao intervalo $1 \leq N_c \leq 6$ (Seborg et al., 1989; Shridhar e Cooper, 1997). Nos ensaios, foi verificado que o valor de $N_c = 2$ proporciona uma resposta "suave" com sobre-sinais máximos pequenos.

O fator de ponderação λ_u serve como penalização para excessivas excursões da ação de controle. Shridhar e Cooper (1997) o calculam analiticamente como:

$$\lambda_u = f K_p^2 \quad (6.3)$$

onde K_p é o ganho de estado estacionário para uma determinada excitação e f é definida como:

$$f = \begin{cases} 0 & N_c = 1 \\ \frac{N_c}{500} \left(\frac{3,5\tau_{dom}}{T_{s,c}} + 2 - \frac{(N_c-1)}{2} \right) & N_c > 1 \end{cases} \quad (6.4)$$

Considerando o maior ganho estacionário e a constante de tempo dominante mais longa pode ser obtido o maior valor previsto do fator de ponderação, $\lambda_u \approx 6, 8$. Durante as simulações se verificou que valores menores provocavam respostas com grandes sobre-sinais. Com base às observações dos ensaios em malha fechada, foi selecionado $\lambda_u = 10$, valor utilizado na avaliação de desempenho do GPC-LRPI tanto para o rastreamento do *set-point* quanto para rejeição de distúrbios externos.

O resumo dos parâmetros de ajuste do controlador mestre GPC-LRPI para o controle do *set-point* de oxigênio dissolvido para remoção de amônio é mostrado na Tabela 6.2.

São utilizados $D(q^{-1}) = \Delta = 1 - q^{-1}$ e $T = (1 - 1.6q^{-1} + 0.8q^{-1})$ para a pré-filtragem LRPI relevante para controle. O estimador recursivo utiliza o fator de esquecimento exponencial $\lambda = 0, 98$ e a magnitude da matriz de covariância igual a 100. O estimador é inicializado com os valores dos parâmetros do modelo identificado da relação (6.2), como indicado anteriormente.

Como explicado na sub-seção 6.2.1, o objetivo é manter o valor da concentração de amônio no valor de 4 mg N/l. Na Figura 6.6 é mostrada a concentração de amônio na última zona aeróbia para mudanças no *set-point*. Inicialmente o *set-point* do controlador mestre GPC-LRPI é fixado

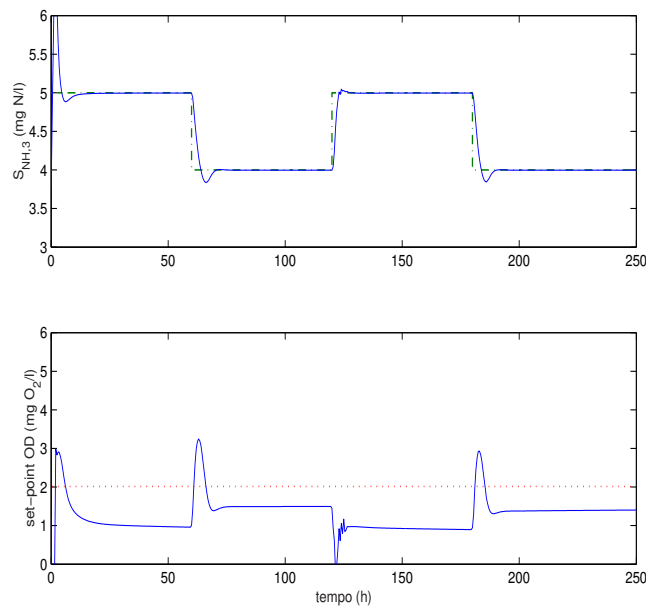


Fig. 6.6: Rastreamento das mudanças do *set-point* da concentração de amônio sem ruído aditivo.

em 5 mg N/l, para depois diminuir seu valor até o valor desejado de 4 mg N/l. Esta seqüência é repetida a cada 50 horas do tempo de simulação. O desempenho do controlador para o rastreamento da referência mostra ser eficaz. Pode-se notar também que a variável manipulada ou *set-point* variável de OD apresenta em média valores menores que 2 mg O_2 /l. Isto significa uma economia nos custos de aeração quando a estratégia de controle supervisorio de OD é utilizada.

O desempenho da estratégia de controle proposta é também testada para o caso em que ruído branco é adicionado na saída do sensor de OD, com média nula e variância 0,01, como mostrado na Figura 6.7. Observa-se novamente um bom rastreamento da referência da concentração de amônio sem que seja necessário reajustar os parâmetros do controlador GPC-LRPI, que adapta constantemente os parâmetros do filtro LRPI conforme aumenta o horizonte de predição do controlador (ver seção 4.4.2).

Distúrbios nas concentrações do afluente

Para mostrar a robustez do controlador na rejeição de distúrbios externos, é simulado o cenário de mudanças na concentração do substrato rapidamente biodegradável e na concentração do amônio, ambas no esgoto afluente. Os valores destes distúrbios nas concentrações para a simulação são mostrados na parte inferior da Figura 6.8. Estas concentrações possuem a mesma forma e amplitudes que as utilizadas na validação cruzada do experimento de identificação.

Para efeito de comparação de desempenho, é simulado também o desempenho dos controladores PI do simulador *benchmark* ASWWTP-USP, com *set-point* em 2 mg O_2 /l. Os valores dos parâmetros de ajuste destes controladores, para as duas zonas aeróbias, são $K_P = 72,8464$ e $K_I = 14,57$. Na Figura 6.8 é mostrada a comparação entre o desempenho da estratégia de *set-point* fixo usando os controladores PI e o desempenho utilizando o controlador GPC-LRPI como controlador mestre com

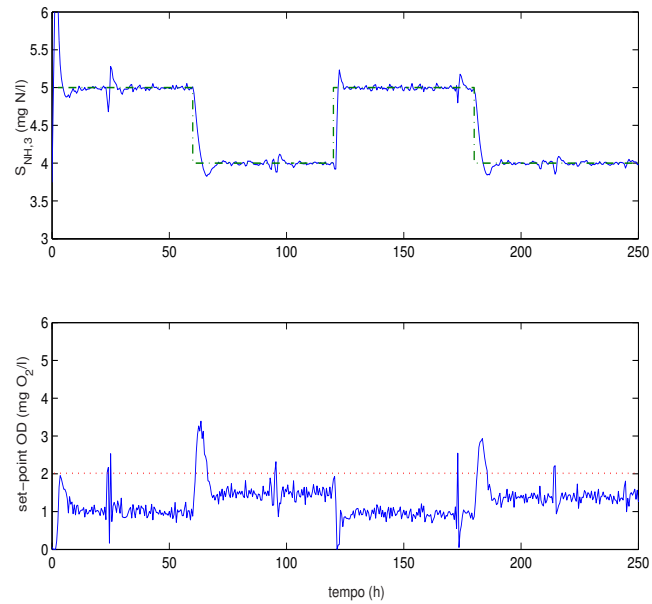


Fig. 6.7: Rastreamento de mudanças do *set-point* da concentração de amônio com ruído aditivo no sensor.

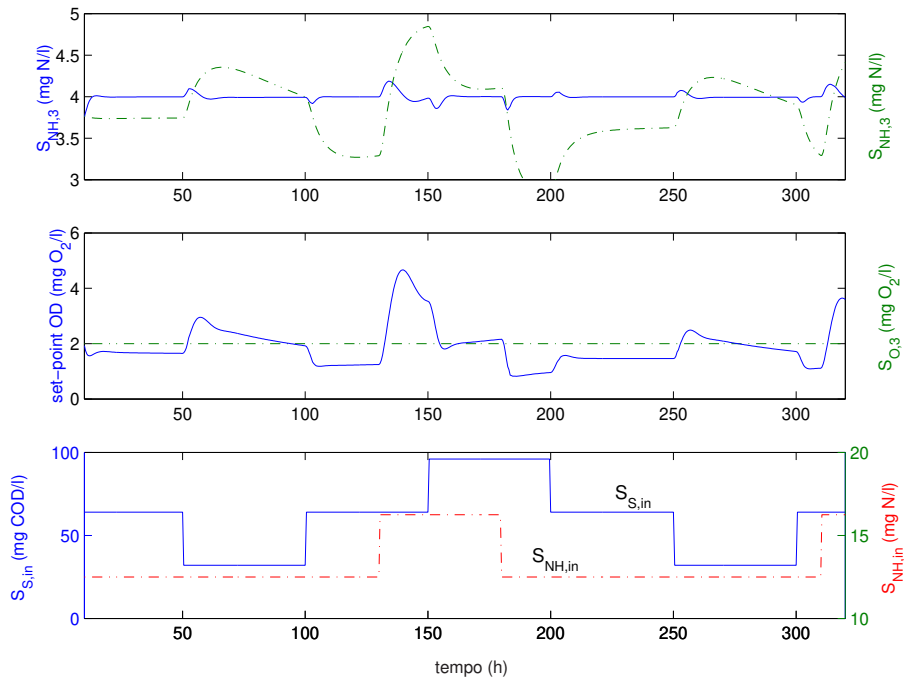


Fig. 6.8: Controle do *set-point* de OD. A linha contínua corresponde ao *set-point* controlado enquanto a linha tracejada corresponde ao *set-point* fixo de 2 mg O₂/l. Os distúrbios nas concentrações de amônio (- -) e de substrato (-) também são mostradas.

set-point variável. Pode-se notar que a estratégia de *set-point* variável é nitidamente superior, tanto na velocidade de resposta como na capacidade para rejeitar distúrbios externos na carga do afluente.

6.3 Controle da desnitrificação

O processo de desnitrificação é principalmente influenciado pela vazão de recirculação interna, rica em nitrato e por uma dosagem externa de carbono (se necessário). Neste processo, a bactéria desnitrificante precisa de quantidades suficientes de carbono rapidamente biodegradável para desenvolver suas atividades metabólicas. Portanto, a relação carbono/nitrogênio (C/N) deve ser tal que permita que o processo de desnitrificação se desenvolva adequadamente. A adição de uma fonte externa de carbono é uma variável de controle eficaz, mas é dispendiosa. Esta deve ser a última opção, devido ao custo adicional e também porque pode incrementar a produção de lodo e os custos de aeração no reator aeróbio. Como foi usada a condição operacional inicial da caracterização do esgoto afluente do *benchmark* ASWWT-USP, como tendo suficiente carga de carbono orgânico (284 mg DQO/l), será adotada a vazão de recirculação interna como a variável manipulada para o controle da desnitrificação.

Na verdade, algoritmos para o controle *on-line* da concentração de nitrato, através da vazão de recirculação interna, têm sido estudados por diferentes pesquisadores. Londong (1992) observou que o grande problema com esta forma de controle era a falta de precisão para pequenos valores de concentração medidos pelo sensor de nitrato. Isto não é mais problema devido às contínuas melhorias na confiabilidade, precisão e facilidade de manutenção dos novos sensores de amônio e nitrato (Yuan et al., 1999; Olsson, 2002b). Sotomayor et al. (2000) propõem um controlador por modelo genérico (GMC), usando um modelo não linear simplificado da dinâmica do nitrato, cujos parâmetros são estimados recursivamente. Devido à natureza variante no tempo do processo de lodos ativados, é desejável que a técnica de controle a projetar adapte os parâmetros do modelo perante possíveis mudanças na região de operação. No entanto, esta abordagem requer sensores adicionais para medição dos distúrbios da configuração *feedforward*. A idéia da parcela *feedforward* é atuar no processo quando algum distúrbio medido aparecer, antes que provoque uma mudança maior na variável controlada. Por outro lado, têm sido propostos algoritmos de controle lineares do tipo PI, usando a hipótese de desacoplamento por escalas de tempo, mediante combinação de configurações *feedback* (Suescon et al., 2001; Vrečko et al., 2002; Ingildsen, 2002) e *feedback-feedforward* (Singman, 1999; Rehns-tröm, 2000; Ingildsen, 2002; Vrečko et al., 2003). Estas abordagens mostram um bom desempenho na rejeição de distúrbios externos, mas requerem uma maior instrumentação para poder atualizar o modelo do termo *feedforward*. Nas abordagens citadas não são feitas comparações de desempenho com estratégias avançadas de controle.

Vale ressaltar que o objetivo aqui é implementar uma estratégia de controle que seja robusta perante distúrbios externos e que ao mesmo tempo utilize o menor número possível de sensores. No presente contexto, o termo robusto é usado para denotar a insensibilidade ou invariância do desempenho do controlador com respeito a incertezas (distúrbios, erros de modelagem devido à redução do modelo e ruídos de medição).

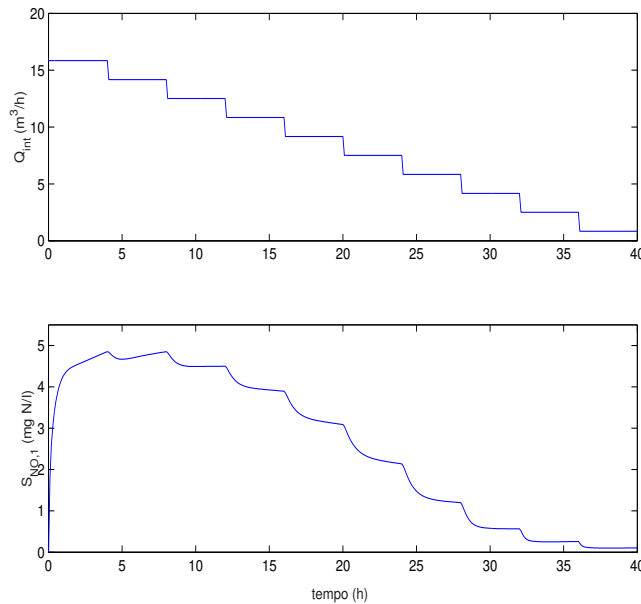


Fig. 6.9: Seqüência decrescente de degraus para determinação do espaço entrada-saída.

6.3.1 Identificação do modelo de ordem reduzida da concentração de nitrato

A síntese do modelo de ordem reduzida é obtida da mesma forma que na seção 6.2.1, mediante o algoritmo ortogonal MGS (ver Apêndice C). Os sinais de excitação que geram os dados de identificação da dinâmica entre a vazão de recirculação interna e a concentração de nitrato, são projetados tal que satisfaçam os requisitos de largura de banda relevante para controle, persistência da excitação e comprimento mínimo, baseados no conhecimento *a priori* do processo de desnitrificação. São usadas também as diretrizes apresentadas no Capítulo 3.

Projeto dos sinais de excitação

As constantes de tempo dominantes são determinadas usando seqüências positivas e negativas de degraus na vazão de recirculação interna, tal como visto na Figura 6.9. Note o comportamento não linear do processo quando degraus da mesma amplitude são injetados. É importante remarcar que o espaço entrada-saída excitado deve conter o *set-point* do controlador a projetar. Seguindo o mesmo procedimento praticado no Capítulo 3, são obtidas a constante de tempo mais longa, $\tau_{dom}^l = 1,12$ h e a mais curta, $\tau_{dom}^c = 0,23$ h.

A largura de banda do sinal de excitação é calculada como sendo $0,1784 \text{ rad/h} \leq w_s \leq 13,06 \text{ rad/h}$, escolhendo-se $\alpha_s = 3$ e $\beta_s = 5$, com o objetivo de satisfazer o requisito de conteúdo de alta e baixa freqüência do sinal. O tempo de comutação relacionado é $T_{sw} = 0,2$ h. Portanto, o comprimento mínimo do sinal a projetar deve satisfazer $N_s \geq 176,14$. Após realizar a análise do requisito de baixa freqüência, juntamente com a consideração do maior conteúdo de harmônicos de potência uniforme na largura de faixa de interesse, é selecionado o sinal MLPRS com $q = 3$ elementos do Campo de Galois e $n_r = 5$ registros de deslocamento. O sinal resultante possui a característica de repetição inversa, pois o maior número de harmônicos presentes em w_s com potência uniforme foram obtidos

para o valor de $\gamma = 1/2$ (supressão de harmônicos múltiplos de 2). O tempo de duração de um ciclo do sinal MLPRS resulta ser 48,4 h que também é gerado mediante o Gerador MLPRS implementado no Capítulo 3.

Para efeitos de geração do sinal, é escolhido o seguinte polinomial primitivo usando o pacote Galois[©]:

$$f(x) = 1 + 2x^4 + x^5 \quad (6.5)$$

cuja equação característica revisada resulta:

$$X = D^4X + 2D^5X \quad (6.6)$$

Seleção da estrutura e estimação dos parâmetros

Na Figura 6.10 é mostrado o registro dos dados ‘medidos’ da concentração de nitrato e da vazão de recirculação interna para o sinal MLPRS com 3 níveis e harmônicos múltiplos de 2 suprimidos. Previamente à aplicação do algoritmo MGS, os sinais são adequadamente escalados e normalizados. Procedimentos de detecção de *outliers* e tendências são recomendados no pré-processamento de dados registrados em plantas de tratamento de esgotos (Olsson e Newell, 1999), mas assume-se a inexistência de irregularidades nas medições. São também aplicadas técnicas de remoção de tendências (se necessário) no modelo polinomial identificado. Para o experimento de identificação, é considerada a dinâmica do sensor de nitrato com atraso de medição de 5 minutos e ruído branco adicionado na saída, com média nula e variância 0,01, para simular ruído de medição de alta frequência. É considerada uma taxa de amostragem inicial de $1/10 \cdot \tau_{dom}^l$ (0,04 h). Mas para a aplicação do algoritmo MGS é usada uma taxa de amostragem de 0,1 h, como resultado da decimação dos dados conforme procedimento de determinação da taxa de amostragem em Aguirre (2000), que utiliza as funções de autocorrelação linear e não linear dos sinais normalizados.

O algoritmo ortogonal MGS seleciona uma estrutura ótima e estima os parâmetros de modelos não lineares autoregressivos com entrada externa (modelos NARX). A vantagem de usar este algoritmo de identificação com sinais de excitação MLPRS relevantes para controle, é que desta forma pode-se obter modelos parcimoniosos lineares para controle. O modelo identificado para a concentração de nitrato, com atrasos máximos na entrada e a saída iguais a 3 e com um tempo morto máximo de 2 é:

$$S_{NO,1}(t) - 1,468S_{NO,1}(t-1) + 0,187S_{NO,1}(t-2) + 0,295S_{NO,1}(t-3) = \\ 0,003 - 0,038Q_{int}(t-3) - 0,053Q_{int}(t-4) \quad (6.7)$$

O procedimento de validação do modelo identificado com os dados medidos é mostrado na parte superior da Figura 6.11. Na parte inferior desta figura, pode ser visto o ajuste do sinal de entrada na região de frequência relevante para controle do processo de desnitrificação.

Para verificar a qualidade de predição do modelo identificado, é realizado outro procedimento de validação, com um conjunto independente de dados medidos no intervalo de 48,4 a 64 horas. É utilizada a predição da concentração de nitrato. Neste caso foi preciso utilizar o procedimento de detecção de tendências mediante o algoritmo de Mann-Kendall e sua respectiva remoção, como explicado em Olsson e Newell (1999). O resultado das predições pode ser visto na Figura 6.12, onde se observa que embora o processo mostre um comportamento não linear, o modelo reduzido linear

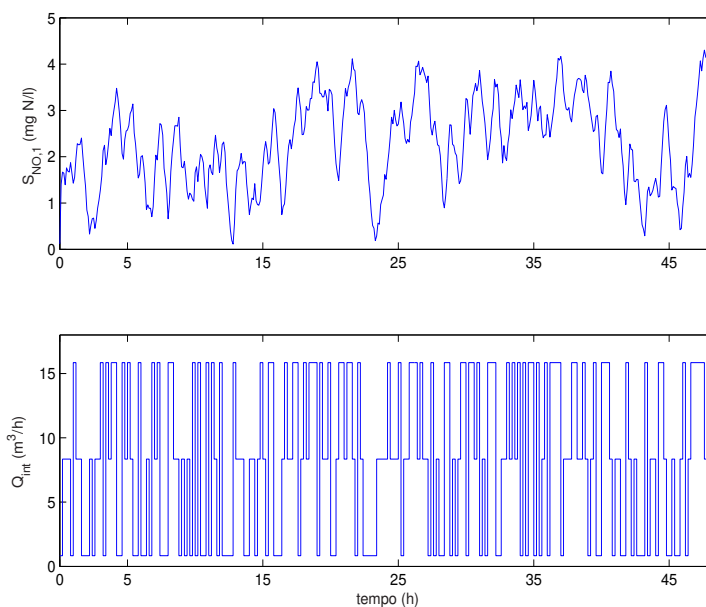


Fig. 6.10: Sinal de saída (concentração de nitrato) e sinal de entrada (vazão de recirculação interna) para o experimento de identificação.

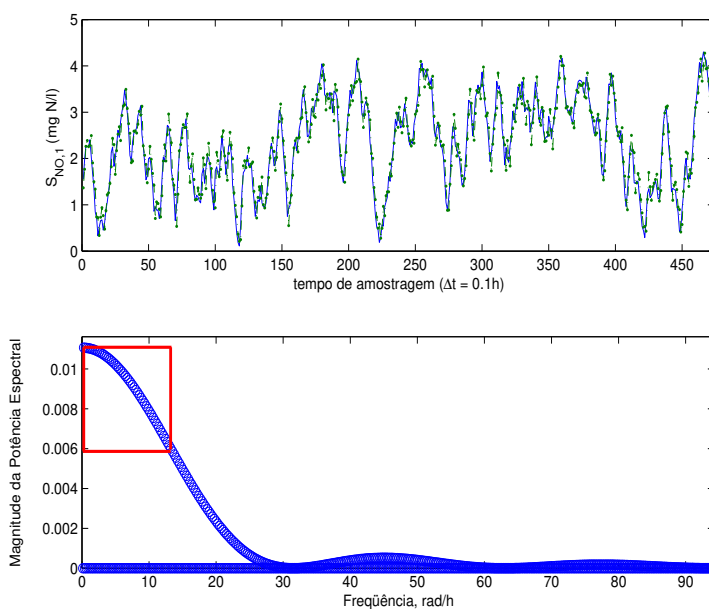


Fig. 6.11: Validação do modelo identificado e potência espectral indicando a região relevante para controle depois do ZOH.

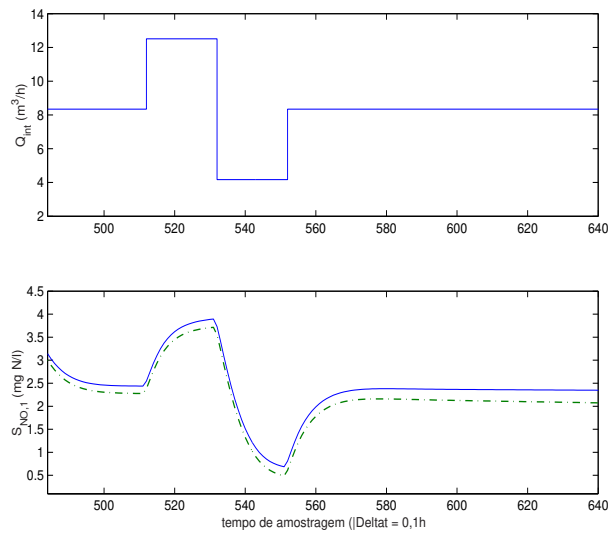


Fig. 6.12: Validação do modelo identificado com o conjunto independente de dados após o processo de remoção de tendências.

consegue reproduzir com eficácia a concentração de nitrato na zona anóxica.

6.3.2 Controle da concentração de nitrato

O controlador proposto é avaliado em um estudo de simulação onde novamente o modelo *benchmark* ASWWTP-USP é usado como planta. Neste procedimento em malha fechada é considerada também a dinâmica do sensor de nitrato, da mesma forma que no experimento de identificação. Os valores iniciais da vazão do esgoto bruto, das vazões de recirculação e das concentrações são os mesmos valores adotados no *benchmark* em regime estacionário.

Parâmetros do controlador

A taxa de amostragem, embora não seja considerada como parâmetro de ajuste no controlador GPC padrão, mostra ter influencia no desempenho de malha fechada do sistema controlado. Portanto, é importante verificar a sensibilidade da resposta para valores diferentes de $T_{s,c}$. Considerando as constantes de tempo dominantes obtidos da etapa de projeto dos sinais de excitação ($\tau_{dom}^l = 1,12$ h e $\tau_{dom}^c = 0,23$ h), a taxa de amostragem recomendada do controlador GPC-LRPI deve ser $T_{s,c} \geq 1,38$ min (segundo a diretriz vista na sub-seção 6.2.2). Porém, como resultado de testes de desempenho de malha fechada, é selecionado o valor $T_{s,c} = 3$ min.

Foi visto que o horizonte de predição de referência deve ser aproximadamente $3\tau_{dom}^c/T_{s,c}$, conforme a garantir a resposta do controlador perante a dinâmicas rápidas. Este valor de referência resulta em um horizonte igual a 20. No entanto, valores menores mostraram levar ao sistema à estabilidade. Após vários testes de tentativa e erro, selecionou-se o horizonte de predição $N_p = 8$, com horizonte de predição mínimo igual a 1.

Recomenda-se um horizonte de controle relativo ao intervalo $1 \leq N_c \leq 6$ (Shridhar e Cooper,

Tab. 6.2: Parâmetros do controlador GPC-LRPI.

N_1	N_p	N_c	λ_u	$T_{s,c}$
1	8	4	8	3 min

1997), sendo que valores maiores aumentam o esforço computacional sem trazer maiores melhorias no desempenho de malha fechada. Um valor inicial pode ser escolhido N_c próximo a 60% do estado estacionário de malha aberta ($N_c \approx t_{60}/T_{s,c}$), como sugerido em Seborg et al. (1989) no caso de ajuste do DMC. Considerando a constante de tempo mais curta e a taxa de amostragem do controlador é obtido o valor de 4,6. Portanto, extrapolando esta recomendação para o ajuste do controlador GPC-LRPI é selecionado o valor de $N_c = 4$, por proporcionar uma resposta "suave" da ação de controle com sobre-sinais máximos pequenos perante distúrbios externos introduzidos nas concentrações de substrato e amônio.

O fator de ponderação λ_u é obtido usando como valor referencial $\lambda_u = fK_p^2$. O ganho de estado estacionário K_p resulta ser entre 0,19 e 1,45, com base aos ensaios dinâmicos realizados na etapa de projeto do sinal de excitação. Usando a equação (6.4) junto com a constante de tempo dominante mais longa ($\tau_{dom}^l = 1,2$ h) é obtido λ_u igual a 1,32. Com base às simulações se verificou que valores menores a 4 provocavam excursões indesejadas na ação de controle (Q_{int}). Desde ponto de vista prático, oscilações abruptas do sinal enviado ao atuador devem ser evitadas. Portanto, é selecionado $\lambda_u = 8$, que consegue um comportamento mais "suave" na ação de controle devido a que é introduzida desta forma uma maior penalização na ação de controle. Valores maiores de λ_u tornam a resposta do processo mais lenta, prejudicando o desempenho no rastreamento do set-point. O valor escolhido é então um compromisso entre o desempenho para a rejeição de distúrbios externos e para o rastreamento do set-point, sem que isto envolva re-ajustar o controlador.

O resumo dos parâmetros do controlador GPC-LRPI para o controle de nitrato na zona anóxica do bioreator é mostrado na Tabela 6.2. O pré-filtro Δ/T da estratégia LRPI utiliza os polinômios $\Delta = 1 - q^{-1}$ e $T = (1 - 0.8q^{-1})$. O estimador recursivo de mínimos quadrados utiliza um fator de esquecimento exponencial de 0,98, com fator inicial para a matriz de covariância igual a 1000. Os valores iniciais dos parâmetros do estimador correspondem aos obtidos no procedimento de identificação em malha aberta, correspondentes à relação (6.7).

Escolha do *set-point*

O valor da concentração de nitrato no final da zona anóxica é um bom indicador da finalização da reação de desnitrificação, com vista à remoção biológica de nitrogênio (Olsson e Newell, 1999). Este valor tem influência também no valor da concentração de nitrogênio total do efluente, junto com a reação de nitrificação, que visa à remoção de amônia no bioreator aeróbio.

No entanto, é frequentemente argumentado que a vazão de recirculação interna possui uma baixa autoridade de controle. O termo autoridade de controle é usado para denotar a habilidade da variável manipulada, em influenciar o sistema a ser controlado, para rejeitar distúrbios e manter o valor desejado da saída (*set-point*). No entanto, em Ingildsen (2002) é demonstrado que a escolha do *set-point*

de nitrato entre os valores de 1 a 3 mg N/l, leva próximo do máximo a remoção de nitrogênio na zona anóxica, quando usada a vazão de recirculação interna como variável manipulada para valores do *set-point* de OD entre 1 a 3 mg O₂/l nas zonas aeróbias. Conforme Rehnström (2000), o *set-point* de nitrato na zona anóxica pode ser duas vezes a constante de média-saturação para o nitrato, cujo valor para o *benchmark* ASWTTP-USP corresponde a $K_{NO_3^-} = 0,5 \text{ g NO}_3\text{-N/m}^3$. É importante notar que a constante de média-saturação determina o nível mais baixo de nitrato em que o processo de desnitrificação permanecerá ativo. Portanto, neste estudo, o *set-point* do controlador GPC-LRPI é fixado para uma concentração de nitrato igual a 1 mg N/l no final da zona anóxica. Isto também satisfaz a condição operacional de manter a zona anóxica "anóxica" (Yuan et al., 2002). As concentrações de OD nas duas zonas aeróbias são controladas em 2 mg O₂/l pelos controladores PI de OD do *benchmark*.

Simulações em malha fechada

O controlador PI de nitrato do *benchmark* ASWWTP-USP tem sido usado como referência na comparação do desempenho do controlador GPC-LRPI, para rejeitar distúrbios e rastrear o *set-point* no controle da concentração de nitrato. Dado que a dinâmica do sensor está sendo considerada, este controlador precisou ser reajustado. Os novos parâmetros de ajuste são $K_P = 0,2361$ e $K_I = 0,3543$.

Com a finalidade de avaliar o desempenho dos controladores PI e GPC-LRPI foram considerados quatro cenários de distúrbios:

1. Adição de um distúrbio externo na vazão de afluente, Q_{in} , para amplitudes variando desde 50% e -50% de seu valor nominal.
2. Adição de distúrbios externos nas concentrações nominais do substrato rapidamente biodegradável, S_S , e de amônio, S_{NH} , no afluente de esgoto bruto. As mudanças são de -50% e 100% respectivamente.
3. Adição de distúrbio externo na vazão da fonte externa de carbono, para mudanças de 0, 2 e 4 l/h. A fonte externa de carbono é composta de metanol puro com uma solução a 33% a uma concentração de 80.000 mg COD/l.
4. Capacidade de rastreamento do *set-point*, introduzindo mudanças na referência da concentração de nitrato na zona anóxica entre 1 a 2 mg N/l.

Na Figura 6.13 é mostrado o resultado para o cenário 1, onde pode ser verificado o desempenho superior do controlador GPC-LRPI para atenuar os distúrbios da vazão de afluente. Se a concentração de nitrato é abaixo do *set-point*, então mais nitrato é convenientemente reciclado à zona anóxica onde esta é consumida. Note que as amplitudes da concentração de nitrato controlada são menores que $\pm 10\%$ do valor da referência, mostrando também uma maior rapidez na rejeição dos distúrbios.

O cenário 2 simula uma redução na concentração de substrato e um incremento na carga de amônio, ambas mudanças na carga do afluente. Mediante comparação pode ser notado que o controlador GPC-LRPI apresenta também uma rejeição ao distúrbio mais rápida com amplitudes menores que 5% do valor da referência de nitrato, como mostrado na Figura 6.14.

Um distúrbio externo na DQO na zona anóxica pode ser simulado através do incremento na vazão da fonte externa de carbono, que corresponde ao cenário 3. Na Figura 6.15 é avaliado o controlador

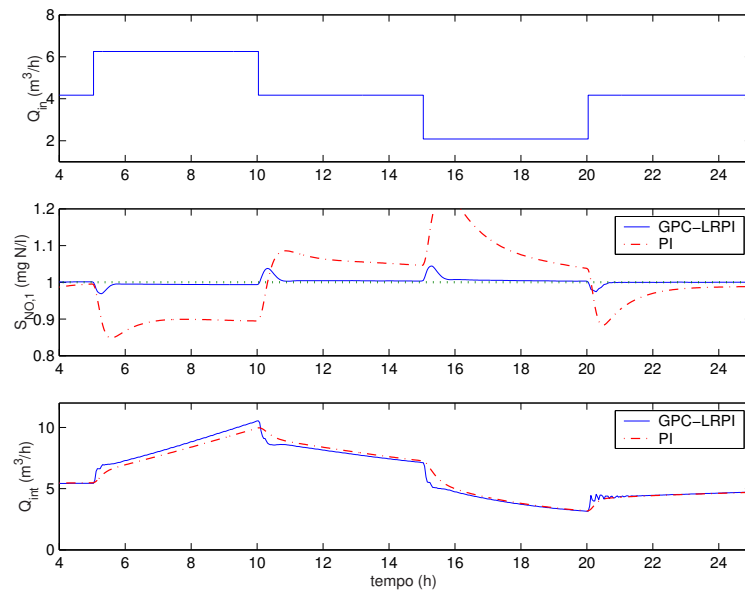


Fig. 6.13: Desempenho dos controladores GPC-LRPI e PI para rejeição de distúrbios na vazão de afluente do esgoto bruto.

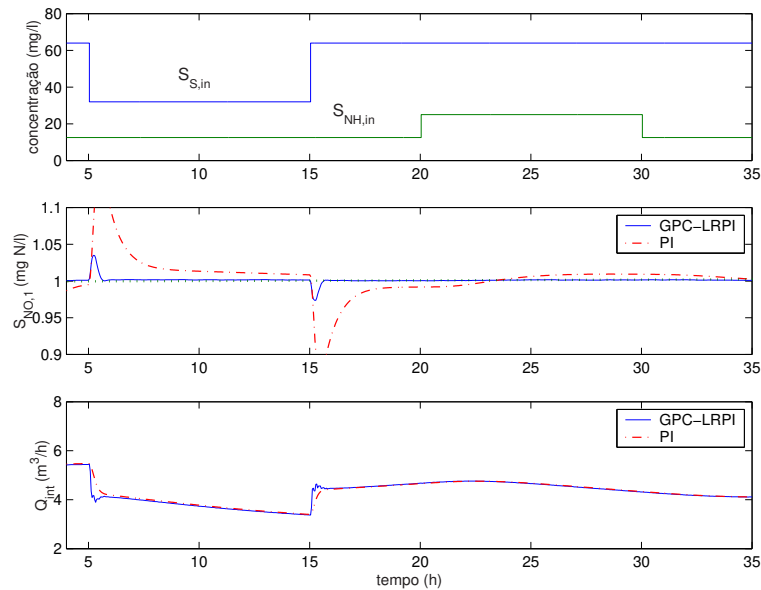


Fig. 6.14: Desempenho dos controladores GPC-LRPI e PI para rejeição de distúrbios nas concentrações de substrato e de amônio no afluente.

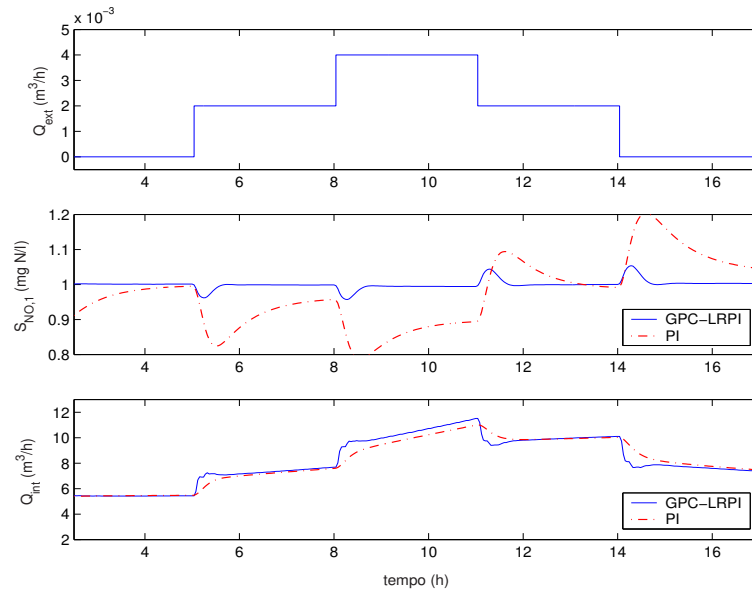


Fig. 6.15: Desempenho dos controladores GPC-LRPI e PI para rejeição de distúrbios na vazão da fonte externa de carbono.

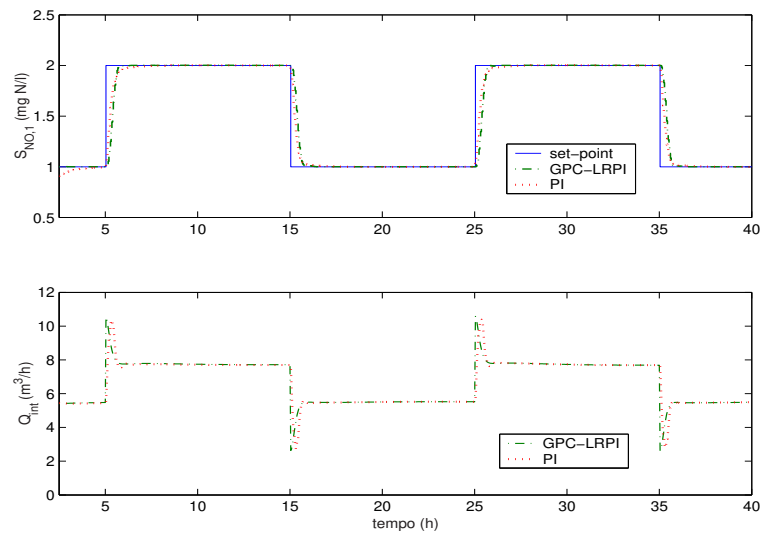


Fig. 6.16: Rastreamento do *set-point* da concentração de nitrato dos controladores GPC-LRPI e PI na última zona anóxica.

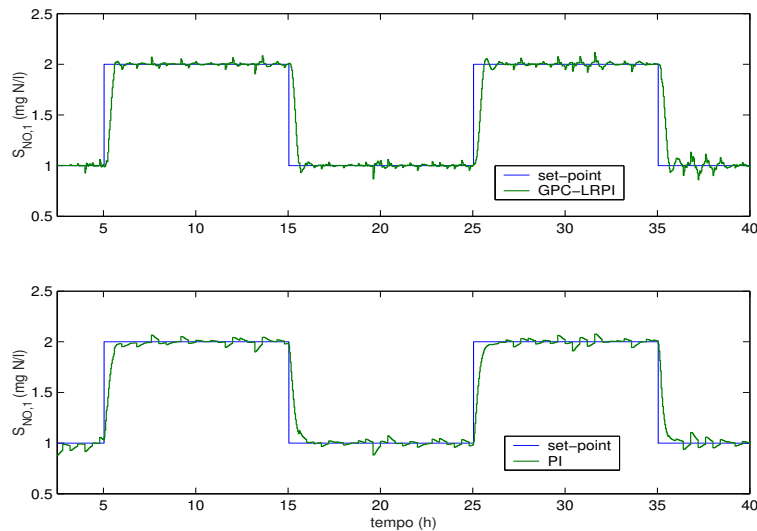


Fig. 6.17: Rastreamento do set-point da concentração de nitrato dos controladores GPC-LRPI e PI com ruído na saída.

GPC-LRPI, onde uma vez mais é verificado seu melhor desempenho se comparado com o desempenho apresentado pelo controlador PI.

Finalmente, a capacidade de rastreamento do *set-point* dos controladores pode ser vista na Figura 6.16, o que constitui o cenário 4. Neste caso o *set-point* da concentração de nitrato na zona anóxica é mudada a 2 mg N/l e após 10 horas é retornado ao valor de 1 mg N/l. O controlador GPC-LRPI mostra um rastreamento ligeiramente mais lento que o controlador PI, como consequência do valor do fator de penalização da entrada adotado. Valores menores de λ_u aumentam a velocidade de resposta do controlador, mas provocam um aumento do sobre-sinal máximo no cenário de rejeição de distúrbios. No entanto, sua capacidade de rastreamento é boa e, o mais importante, o controlador GPC-LRPI apresenta maior rapidez na rejeição dos distúrbios, objetivo altamente desejável no controle de plantas de tratamento de esgotos.

O desempenho no rastreamento do set-point do controlador GPC-LRPI é também testado e comparado com o controlador PI, adicionando ruído branco de média nula e variância 0,01 na saída da medição no sensor de nitrato. Este teste de desempenho pode ser visto na Figura 6.17. Observa-se também um bom desempenho para mudanças na referência do controlador GPC-LRPI.

6.4 Conclusões

Um controlador preditivo adaptativo baseado em modelos de identificação preditiva de horizonte estendido (GPC-LRPI) foi desenvolvido para controlar *on-line* a concentração de amônio na última zona aeróbia e a concentração de nitrato na saída da zona anóxica na planta de lodos ativados com pré-desnitrificação. Os esquemas propostos utilizaram modelos de ordem reduzida com uma pequena demanda de instrumentação, diferentemente de esquemas *feedback-feedforward* propostos na literatura, que requerem medidores de nitrato/amônio para os termos antecipatórios.

Foi mostrado que as diretrizes para o projeto de sinais de excitação relevantes para controle, utilizadas em conjunto com o algoritmo ortogonal de Gram-Schmidt Modificado, são efetivas para identificar modelos lineares de ordem reduzida. É importante ressaltar que incertezas de modelagem oriundas do procedimento de redução do modelo para controle, podem ser atenuadas usando a técnica de identificação preditiva de horizonte estendido.

O ajuste do controlador GPC-LRPI resulta particularmente complexo para o controle de nitrato e de amônio do processo de remoção de nitrogênio. A característica do processo de ter comportamento não linear e apresentar dinâmicas do tipo *stiff* impõem dificuldades ao ajuste do controlador. No entanto, foi mostrado que diretrizes de ajuste do controlador DMC encontradas na literatura de controle preditivo são um valioso referencial para o posterior ajuste baseado em ensaios de tentativa e erro. O compromisso proposto desta estratégia de ajuste entre o desempenho na rejeição a distúrbios externos e ao rastreamento do *set-point* demonstrou ser eficiente. O controlador GPC-LRPI mostrou ser mais estável e insensível às perturbações introduzidas na planta, se comparado ao desempenho do controlador PI para ambos processos. Portanto, o controlador GPC-LRPI constitui-se em uma potencial alternativa para a abordagem no controle da remoção de nitrogênio em plantas de lodos ativados.

Capítulo 7

Conclusões e Tópicos para Futuras Pesquisas

7.1 Introdução

No presente capítulo enumera-se as contribuições e conclusões desta tese divididas em cinco tópicos: geração de sinais relevantes para controle, identificação do processo de remoção de nitrogênio, modelagem para controle preditivo, controle preditivo generalizado baseado em identificação de preditores e o controle do processo de remoção de nitrogênio.

Finalmente, são apresentadas algumas recomendações e tópicos para futuras pesquisas dentro do contexto da estrutura assumida nas duas primeiras seções.

7.2 Contribuições do trabalho

As principais contribuições da tese são descritas resumidamente a seguir.

7.2.1 Geração de sinais

1. Realizou-se um estudo do estado da arte de sinais de excitação para a identificação de sistemas lineares e não lineares. Sinais de excitação estocásticos e determinísticos foram incluídos junto com a informação para a obtenção dos programas computacionais geradores destes sinais.
2. Implementou-se em Simulink um bloco gerador de sinais de excitação pseudo-aleatórios de sequência m multi-nível, o Gerador MLPRS. Características importantes como comprimento mínimo do sinal, persistência da excitação e potência espectral do sinal podem ser incorporadas introduzindo o número de elementos do Campo de Galois, a equação característica revisada e o mapeamento para supressão de harmônicos.
3. Demonstrou-se mediante simulações que diretrizes de projeto de sinais MLPRS baseadas no conhecimento *a priori* do processo são eficazes para projetar sinais com características *plant-friendly*, isto é, sinais cuja resposta em frequência abrangem a faixa de frequência útil para o controle e que ao mesmo tempo sejam apropriados para atuadores em plantas industriais (ver definição na seção 3.1).

7.2.2 Identificação

1. Foi proposto um procedimento sistemático para reduzir os modelos dos processos de nitrificação e desnitrificação via identificação, como resposta à necessidade de modelos que sejam dependentes dos objetivos de controle e da informação *a priori* disponível.
2. Identificou-se modelos com características relevantes para controle dos seguintes processos:
 - Processo de neutralização de pH. Processo de difícil modelagem por sua característica estática não linear.
 - Processo de concentração de oxigênio dissolvido no último reator aeróbio.
 - Processo de nitrificação caracterizado pela dinâmica da concentração de amônio na última zona aeróbia.
 - Processo de desnitrificação caracterizado pela dinâmica da concentração de nitrato na zona anóxica.

Cabe ressaltar que modelos parcimoniosos foram obtidos, potenciais candidatos para aplicação em controle baseado em modelo.

7.2.3 Modelagem para controle preditivo

1. Realizou-se um estudo comparativo na modelagem para preditores orientados ao projeto de controladores preditivos, visando reduzir a limitação do modelo preditor por iteração sucessiva padrão, no tocante à presença de dinâmicas não modeladas nas altas frequências e a redução da propagação do erro da predição k -passos-à frente convencional.
2. Realizou-se a formulação e implementação do modelo preditor de múltiplos-passos-à frente como alternativa ao modelo preditor por iteração sucessiva usado geralmente na síntese do controlador preditivo generalizado.

7.2.4 Controle GPC baseado em modelos preditores

1. Foi proposta uma formulação do controlador GPC baseado no modelo preditor de múltiplos-passos-à frente como um equivalente de um controlador linear por alocação de pólos especialmente útil para a implementação por blocos em Simulink[®], mediante *s-functions* do Matlab[®]. Esta implementação permite acompanhar a atualização dos parâmetros do controlador a cada passo de cálculo do algoritmo.
2. Realizou-se um estudo comparativo por simulação do desempenho em malha fechada dos controladores GPC baseados no modelo preditor por identificação preditiva de horizonte estendido e no modelo preditor de múltiplos-passos-à frente.

7.2.5 Controle do processo de remoção de nitrogênio

1. Demonstrou-se que a abordagem de desacoplamento do processo de remoção de nitrogênio considerando condições operacionais apropriadas e escalas de tempo similares é aplicável também ao modelo *benchmark* ASWWTP-USP.
2. Implementou-se a estratégia de controle local do processo de nitrificação usando um esquema de controle em cascata, onde o controlador GCP-LRPI atua como controlador mestre. Vale ressaltar que o objetivo de controle é levar o valor da concentração de amônio abaixo do valor recomendado na resolução CONAMA No. 20.
3. Implementou-se também a estratégia de controle local do processo de desnitrificação usando o controlador GPC-LRPI, que ajusta a vazão de recirculação interna com base a medições da concentração de nitrato na zona anóxica a valores desejados.

7.3 Conclusões

Nesta seção são listadas a seguir as principais conclusões obtidas do estudo e implementação dos procedimentos de identificação e controle propostos para o problema de remoção de nitrogênio na planta de tratamento de esgotos por lodos ativados com pré-desnitrificação.

7.3.1 Geração de sinais

1. Foi verificado que a supressão de harmônicos por mapeamento de sinais de excitação pseudo-aleatórios constitui também uma ferramenta útil para a identificação de modelos paramétricos não lineares ou mesmo na redução de distorções por não linearidades de sistemas não lineares para a identificação de sistemas lineares.
2. Foi mostrado como sinais de excitação pseudo-aleatórios com base em seqüências de comprimento máximo multi-nível (sinais MLPRS) são gerados mediante Campos de Galois, o que suportou a implementação do Gerador de Sinais MLPRS em Simulink[®].
3. Evidenciou-se a utilidade das diretrizes para o projeto de sinais MLPRS com características relevantes para controle, baseados em conhecimento *a priori* do processo, na identificação de processos industriais simulados e que ao mesmo tempo são apropriadas para atuadores (*plant-friendly*).

7.3.2 Identificação

1. A identificação do processo de neutralização de pH demonstrou a eficácia de um sinal MLPRS, com 3 níveis e supressão de harmônicos múltiplos de 2, na identificação de um modelo com grau de não linearidade igual a 2. A comparação com sinais estocásticos demonstrou a superioridade das predições do modelo usando o sinal MLPRS.

2. Na identificação do processo de Oxigênio Dissolvido foram determinados 2 modelos não lineares, usando um sinal gerado por um Campo de Galois de 7 elementos e harmônicos múltiplos de 2 suprimidos. Um destes modelos é adequado para aplicações de controle baseado em modelo, enquanto o outro pode ser empregado para o entendimento deste processo, devido à sua qualidade de predição superior. Foi identificado também um modelo linear parcimonioso, com um sinal de 3 níveis e harmônicos múltiplos de 2 suprimidos, candidato para projeto de controladores lineares baseados em modelo.

7.3.3 Modelagem para controle preditivo

Foi realizado um estudo comparativo de modelagem via identificação de preditores orientados ao projeto de controladores preditivos baseados em modelo, envolvendo três técnicas de modelagem: identificação preditiva com múltiplos-passos-à frente (MSPI), identificação preditiva de horizonte estendido (LRPI) e identificação de modelos um-passo-à frente (LS). As conclusões foram:

1. Os modelos MSPI e LRPI apresentam menor propagação de erros que o modelo LS.
2. O modelo MSPI é eficaz para horizontes curtos predição.
3. Para horizontes longos de predição, do ponto de vista do erro de predição, os modelos LRPI e MSPI são equivalentes, mas o modelo LRPI tem complexidade computacional inferior ao MSPI.

7.3.4 Controle GPC baseado em modelos preditores

1. A formulação modificada do controlador preditivo generalizado (GPC) baseada no modelo preditor múltiplos-passos-à frente, que utiliza o algoritmo recursivo de mínimos quadrados de Bierman para estimar os parâmetros do modelo *on-line*, mostrou se também adequada para a análise das propriedades do controlador em malha fechada, por ter a forma de um controlador linear por realimentação, como acontece com o controlador por alocação de pólos (Wellstead e Zarrop, 1991). Esta forma se mostrou também útil para a implementação do GPC baseado em modelo MSPI em Simulink[®].
2. Da análise comparativa do desempenho dos controladores GPC baseados nos modelos preditores por iteração sucessiva (GPC-LS), no modelo por identificação de horizonte estendido (GPC-LRPI) e por identificação de múltiplos-passos-à frente (GPC-MSPI) usando como planta o modelo de Rohrs da mesma forma que em Shook et al. (1991), observou-se que o desempenho no rastreamento do set-point para os controladores implementados é boa quando não é considerado ruído na medição. Notou-se a superioridade na adaptabilidade e robustez do controlador GPC-MSPI quando ruído aditivo é adicionado na variável de controle, comparado com o desempenho mostrado pelo controlador GPC-LRPI e PI. Em contra partida, o custo computacional do GPC-MSPI é maior devido ao aumento da ordem do modelo MSPI com o aumento do horizonte de predição.

3. Como resultado da análise do item anterior, selecionou-se o controlador GPC-LRPI para ser aplicado no controle de remoção de matéria nitrogenada na planta de lodos ativados. O preditor LRPI estima um modelo relevante para o controle preditivo conservando a ordem do modelo, o que significa um menor custo computacional e erros de predição próximos ao do controlador GPC-MSPI, para horizontes de predição longos (maiores que 10).

7.3.5 Controle do processo de remoção de nitrogênio

1. As diretrizes para o projeto de sinais de excitação relevantes para controle, utilizadas em conjunto com o algoritmo ortogonal de Gram-Schmidt Modificado, mostraram-se eficazes para identificação de modelos lineares de ordem reduzida das dinâmicas de nitrato e amônio na planta de lodos ativados.
2. Verificou-se nas simulações com o modelo *benchmark* ASWWTP-USP, usado como a planta de tratamento de esgotos por lodos ativados, que a estratégia de decompor o processo de remoção de matéria nitrogenada no controle de nitrato na zona anóxica e no controle de amônia na zona aeróbia, é uma alternativa válida quando consideradas constantes de tempo na escala média (na casa de horas a dias).
3. O controlador preditivo adaptativo GPC-LRPI foi aplicado como controlador mestre para controlar *on-line* a concentração de amônio na última zona aeróbia da planta de lodos ativados com pré-desnitrificação. A estratégia de controle em cascata mostrou ser eficiente para manter a concentração de amônio no nível especificado, de 4 mg N/l, mostrando um rastreamento satisfatório do set-point de amônio, mesmo no caso de ser adicionado ruído à saída do sensor de amônio. Para este caso, foi comprovada uma demanda de oxigênio dissolvido no reator nitrificante menor que 2 mg O₂/l, valor de referência da estratégia de controle de OD fixo. Com isto foi obtida uma economia nos custos de aeração, custo este que representa uma parte importante dentro do total de custos operacionais da planta. A estratégia de controle implementada mostrou também ser robusta na rejeição de distúrbios externos representados por mudanças na carga de substrato e amônio do afluente, quando comparada à estratégia de controle de OD fixo.
4. O controlador GPC-LRPI também foi aplicado no controle da concentração de nitrato na saída da zona anóxica na planta. Quando a carga de matéria orgânica é suficientemente alta, a vazão de recirculação interna é uma variável de controle eficaz para manter a concentração de nitrato em 1 mg N/l. O desempenho do controlador para rastreamento do set-point e da rejeição de distúrbios mostrou ser satisfatório, com maior insensibilidade aos distúrbios se comparado com o controle PI. Isto ocorreu porque incertezas de modelagem oriundas do procedimento de redução do modelo para controle são atenuadas usando a técnica de identificação preditiva de horizonte estendido.
5. Os esquemas propostos utilizaram modelos de ordem reduzida com uma pequena demanda de instrumentação, diferentemente de esquemas *feedback-feedforward*, que requerem um modelo e instrumentação adicional para os termos antecipatórios.

7.4 Tópicos para futuras pesquisas

A seguir são apresentados resumidamente alguns tópicos sugeridos para futuras pesquisas.

7.4.1 Geração de sinais de excitação

Verificou-se ao longo do presente trabalho a importância da geração de sinais de excitação com características relevantes para controle. Modelos paramétricos foram obtidos para representar processos industriais mediante estruturas lineares e não lineares. No entanto, a identificação de modelos não paramétricos, no domínio da frequência, oferece um grande potencial na avaliação das propriedades do modelo e constituem um valioso complemento aos métodos paramétricos, especialmente na identificação de modelos não lineares. O procedimento de supressão de harmônicos no sinal de excitação permitiu uma separação dos componentes de saída do sistema devido a não linearidades de ordem par e ímpar e mesmo a separação parcial de outros componentes. Portanto, a pesquisa de outros tipos de sinais com apelo prático torna-se importante.

Sinais de excitação com um número ótimo de níveis (Barker et al., 2003), constituem uma alternativa para gerar sinais pseudo-aleatórios com espectro uniforme e supressão de harmônicos como sugerido para a identificação de sistemas não lineares. No entanto, sinais pseudo-aleatórios multi-nível não proporcionam algumas vezes os limites de frequência requeridos. Isto pode originar sinais com comprimentos maiores do que o necessário. Sinais multi-nível podem não prover a variância de níveis requerida, dado que comutam em intervalos fixos originando mudanças abruptas. Este pode ser um efeito indesejado na propriedade de sinais *plant-friendly*. Uma alternativa são os sinais multi-senoidais (Braun et al., 2000; Rivera et al., 2003) que também são determinísticos com fator de pico mínimo e não têm restrições no número de níveis.

7.4.2 Identificação para controle preditivo

No desenvolvimento da metodologia de identificação de modelos para controle preditivo foi considerado o enfoque em malha aberta. No entanto, constitui campo de ativa pesquisa o enfoque na identificação de modelos para controle preditivo em malha fechada, que apresenta um envolvimento matemático mais complexo mas que ao mesmo tempo integra dentro do índice de desempenho o requisito de desempenho do sistema de controle e da tarefa de identificação.

Uma forma de enunciar uma medida de desempenho é estabelecer a função $K(G_o, C)$ que pode ser medida usando a norma $\|J(G_o, C)\|_p$ (Van den Hof e Schrama, 1995). A função $K(G_o, C)$ depende de G_o (o sistema real) e C (o controlador), mas também depende das propriedades do sinal. O propósito da identificação relevante para controle é encontrar um modelo \tilde{G} , tal que uma vez que o controlador C seja projetado conforme \tilde{G} , o desempenho do sistema em malha fechada seja otimizado. Uma forma de enunciar uma função de desempenho é mediante o critério de controle preditivo:

$$J(u, t) = E \left\{ \sum_{j=N_1}^{N_2} [\hat{y}(t + k/t) - r(t + k)]^2 + \lambda_u \sum_{j=1}^{N_c} [\Delta u(t + k - 1)]^2 \right\}$$

que pode ser inserida em $\|J(G_o, C)\|_2^2$ com (Van den Hof e Schrama, 1995):

$$\|J(G_o, C)\|_2^2 = [y(t) - r(t) \quad \sqrt{\lambda}u(t)]^T$$

como sendo um sinal quase-estacionário bidimensional. Uma abordagem interessante a ser pesquisada é a identificação em malha fechada baseada em múltiplos-passos-à frente dos erros de predição (Gopaluni et al., 2003).

7.4.3 Controle preditivo baseado em multimodelos

Os processos industriais na prática são inerentemente não lineares. A suposição adotada nos procedimentos de identificação e controle é que estes processos estão dentro de vizinhança do ponto de operação desejado. É desejável que o modelo seja capaz de descrever a operação da planta para regimes de operação adversos, como são as grandes mudanças na carga de afluente na planta de tratamento de esgotos. Uma alternativa de modelagem para diferentes regimes de operação constitui a abordagem por multimodelos (Johansen e Foss, 1997). Com esta consideração é possível construir simultaneamente k -passos modelos de predição separados, um para cada horizonte de predição, que resolveria também o problema proposto pelo método de identificação preditiva de horizonte estendido. O modelo poderia ser selecionado por uma função de interpolação que considere o erro de predição de cada modelo para mudanças no regime de operação.

7.4.4 Controle da remoção de nutrientes

O controlador preditivo baseado na identificação relevante para controle de modelos preditores proposto é eficiente para os casos de distúrbios externos relativamente grandes e com carga orgânica carbonácea suficiente. Existem casos extremos de mudanças por chuvas ou por mudanças maiores nas concentrações dos compostos orgânicos. Uma alternativa para contornar estes casos é inserir controle *feedforward* na lei do controlador preditivo generalizado. Desta forma o modelo CARIMA seria formulado como:

$$A(q^{-1})y(t) = B(q^{-1})u(t - d) + D(q^{-1})\omega(t - d) + \frac{C(q^{-1})}{\Delta}\xi(t) \quad (7.1)$$

com os seguintes polinomiais:

$$\begin{aligned} A(q^{-1}) &= 1 + a_1q^{-1} + \dots + a_{n_a}q^{-n_a} \\ B(q^{-1}) &= b_0 + b_1q^{-1} + \dots + b_{n_b}q^{-n_b} \\ D(q^{-1}) &= d_0 + d_1q^{-1} + \dots + d_{n_d}q^{-n_d} \\ C(q^{-1}) &= 1 + c_1q^{-1} + \dots + c_{n_c}q^{-n_c} \end{aligned}$$

onde ω representa sinais conhecidos tais como distúrbios medidos. O preditor k -passos-à frente correspondente:

$$\hat{y}(t + k/t) = G'_k \Delta u(t + k - 1) + y_o(t + k/t)$$

onde:

$$y_o(t + k/t) = \frac{F_k(q^{-1})}{C(q^{-1})}y(t) + \frac{H_k(q^{-1})}{C(q^{-1})}\Delta u(t - 1) + \frac{E'_k(q^{-1})}{C(q^{-1})}D(q^{-1})\Delta\omega$$

e E'_k , F_k , G'_k e H_k são soluções de:

$$\begin{aligned} C(q^{-1}) &= E'_k(q^{-1})A(q^{-1})\Delta + q^{-k}F_j(q^{-1}) \\ E'_k(q^{-1})B_{t+k}q^{-d+1} &= C(q^{-1})G'_k(q^{-1}) + q^{-k}H_k(q^{-1}) \end{aligned}$$

O termo *feedforward* inserido é uma estratégia eficiente devido à atrasos de transporte hidráulico. A idéia é atuar no processo quando os distúrbios no afluente apareçam e antes que provoquem maiores mudanças no efluente. Esta estratégia pode ser factível com o algoritmo de identificação ortogonal MGS, que incorpora dentro da estrutura NARX o termo de distúrbio medido (ver Apêndice C). Assim sendo, podem ser consideradas as seguintes modificações à estrutura de controle propostas.

Nitrificação

O controlador mestre GPC-LRPI poderia ser usado considerando a carga de amônio na entrada da primeira zona aeróbia, que implicaria em uma maior rejeição ao distúrbio, tal como comprovado por Ingildsen (2002) com controladores PI em cascata. Obviamente isto implicaria um aumento no custo de instrumentação, por ser agora necessário mais um sensor de amônio.

Na desnitrificação

Se o requisito de qualidade de efluente não é satisfeito para máxima remoção de nitrato na zona anóxica, utilizando a vazão de recirculação interna como variável de controle, então a alternativa seria usar uma fonte externa de carbono. A dosagem de carbono externo poderia ser combinada com o controle da vazão de recirculação interna. Desta forma a vazão externa de carbono seria a ação de controle usando como variável de referência uma média da concentração de nitrato e amônio. Isto leva em consideração o nitrogênio total na saída do reator aeróbio.

O controle da concentração de nitrato mediante a vazão de recirculação interna asseguraria uma utilização completa do volume anóxico. Vale ressaltar que é recomendável inserir uma restrição na dosagem de carbono, devido a presença de agentes tóxicos inibidores que causariam um aumento na vazão de carbono.

Finalmente, sugere-se uma implementação da lei de controle preditivo baseada em modelos de identificação preditiva multivariável. Desta forma seria considerada uma função custo que incorporasse mais variáveis, principalmente as que mostrem maior acoplamento. Uma valiosa referência para a implementação do GPC multivariável é encontrada em Camacho e Bordons (1995).

Referências Bibliográficas

- Aguirre, L. A. (2000). *Introdução à identificação de sistemas: técnicas lineares e não-lineares aplicadas a sistemas reais*, Editora UFMG, Belo Horizonte.
- Andrews, J. F. (1975). Dynamics models and control strategies for biological wastewater treatment processes, in K. a. Wanielista (ed.), *Mathematical Modeling for Water Pollution Control Processes*, Ann Arbor Science Publishing Inc, chapter 9, pp. 319–378.
- Arceivala, S. J. (1981). *Wastewater treatment and disposal : engineering and ecology in pollution control*, Marcel Dekker, New York.
- Ayesa, E., Carstensen, J., Jeppsson, U. e Vanrolleghem, P. A. (1994). Identification of the dynamic processes in WWTP and control of WWTP, chapter 10, *Technical report*, COST-682 Working Group Meeting, San Sebastian, Spain.
- Barker, H. A. (1986). Sum and product tables for galois fields, *International Journal of Mathematics and Educational Science Technology* **17**(4): 473–485.
- Barker, H. A. (1993). Design of multilevel-pseudo-random signals for system identification, in K. R. Godfrey (ed.), *Perturbation Signals for Systems Identification*, Prentice Hall Int, U.K., chapter 11, pp. 321–347.
- Barker, H. A. e Godfrey, K. R. (1999). System identification with multi-level periodic perturbation signals, *Control Engineering Practice* **7**: 717–726.
- Barker, H. A., Godfrey, K. R. e Tucker, A. J. (2000). Nonlinear system identification with multilevel perturbation signals, *Proceedings of the IFAC System Identification, SYSID 2000*, San Diego, CA.
- Barker, H. A., Tan, H. A. e Godfrey, K. R. (2003). The performance of multilevel perturbation signals for nonlinear system identification, *Proceedings of the IFAC System Identification, SYSID 2003*, Rotterdam, Netherland.
- Barker, H. A. e Zhuang, M. (1997a). Design of pseudo-random perturbation signals for frequency-domain identification of nonlinear systems, *Proceedings of the IFAC Symposium on System Identification, SYSID 1997*, Japan, pp. 1635–1640.
- Barker, H. A. e Zhuang, M. (1997b). Galois: A program for generating pseudo-random perturbation signals, *Proceedings of the IFAC Symposium on System Identification, SYSID 1997*, Japan, p. 1641.

- Benchiheb, Y. (1997). *Dynamic modelling and optimization of activated sludge wastewater treatment in relation to investment and operating costs*, Master's thesis, University of Gent, Belgium.
- Bitmead, R. R., Gevers, M. e Wertz, V. (1990). *Adaptive Optimal Control: The Thinking Man's*, Prentice Hall, Englewood Cliffs, NJ.
- Blondel, V., Gevers, M. e Bitmead, R. R. (1997). When is model good for control design?, *Proceedings of the 36th IEEE Conference on Decision and Control*, San Diego, California, pp. 1283–1288.
- Bordons, C. e Camacho, E. F. (1998). A generalized predictive controller for a wide class of industrial processes, *IEEE Transactions on Control System Technology* **6**(3): 372–387.
- Braun, M. W. (2001). *Model-on-Demand nonlinear estimation and model predictive control: novel methodologies for process control and supply chain management*, PhD thesis, Arizona State University, Arizona State University.
- Braun, M. W., Ortiz–Mojica, R. e Rivera, D. E. (2000). Design of minimum crest factor multisensoidal signals for “plant-friendly” identification of nonlinear process systems, *Proceedings of the IFAC System Identification, SYSID 2000*, Santa Barbara, CA.
- Braun, M. W., Rivera, D. E. e Stenman, A. (2001). A ‘Model-on-Demand’ identification methodology for nonlinear process systems, *International Journal of Control* **74**(18): 1708–1717.
- Braun, M. W., Rivera, D. E., Stenman, A., Foslien, W. e Hrenya, C. (1999). Multilevel pseudo-random signal design and “Model on Demand” estimation applied to nonlinear identification of a rtp wafer reactor, *Proceedings of the American Control Conference*, San Diego, CA.
- Camacho, E. F. e Bordons, C. (1995). *Model Predictive Control in the Process Industry*, Springer-Verlag London Ltd.
- Carstensen, J. (1994). *Identification of Wastewater Processes*, PhD thesis, Technical University of Denmark, Denmark.
- Chen, S. e Billings, S. A. (1989). Representation of non linear systems: the NARMAX model, *International Journal of Control* **49**(3): 1013–1032.
- Chen, S., Billings, S. A. e Luo, W. (1989). Orthogonal least square methods and their application to nonlinear system identification, *International Journal of Control* **50**(5): 1873–1896.
- Clarke, D. W. (1988). Application of generalized predictive control to industrial process, *IEEE Control System Magazine* **122**: 49–55.
- Clarke, D. W. (1996). Adaptive predictive control, *A. Rev. Control* **20**: 83–94.
- Clarke, D. W. e Mohtadi, C. (1989). Properties of generalized predictive, *Automatica* **25**: 859–875.
- Clarke, D. W., Mohtadi, C. e Tuffs, P. S. (1987). Generalized predictive control - part I. the basic algorithm, *Automatica* **23**(2): 137–148.

- Copp, J. (2000). *The Cost Simulation Benchmark: Description and Simulator Manual*, a product of COST Action 624 and COST Action 682, COST: European Cooperation in the field of Scientific and Technical Research, Luxemburgo.
- Cormack, A. S., Godfrey, K. R. e Flower, J. O. (1995). Design of multilevel multiharmonic signal for system identification, *IEE Proceedings Control Theory and Applications* **142**(3): 247–252.
- Cott, B. J. (1995). Introduction to the Process Identification Workshop at the 1992 Canadian Chemical Engineering Conference, *Journal of Process Control* **5**(2): 67–69.
- Cutler, C. R. e Ramaker, B. L. (1980). Dynamic matrix control - a computer control algorithm, *Joint Automatic Control Conference, JACC*, Proceedings, San Francisco, CA.
- Davies, W. (1970). *System Identification for Self-Adaptive Control*, John Wiley and Sons, U.K.
- Doyle, F. J., Ogunnaike, B. A. e K. Pearson, R. (1999). Nonlinear model-based control using second order volterra models, *Automatica* **31**: 697–714.
- Fardin, J. F. (2000). *Filtragem relevante para controle: uma abordagem baseada na identificação de preditores*, Tese de Doutorado, DAC/FEEC - Universidade Estadual de Campinas, Campinas, Brasil.
- Freitas, F. F., Henrique, H. M. e Xavier, A. F. (2002). Construção e automação de uma planta de lodos ativados, *Congresso Brasileiro de Engenharia Química, COBEQ 2002*, Universidade Federal do Rio Grande do Norte, Proceedings, CD Rom, Natal, RN.
- Gaikwad, S. V. e Rivera, D. E. (1996). Control-relevant input signal design for multivariable system identification: Application to high purity distillation, *Proceedings of the IFAC World Congress*, San Francisco, CA., pp. 349–354.
- Gevers, M. (1997). Identification for control, *A. Rev. Control* **20**: 95–106.
- Godfrey, K. R. (1993). Introduction to perturbation signals for time-domains system identification, in K. R. Godfrey (ed.), *Perturbation Signals for Systems Identification*, Prentice Hall Int, U.K., chapter 1, pp. 1–59.
- Godfrey, K. R., Barker, H. A. e Tucker, A. J. (1999). Comparison of perturbation signals for linear systems identification in the frequency domain, *IEE Proceedings Control Theory and Applications* **146**(6): 535–548.
- Godfrey, K. R., Tan, A. H. e Barker, H. (2003). A survey of readily accesible perturbation signals, *Proceedings of the 13th IFAC Symposium on System Identification*, Rotterdam, The Netherlands.
- Goodwin, G. C. e Payne, R. L. (1977). *Dynamic System Identification: Experiment Design and Data Analysis*, Academic Press, Inc.
- Goodwin, G. C. e Sin, K. S. (1984). *Adaptive Filtering: Prediction and Control*, Prentice Hall, Englewood Cliff.

- Gopaluni, R. B., Patwardhan, R. S. e Shah, S. L. (2002). Experiment design for MPC relevant identification, *Proceedings of the American Control Conference*, Anchorage, Alaska, pp. 2713–2718.
- Gopaluni, R. B., Patwardhan, R. S. e Shah, S. L. (2003). The nature of data pre-filters in MPC relevant identification-open- and closed-loop issues, *Automatica* **39**: 1617–1626.
- Greco, C., Menga, G., Mosca, E. e G.Zappa (1984). Performance improvements of self-tuning controllers by multistep horizons: The MUSMAR approach, *Automatica* **20**(5): 681–699.
- Gujer, W., Henze, M., Mino, T. e Van Loosdrecht, M. (1999). Activated sludge model no. 3, *Water Science and Technology* **39**(1): 183–193.
- Gustafsson, T. K. e Waller, K. V. (1983). Dynamic modelling and reaction invariant control of ph, *Chemical Engineering Science* **38**(3): 389–398.
- Haber, R., Schmitz, U. e Bars, R. (2003). Long range optimal model and multi-step-ahead prediction for predictive control, *Proceedings of the 13th IFAC Symposium on System Identification, SYSID 2003*, Rotterdam, The Netherlands.
- Henson, M. A. (1998). Nonlinear model predictive control: current status and future directions, *Computers and Chemical Engineering* **23**: 187–202.
- Henson, M. A. e Seborg, D. E. (1997). Adaptive input-output linearization of a ph neutralization process, *International Journal of Adaptive Control and Signal Processing* **11**(3): 171–200.
- Henze, M., Grady Jr., C. P., Gujer, W., Marais, G. V. R. e Matsuo, T. (1987a). Activated sludge model no. 1, *IAWQ Scientific and Technical Report no. 1*, IAWQ, London, Great Britain.
- Henze, M., Grady Jr., C. P., Gujer, W., Marais, G. V. R. e Matsuo, T. (1987b). A general model for single-sludge wastewater treatment systems (abbreviated report), *Water Resources* **21**(5): 505–515.
- Henze, M., Gujer, W., Mino, T., Matsuo, T., Wentzel, M. C., Marais, G. V. R. e Van Loosdrecht, M. (1999). Activated sludge model no. 2d, *Water Science and Technology* **39**(1): 165–182.
- Henze, M., Gujers, W., Mino, T., Matsuo, T., Wentzel, M. C. e Marais, G. V. R. (1995). Activated sludge model no. 2, *IAWQ Scientific and Technical Report no. 3*, IAWQ, London, Great Britain.
- Hjalmarsson, H., Gevers, M. e De Bruyne, F. (1996). For model-based control design, closed-loop identification gives better performance, *Automatica* **32**(12): 1659–1673.
- IBGE (2002). *Indicadores de Desenvolvimento Sustentável: Brasil 2002*, Diretoria de Geociências, Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, Brasil.
- Ingildsen, P. (2002). *Realising Full-Scaling Control in Wastewater Treatment Systems Using In Situ Nutrient Sensors*, PhD thesis, Lund University, Lund, Sweden.

- Jeppsson, U. (1996). *Modelling Aspects of Wastewater Treatment Processes*, PhD thesis, Lund Institute of Technology (LTH), Lund, Sweden.
- Jeppsson, U., J. Alex, H. S. e Vanrolleghem, P. (2002). Status and future trends of ICA in wastewater treatment - a european perspective, *Water Science and Technology* **45**(4-5): 485–494.
- Jerônimo, R. A., Sotomayor, O. A. Z., Garcia, C. e Park, S. W. (2000). Polynomial nonlinear modeling of biological nutrients removal (bnr) activated sludge processes, *Anais do XIII Congresso Brasileiro de Automática, CBA 2000*, Florianópolis, Santa Catarina, pp. 366–371.
- Johansen, T. A. e Foss, B. A. (1997). Operating regime process modeling and identification, *Computers Chemical Engng* **21**(2): 159–176.
- Lara, J. M. V. (1999). *Controle preditivo adaptativo não linear aplicado a um processo de neutralização de pH*, Tese de Mestrado, LAC/PTC - Escola Politécnica da Universidade de São Paulo, São Paulo, Brasil.
- Lara, J. M. V. e Milani, B. E. A. (2003). Identification of neutralization process using multi-level pseudo-random signals, *Proceedings of the 2003 American Control Conference*, Denver, Colorado, pp. 3822–3827.
- Lara, J. M. V. e Milani, B. E. A. (2005a). Adaptive predictive control of ammonium level based on dissolved oxygen variable set-point control in an activated sludge process, *4th Mercosur Congress on Process Systems Engineering, ENPROMER 2005*, Rio de Janeiro, Brazil. (abstract accepted).
- Lara, J. M. V. e Milani, B. E. A. (2005b). Control of nitrate flow in pre-denitrification systems using identification for long-range predictive control, *7th IASTED International Conference on Control and Applications, CA 2005*, Cancun, México. (paper submitted).
- Lee, J. H. (2000). Modeling and identification for nonlinear model predictive control: Requirements, current status and future research needs, in F. Allgower e A. Zheng (eds), *Nonlinear Model Predictive Control*, Vol. 26, Birkhauser Verlag, Basel, Birkhauser Verlag, Basel.
- Lim, R. K. e Phan, M. Q. (1997). Identification of a multistep-ahead observer and its application to predictive control, *Journal of Guidance, Control, and Dynamics* **20**(6): 1200–1206.
- Lindberg, C. F. (1997). *Control and Estimation Strategies Applied to the Activated Sludge Processes*, PhD thesis, Uppsala University, Sweden.
- Liu, D., Shah, S. e Fisher, D. G. (1999). Multiple prediction models for long range predictive control, *Proceedings of the 14th World Congress of IFAC*, Beijing, China, pp. 175–180.
- Ljung, L. (1987). *System Identification: Theory for the User*, Prentice Hall International.
- Ljung, L. (1996). Development of system identification, *Technical Report LiTH-isy-R-1910*, Dept of EE. Linköping University, S-581 83 Linköping, Sweden.
- Ljung, L. (1999). *System Identification: Theory for the User, second edition*, Prentice Hall PTR.

- Ljung, L. e Glad, T. (1994). *Modeling of Dynamic Systems*, Prentice Hall PTR.
- Londong, J. (1992). Strategies for optimized nitrate reduction with primary denitrification, *Water Science and Technology* **26**(5-6): 1087–1096.
- Lukasse, L. J. S. (2002). *Control and Identification in Activated Sludge Processes*, PhD thesis, Wageningen University, Wageningen, The Netherlands.
- Maciejowsky, J. M. (2002). *Predictive Control with Constraints*, Pearson Education Limited, London.
- Marsili-Libelli, S. (1990). Adaptive estimation of bioactivities in the activated sludge process, *IEE Proceedings, Pt. D* **137**(6): 349–356.
- Marsili-Libelli, S. e Vaggi, A. (1997). Estimation of respirometric activities in bioprocesses, *Journal of Biotechnology* **52**: 181–192.
- Mayne, D. Q., Rawlings, J. B., Rao, C. V. e Sokaert, P. O. M. (2000). Constrained model predictive control: Stability and optimality, *Automatica* **36**: 789–814.
- McIntosh, A. R., Shah, S. L. e Fisher, D. G. (1991). Analysis and tuning of adaptive generalized predictive control, *Can. J. Chem. Eng.* pp. 69–97.
- Meadows, E. S. e Rawling, J. B. (1997). Model predictive control, in M. A. Henson e D. E. Seborg (eds), *Nonlinear Process Control*, Prentice Hall, Inc., pp. 233–310.
- Mohtadi, C. (1988). On the role of prefiltering in parameter estimation and control, in S. L. Shah e G. Dumont (eds), *Lecture Notes in Control and Information Sciences*, Vol. 137, Springer-Verlag, Berlin, pp. 121–144.
- Monod, J. (1949). The growth of bacterial cultures, *Annual Review in Microbiology* **3**: 371–394.
- Morari, M. e Lee, J. H. (1999). Model predictive control: past, present and future, *Computers and Chemical Engineering* **23**: 667–682.
- Mosca, E. (1995). *Optimal, Predictive and Adaptive Control*, Prentice Hall Information and System Sciences Series.
- Nikolaou, M. (2001). Model predictive controllers: A critical synthesis of theory and industrial needs, *Advances in Chemical Engineering Series*, Academic Press.
- Norton, J. P. (1986). *An introduction to Identification*, Academic Press: New York.
- Oliveira, G. H. C., do Amaral, W. C. e Latre, L. G. (1996). A robustness approach for prefiltering the gpc, *Computers Electr. Engng.* **22**(5): 315–324.
- Olsson, G. (1993). Advancing ICA technology by eliminating the constraints, *Water Science and Technology* **28**(11-12): 1–7.

- Olsson, G. (2002a). Lectures notes of control of biological wastewater treatment course. Industrial Electrical Engineering and Automation, Lund University, http://www.iea.lth.se/sbr/lectures/L1_WWT_introduction.pdf, último acesso 20 Outubro 2004.
- Olsson, G. (2002b). Lessons learnt at ica2001, *Water Science and Technology* **45**(4-5): 1–8.
- Olsson, G. e Newell, B. (1999). *Wastewater Treatment Systems: Modelling, Diagnosis and Control*, 1th edn, IWA Publishing, London, UK.
- Pearson, R. K., Menold, P. H. e Allgöwer, F. (1998). Practically-motivated input sequences for nonlinear model identification, *Technical Report AUT98-08*, Dept. of Information Technology and Electrical Engineering, Swiss Federal Institute of Technology (ETH), Switzerland.
- Pearson, R. K. e Ogunnaike, B. A. (1997). Nonlinear process identification, in M. A. Henson e D. E. Seborg (eds), *Nonlinear Process Control*, Prentice Hall, Inc., chapter 2, pp. 11–110.
- Pröll, T. (1993). *Model predictive control based on nonlinear autoregressive and neural networks models*, PhD thesis, Colorado State University, Colorado, USA.
- Quin, S. J. e Badgwell, T. A. (1996). An overview of industrial model predictive control technology, *Proceedings of the 5th Conference on Chemical Process Control -CPC V*, Tahoe City, California.
- Quin, S. J. e Badgwell, T. A. (2000). An overview of nonlinear model predictive control applications, in F. Allgower e A. Zheng (eds), *Nonlinear Model Predictive Control*, Vol. 26, Birkhauser Verlag, Basel, Birkhauser Verlag, Basel.
- Quin, S. J. e Badgwell, T. A. (2003). A survey of industrial model predictive control technology, *Control Engineering Practice* **11**(7): 733–764.
- Rawlings, J. B. (1999). Tutorial: Model predictive control technology, *Proceedings of the American Control Conference*, San Diego, California, pp. 662–676.
- Rehnström, A. (2000). *Automatic control of an activated sludge process in a wastewater treatment plant - a benchmark study*, Master's thesis, Uppsala University, Uppsala, Sweden.
- Rivera, D. E. e Flores, M. E. (2000). Beyond step testing and process reaction curves: introducing meaningful system identification concepts in the undergraduate chemical engineering curriculum, *Proceedings of the IFAC System Identification*, Santa Barbara, California, pp. 815–820.
- Rivera, D. E., Lee, H. e Braun, M. W. (2003). "Plant-friendly" system identification: a challenge for the process industries, *Proceedings of the 12th IFAC Symposium on System Identification, SYSID 2003*, Rotterdam, The Netherlands, pp. 917–922.
- Rivera, D. E., Pollard, J. F. e Garcia, C. E. (1992). Control-relevant prefiltering: A systematic design approach and case study, *IEEE Transactions on Automatic Control* **37**(7): 964–974.
- Robinson, B. D. e Clarke, D. W. (1991). Robustness effects of a prefilter in generalised predictive control, *IEE Proceedings-Pt.D* **138**(1): 2–8.

- Rohrs, C. E., Athans, M., Valavani, L. e Stein, G. (1984). Some design guidelines for discrete-time adaptive controllers, *Automatica* **20**(7): 991–994.
- Rossiter, J. A. (2000). Identification of models for predictive control, *IEE Workshop Model Validation for Plant Control and Condition Monitoring*, London, UK, pp. 1–7.
- Rossiter, J. A. e Kouvaritakis, B. (2001). Modelling and implicit modelling for predictive control, *International Journal of Control* **74**(11): 1085–1095.
- Rossiter, J. A., Yao, L. e Kouvaritakis, B. (2000). Identification of predictions models for a non-linear power generation model, *Proceedings of the IEE International Conference on Control*, Cambridge, UK.
- SABESP (2002). Volumes de Água e esgoto: 2002. Companhia de Saneamento Básico do Estado de São Paulo, http://www.sabesp.com.br/relacao_investidores/informacoes_operacionais/vol_agua_regiao.htm, último acesso 08 de julho 2003.
- Samuelsson, P. e Carlsson, B. (2001). Feedforward control of the external carbon flow rate in an activated sludge process, *Water Science and Technology* **43**(1): 115–122.
- Schön, T. (2001). *Identification for predictive control - a multiple model approach*, Master's thesis, Dept. of Electrical Engineering - Linköping University, Linköping, Sweden. Reg. Nr.LiTH-ISY-3164.
- Seborg, D. E. (1999). A perspective on advances strategies for process control (Revisited), in P. M. Frank (ed.), *Advances in Control-Highlights of ECC'99*, London: Springer-Verlag, pp. 103–134.
- Seborg, D. E., Edgar, T. F. e Mellichamp, D. A. (1989). *Process Dynamics and Control*, John Wiley and Sons.
- Shook, D. S., Mohtadi, C. e Shah, S. L. (1991). Identification for long-range predictive control, *IEE Proceedings-D* **138**(1): 75–84.
- Shook, D. S., Mohtadi, C. e Shah, S. L. (1992). A control-relevant identification strategy for gpc, *IEEE Transactions on Automatic Control* **37**(7): 975–980.
- Shridhar, R. e Cooper, D. J. (1997). A tuning strategy for unconstrained siso model predictive control, *Ind. Eng. Chem. Res.* **36**: 729–746.
- Singman, J. (1999). *Efficient control of wastewater treatment plant - a benchmark study*, Master's thesis, Uppsala University, Uppsala, Sweden.
- Sánchez, J. M. M. e Rodellar, J. (1996). *Adaptive Predictive Control: From the concepts to plant optimization*, Prentice Hall International, Series in Systems and Control Engineering, London, UK.

- Sotomayor, O. A. Z. (1997). *Modelamento, simulação e controle de um processo de neutralização de pH*, Tese de Mestrado, LAC/PTC - Escola Politécnica da Universidade de São Paulo, São Paulo, Brasil.
- Sotomayor, O. A. Z., Park, S. e Garcia, C. (2000). Nitrate concentration-based control of a pre-denitrifying activated sludge system, *IFAC International Symposium on Advanced Control of Chemical Processes, ADCHEM 2000*, Pisa, Italy, pp. 213–218.
- Sotomayor, O. Z., Park, S. W. e Garcia, C. (2001). A simulation benchmark to evaluate the performance of advanced control techniques in biological wastewater treatment plant, *Brazilian Journal of Chemical Engineering*, **18**(1): 81–101.
- Spanjers, H. (2000). Modelling and control in sustainable wastewater treatment, *New World Water*, pp. 25–31.
- Srinivasan, R. e Rengaswamy, R. (1999). Use of inverse repeat sequence (IRS) for identification in chemical process systems, *Industrial and Engineering Chemical Research* **38**: 3420–3429.
- Steffens, M. A. (1997). *Advanced Process Control for Nutrient Removal Activated Sludge Processes*, PhD thesis, University of Queensland, Australia.
- Steffens, M. A., Lant, P. A. e Newell, R. B. (1997). A systematic approach for reducing complex biological wastewater treatment models, *Water Research* **31**(3): 590–606.
- Suescon, J., Irizar, I., Ostalaza, X. e Ayesa, E. (1998). Dissolved oxygen control and simultaneous estimation of oxygen uptake rate in activated-sludge plants, *Water Environment Research* **70**(3): 316–322.
- Suescon, J., Ostalaza, X., Garcia-Sans, M. e Ayesa, E. (2001). Real-time control strategies for predenitrification-nitrification activated sludge plants biodegradation control, *Water Science and Technology* **43**(1): 209–216.
- Takács, I., Patry, G. G. e Nolasco, D. (1991). A dynamic model of clarification-thickening process, *Water Research*, **25**(10): 1263–1271.
- Tan, A. H. e Godfrey, K. R. (2001). The generation of binary and near-binary pseudo-random signals: an overview, *IEEE Instrumentation and Measurement Technology Conference*, Budapest, Hungary, pp. 766–771.
- Van den Hof, P. e Schrama, R. J. P. (1995). Identification and control: Closed-loop issues, *Automatica* **31**(12): 1751–1770.
- Von Sperling, M. (1995). *Princípios do Tratamento Biológico de Águas Residuárias: Introdução à qualidade das águas e ao tratamento de esgotos*, Vol. 1, 1a edn, Departamento de Engenharia Sanitária e Ambiental, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.
- Von Sperling, M. (1996). *Princípios do Tratamento Biológico de Águas Residuárias: Princípios básicos do tratamento de esgotos*, Vol. 2, 1a edn, Departamento de Engenharia Sanitária e Ambiental, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.

- Von Sperling, M. (2000). *Princípios do Tratamento Biológico de Águas Residuárias: Lodos Ativados*, Vol. 4, 2a edn, Departamento de Engenharia Sanitária e Ambiental, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.
- Vrečko, D., Hvala, N. e Carlsson, B. (2003). Feedforward-feedback control of an activated sludge process: a simulation study, *Water Science and Technology* **47**(12): 19–26.
- Vrečko, D., Hvala, N. e Kocijan, J. (2002). Wastewater treatment benchmark: What can be achieved with simple control?, *Water Science and Technology* **45**(4-5): 127–134.
- Wahlberg, B. e Ljung, L. (1986). Design variables for bias distribution in transfer function estimation, *IEEE Transactions on Automatic Control* **31**(2): 134–144.
- Weijers, S. R. (2000). *Modelling, Identification and Control of Activated Sludge Plants for Nitrogen Removal*, PhD thesis, Technische Universiteit Eindhoven, Eindhoven, The Netherlands.
- Wellstead, P. E. e Zarrop, M. B. (1991). *Self-tuning Systems, Control and Signal Processing*, John Wiley and Sons Ltda.
- Yarmolik, V. N. e Demidenko, S. N. (1988). *Generation and application of pseudorandom sequences for random testing*, John Wiley and Sons Ltda.
- Yuan, Z., Keller, J. e Lant, P. (1999). Optimization and control of nitrogen removal, activated sludge processes: a review of recent developments, *9th European Conference on Biotechnology*, Brussels, Belgium.
- Yuan, Z., Oehmen, A. e Ingildsen, P. (2002). Control of nitrate recirculation flow in predenitrification systems, *Water Science and Technology* **45**(4-5): 29–36.
- Zhu, Y. (1998). Multivariable process identification for MPC: the asymptotic method and its application, *Journal of Process Control* **8**(2): 101–115.
- Zierler, N. (1959). Linear recurring sequences, *Journal of the Society in Industrial Application and Mathematics* **7**(1): 31–48.

Apêndice A

Indicadores para Caracterização de Afluentes em ETE's

Para a caracterização dos esgotos são utilizados indicadores de poluição determinados através de parâmetros globais. A utilização destes parâmetros serve como base para a definição da medida do grau de poluição comum a todos os componentes da mistura em que esgoto está constituído. São apresentados segundo sua função como parâmetros indicadores da matéria orgânica e dos nutrientes (Von Sperling, 1995).

A.1 Matéria Orgânica

Os métodos globais para medida indireta do teor de poluente, em termos de matéria orgânica, de um dado despejo são:

1. Demanda Química de Oxigênio (DQO): é amplamente usada para indicar o teor de matéria orgânica presente no esgoto. Corresponde à quantidade de oxigênio necessária para oxidar quimicamente uma dada substância através de um forte oxidante em meio ácido. O teste não é afetado pela nitrificação, dando uma indicação da oxidação apenas da matéria orgânica carbonácea (e não a nitrogenada). Se expressa em mg/l, podendo variar entre 450 mg/l e 700 mg/l nos esgotos domésticos.
2. Demanda Bioquímica de Oxigênio (DBO): é um parâmetro que fornece a biodegradabilidade de um efluente e corresponde à quantidade de oxigênio consumida pelos microrganismos aeróbios, presentes ou introduzidos na amostra, para realizar a biodegradação das substâncias biodegradáveis presentes no despejo. Retrata a quantidade de oxigênio requerida para estabilizar, através de processos bioquímicos, a matéria orgânica carbonácea biodegradável. No teste da DBO são oxidadas tanto a fração biodegradável quanto a fração inerte do despejo, podendo superestimar o consumo de oxigênio. Se expressa em mg/l, pode variar entre 250 mg/l e 350 mg/l nos esgotos domésticos.
3. Carbono Orgânico Total (COT): neste teste a matéria orgânica é totalmente oxidada, medindo todo o carbono liberado na forma de CO_2 . Para garantir que o carbono a medir seja realmente o carbono orgânico, as formas inorgânicas de carbono (como CO_2 , HCO_3^-) devem ser

removidas antes da análise ou corrigidas posteriormente. Também é expressa em unidades de concentração, mg/l, sendo que nos esgotos domésticos esta concentração varia de 170 mg/l a 350 mg/l.

A.2 Nutrientes

Outros parâmetros importantes na caracterização dos esgotos são os relacionados aos nutrientes, como o nitrogênio e o fósforo.

A.2.1 Nitrogênio

O nitrogênio é indispensável para o crescimento dos microrganismos mas também é o responsável pelo fenômeno de eutrofização em lagoas e represas. É caracterizado pelo Nitrogênio Total, que inclui o nitrogênio orgânico, amônia, nitrito e nitrato. O nitrogênio orgânico e a amônia compreendem o denominado Nitrogênio Total Kjeldahl (NTK).

1. Nitrogênio orgânico: nitrogênio na forma de proteínas, aminoácidos e uréia. A concentração típica nos esgotos domésticos está entre 15 a 30 mgN/l.
2. Amônia: produzida como primeiro estágio da decomposição do nitrogênio orgânico, está presente nos esgotos na forma livre (NH_3) ou na forma ionizada (NH_4^+). Pode-se esperar uma concentração de 20 a 40 mgNH₃-N/l.
3. Nitrito (NO_2^-): Estágio intermediário da oxidação da amônia, praticamente ausente no esgoto doméstico bruto. A concentração em mgNO₂-N/l é aproximadamente nula.
4. Nitrato (NO_3^-): Produto final da oxidação da amônia. A concentração típica em esgotos domésticos está entre 0 a 2 mgNO₃-N/l.

A.2.2 Fósforo

O fósforo é um nutriente essencial para o crescimento dos microrganismos responsáveis pela estabilização da matéria orgânica, mas também o excesso deste nutriente pode provocar a eutrofização das águas em lagoas e represas. O fósforo total existe na forma orgânica e inorgânica. Na forma inorgânica apresenta-se como ortofosfatos e polifosfatos.

A.3 Outros Parâmetros

A concentração de substâncias classificadas como poluentes são especificadas também mediante parâmetros físicos e químicos, estipuladas nos padrões de lançamento. Os principais critérios de qualidade de despejos são

1. **Concentração de Oxigênio Dissolvido:** é o principal parâmetro que caracteriza os efeitos de poluição nos despejos orgânicos. É expressa em unidades de concentração mg/l, sendo usual uma concentração de 2 mg/l no reator biológico de plantas de lodos ativados para manter uma população adequada de microorganismos.
2. **pH:** representa a concentração de íons hidrogênio, dando uma indicação sobre a condição de acidez ou alcalinidade da água.
3. **Temperatura:** Aumento de temperatura do despejo coletado implica em um maior consumo de oxigênio dissolvido no corpo receptor, em vista da maior atividade química e bioquímica, prejudicando a eficiência do processo de remoção de matéria orgânica.

Apêndice B

Definições importantes para o experimento de identificação

Nesta seção são fornecidos conceitos relacionados à geração de sinais como sendo parte importante do experimento de identificação.

B.1 Estabilidade de uma função de transferência

Definição B.1.1. Uma função de transferência $G(q)$ (ou o “sistema G” ou o “filtro G”) é dita *estável* se

$$G(q) = \sum_{k=1}^{\infty} g(k)q^{-k}, \quad \sum_{k=1}^{\infty} |g(k)| < \infty \quad (\text{B.1})$$

Esta definição coincide com a definição teórica de estabilidade BIBO (*bounded-input, bounded-output*), o seja, uma entrada $u(t)$ em $G(q)$ está sujeita a $|u(t)| \leq C$, então a saída correspondente $z(t) = G(q)u(t)$ será também limitada, $|z(t)| \leq C'$.

Definição B.1.2. Uma família de filtros $G_{\alpha}(q)$ é dita *uniformemente estável* se

$$|g_{\alpha}(k)| \leq g(k), \quad \forall \alpha \in \mathcal{A}, \quad \sum_{k=1}^{\infty} g(k) < \infty \quad (\text{B.2})$$

B.2 Funções de autocorrelação e de potência de um sinal

Definição B.2.1. A *função de autocorrelação discreta* de um sinal fracamente estacionário (e.g. cujo valor médio não varia com o tempo), é definida por:

$$R_{uu}(k) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N u(t) \cdot u(t-k) \quad (\text{B.3})$$

similarmente, a *função de correlação cruzada discreta* entre dois sinais $u(t)$ e $y(t)$ com a mesma propriedade é definida por:

$$R_{uy}(k) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N u(t) \cdot y(t - k) \quad (\text{B.4})$$

Definição B.2.2. A *função de densidade do espectro de potência* (potência) do sinal $u(t)$ define-se como:

$$\phi_u(w) = \sum_{k=-\infty}^{\infty} R_{uu}(k) e^{-jw} \quad (\text{B.5})$$

B.3 Experimentos informativos

O conceito de experimentos informativos deriva da definição de sinais persistentemente excitantes, mostrado a seguir.

B.3.1 Persistência de excitação

A definição de persistência da excitação de um sinal permite quantificar o quão ativo é um sinal para provocar mudanças tais que garantam que as dinâmicas do processo sejam determinadas a partir do experimento e assim garantir que este sinal seja adequado para identificação de sistemas. Portanto, uma definição formal neste sentido pode ser enunciada em termos da função covariância:

Definição B.3.1. Um sinal $u(t)$ é persistentemente excitante de ordem n se as seguintes três condições se cumprem (Godfrey (1993) *apud* Norton (1986))

- A média \bar{u} exista, onde

$$\bar{u} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N u(t) \quad (\text{B.6})$$

- A função autocovariância¹ exista, onde

$$C_{uu}(k) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N (u(t) - \bar{u})(u(t + k) - \bar{u}) \quad (\text{B.7})$$

- A matriz de autocovariância com elementos i, j dados por $C_{uu}(i - j)$, para $i = 1, \dots, n; j =$

¹Note que $R_{uu}(k) = C_{uu}(k) + \bar{u}^2$ para um sinal $u(t)$ fracamente estacionário e ergódico.

$1, \dots, n$ dada por

$$\bar{C}_{uu}^n = \begin{bmatrix} C_{uu}(0) & C_{uu}(1) & \dots & C_{uu}(n-1) \\ C_{uu}(1) & C_{uu}(0) & \dots & C_{uu}(n-2) \\ \vdots & \vdots & \dots & \vdots \\ C_{uu}(n-1) & C_{uu}(n-2) & \dots & C_u(0) \end{bmatrix} \quad (\text{B.8})$$

for positiva definida.

No caso de $u(t)$ ser ergódico, a esperança matemática substitui as médias. Resumindo, o sinal será persistentemente excitante se sua média e função de autocovariância existirem e se a matriz de autocovariância de $u(t)$, $\bar{C}_{uu}^n \in \mathbb{R}^{n \times n}$ for não singular.

É também útil considerar a definição de persistência da excitação no domínio da frequência.

Definição B.3.2. Um sinal $u(t)$, com espectro $\phi_u(w)$ é dito persistentemente excitante de ordem n se, para todo filtro da forma:

$$M_n(q) = m_1(q^{-1}) + \dots + m_n(q^{-1}) \quad (\text{B.9})$$

a relação

$$|M_n(e^{jw})|^2 \phi_u(w) \equiv 0 \quad (\text{B.10})$$

implicar que $M_n(e^{jw}) \equiv 0$.

Isto implica que $u(t)$ é persistentemente excitante de ordem n , se $\phi_u(w)$ for diferente de zero em pelo menos n pontos no intervalo $-\pi \leq w \leq \pi$.

Definição B.3.3. Um sinal $u(t)$ com espectro $\phi_u(w)$ é persistentemente excitante se

$$\phi_u(w) > 0, \quad \text{para quasi toda } w. \quad (\text{B.11})$$

"Para quase todas" as frequências significa que o espectro de potência pode ser zero para um conjunto de medida zero (em um número contável de pontos).

Corolário B.3.1. Um experimento em malha aberta é informativo se a entrada é persistentemente excitante.

Apêndice C

Algoritmo de Identificação Não Linear Multivariável

C.1 Representação do processo

Baseado na definição do modelo NARMAX (*Non linear AutoRegressive with Moving Average and eXogeneous input*) de Chen e Billings (1989), pode-se obter o modelo NARX multivariável, para m saídas, r entradas e s variáveis de distúrbios medidos:

$$\mathbf{y}(t) = \mathbf{b} + \mathbf{f}^\ell[\mathbf{y}(t-1), \dots, \mathbf{y}(t-N_y), \mathbf{u}(t-d-1), \dots, \mathbf{u}(t-d-N_u), \mathbf{d}(t-1), \dots, \mathbf{d}(t-N_d)] + \mathbf{e}(t) \quad (\text{C.1})$$

mediante a não incorporação dos erros passados ou dos resíduos do processo de identificação $\mathbf{e}(t)$ em \mathbf{f}^ℓ ; sendo os vetores de saída, entrada, distúrbio medido, de polarização e ruído aditivo do sistema para um instante de amostragem definidos como:

$$\mathbf{y}(t) = \begin{bmatrix} y_1(t) \\ \dots \\ y_m(t) \end{bmatrix}, \quad \mathbf{u}(t) = \begin{bmatrix} u_1(t) \\ \dots \\ u_r(t) \end{bmatrix}, \quad \mathbf{d}(t) = \begin{bmatrix} d_1(t) \\ \vdots \\ d_s(t) \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \quad (\text{C.2})$$

com N_y , N_u e N_d atrasos máximos da saída, entrada e distúrbio medido, respectivamente. \mathbf{f} é uma função não linear com grau de não linearidade ℓ . O modelo NARX caracteriza cada saída do processo separada e independentemente. Note na relação (C.1) que $\mathbf{y}(t)$ é considerado dependente da entrada atrasada de um número fixo de amostras, d .

C.1.1 Aproximação polinomial

O vetor função não linear $\mathbf{f}^\ell = [f_1, \dots, f_m]^T$ pode ser aproximado mediante funções polinomiais (Chen e Billings, 1989). Deste modo, o número total de termos é gerado a partir da multiplicação dos seguintes regressores lineares:

$$\begin{aligned}
x_1(t) &= 1 \\
x_2(t) &= y_1(t-1) \\
\vdots &= \vdots \\
x_{m.N_y+1}(t) &= y_m(t-N_y) \\
x_{m.N_y+2}(t) &= u_1(t) \\
\vdots &= \vdots \\
x_{m.N_y+r.N_u+1}(t) &= u_r(t-N_u) \\
x_{m.N_y+r.N_u+2}(t) &= d_1(t-1) \\
\vdots &= \vdots \\
x_{m.N_y+r.N_u+s.N_d+1}(t) &= d_s(t-N_d)
\end{aligned} \tag{C.3}$$

O número total de regressores lineares pode ser calculado pela seguinte equação:

$$n = m.N_y + r.N_u + s.N_d + 1 = M + 1 \tag{C.4}$$

onde esses n regressores lineares são um subconjunto do modelo polinomial não linear e correspondem ao subconjunto de $\mathbf{f}^\ell = [f_1, \dots, f_m]^T$, quando o grau de não linearidade é $\ell = 1$, que equivale ao modelo NARX estendido linear. M representa o número de termos regressores sem incluir o termo de polarização. Cada termo adicional de saída na aproximação polinomial consegue-se com a multiplicação destes n -elementos básicos, sem incluir o termo $x_1(t)$, até o grau polinomial ℓ .

Finalmente, a aproximação polinomial pode ser escrita para $i = 1, \dots, m$ saídas na seguinte forma compacta:

$$\begin{aligned}
y_i(t) = & \theta_0^{(i)} + \sum_{j_1=2}^n \theta_{j_1}^{(i)} x_{j_1}(t) + \sum_{j_1=2}^n \sum_{j_2=j_1}^n \theta_{j_1 j_2}^{(i)} x_{j_1}(t) x_{j_2}(t) + \dots + \\
& \sum_{j_1=2}^n \dots \sum_{j_\ell=j_{\ell-1}}^n \theta_{j_1 \dots j_\ell}^{(i)} x_{j_1}(t) \dots x_{j_\ell}(t) + e_i(t) \quad , \forall i = 1, \dots, m
\end{aligned} \tag{C.5}$$

Note que os parâmetros a serem estimados para o ajuste da estrutura são lineares.

C.1.2 Conjunto de termos candidatos

A união de todos os possíveis termos em um modelo polinomial é denominado conjunto de termos candidatos e pode ser calculado para um determinado grau de não linearidade segundo:

$$\begin{aligned}
 \ell = 1 : n &= m \cdot N_y + r \cdot N_u + s \cdot N_d + 1 = M + 1 \quad (n = n_\theta \text{ no caso linear}) \\
 \ell = 2 : n_\theta &= m \cdot N_y + r \cdot N_u + s \cdot N_d + \sum_{i=1}^n i + 1 = M + 1 \\
 \ell = 3 : n_\theta &= m \cdot N_y + r \cdot N_u + s \cdot N_d + \sum_{i=1}^n i + \sum_{j=1}^n \sum_{i=1}^j i + 1 = M + 1
 \end{aligned} \tag{C.6}$$

onde n_θ representa o conjunto de termos candidatos. Como a estrutura do modelo não é conhecida, tem que se considerar o conjunto completo de termos candidatos durante todo o processo de identificação. A informação que precisa ser estimada *a priori* são os atrasos máximos e o grau de não linearidade esperado.

C.1.3 Redução do modelo

A redução do modelo é um passo importante na obtenção do modelo parcimonioso. Pode-se escrever a equação C.5 em forma de regressão linear matricial, tal que permita combinar os procedimentos de detecção de estrutura com a estimação de parâmetros:

$$\mathbf{Z}^i = \mathbf{P} \times \Theta^i + \Xi^i \quad , \forall i = 1, \dots, m \tag{C.7}$$

sendo

- \mathbf{Z}^i = vetor de variável dependente relativa à i -ésima saída.
- \mathbf{P} = matriz de regressores.
- Θ^i = vetor de parâmetros relativa a i -ésima saída.
- Ξ^i = vetor erro de modelagem relativa a i -ésima saída.

A matriz de regressores será a mesma para todas as saídas. No início do processo de identificação tem que ser considerada a matriz de regressores completa. Para $(N + 1)$ dados entrada-saída, desde o início do experimento de identificação *off-line*, instante $t - N$, até o instante atual, t e para n_θ possíveis regressores até o grau ℓ , a matriz de regressores é dada por:

$$\mathbf{P} = \begin{bmatrix} 1 & x_2(t-N) & \dots & x_n(t-N) & x_2^2(t-N) & x_2(t-N)x_3(t-N) & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_2(t) & \dots & x_n(t) & x_2^2(t) & x_2(t)x_3(t) & \dots \\ x_2(t-N)x_n(t-N) & \dots & x_2^\ell(t-N) & \dots & x_2^{\ell-1}(t-N)x_n(t-N) & \dots & x_n^\ell(t-N) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_2(t)x_3(t) & \dots & x_2^\ell(t) & \dots & x_2^{\ell-1}(t)x_n(t) & \dots & x_n^\ell(t-N) \end{bmatrix} \tag{C.8}$$

\mathbf{P} pode ser representada abreviadamente por $\mathbf{P} = [p_1 \dots p_{n_\theta}]$. Note que $p_1 = 1$. O vetor de variáveis dependentes, o vetor de parâmetros e o vetor erro de modelagem são denotados como segue:

$$\mathbf{Z}^i = \begin{bmatrix} y_i(t-N) \\ \dots \\ y_i(t) \end{bmatrix}, \quad \Theta^i = \begin{bmatrix} \theta_1^i \\ \dots \\ \theta_{n_\theta}^i \end{bmatrix}, \quad \Xi^i = \begin{bmatrix} \xi_i(t-N) \\ \dots \\ \xi_i(t) \end{bmatrix} \quad (\text{C.9})$$

onde: \mathbf{P} tem dimensão $(N+1) \times n_\theta$, \mathbf{Z} tem dimensões $(N+1) \times 1$, Θ tem dimensões $n_\theta \times 1$ e Ξ tem dimensões $(N+1) \times 1$.

C.2 Estimação de parâmetros

Os modelos NARX são estruturas lineares em seus parâmetros, portanto o problema de estimação de parâmetros pode ser resolvido mediante algum método baseado nos Mínimos Quadrados Lineares (MQL). Uma classe de algoritmos que combina eficientemente a seleção da estrutura mínima com a estimação dos parâmetros para modelos polinomiais NARX é a dos algoritmos ortogonais. Chen e Billings (1989) demonstraram que dentre os métodos de decomposição ortogonal, o Método de Gram-Schmidt Modificado é o mais apropriado com relação a outros métodos (e.g. Gram-Schmidt clássico ou a transformação de Householder).

C.2.1 Método de Gram-Schmidt Modificado

O problema MQL corresponde a calcular o vetor de parâmetros $\hat{\Theta}$ tal que minimiza a norma euclidiana $\|\mathbf{Z} - \mathbf{P} \times \Theta\|$. A solução deste problema deve satisfazer a equação normal:

$$\mathbf{P}^T \times \mathbf{P} \times \hat{\Theta} = \mathbf{P}^T \times \mathbf{Z} \quad (\text{C.10})$$

Note que foi omitido o índice i que denota a i -saída, para facilitar a notação. \mathbf{P} representa o conjunto total de termos e o objetivo é calcular um subconjunto \mathbf{P}_s que corresponda ao modelo mínimo definido anteriormente. É importante ressaltar que é recomendável não efetuar o produto $\mathbf{P}^T \mathbf{P}$ na solução do problema MQL para evitar mal-condicionamento numérico. É sempre melhor usar uma decomposição de \mathbf{P} .

A decomposição é obtida através da ortogonalização da matriz de regressores em (C.7), resultando no seguinte modelo ortogonal:

$$\mathbf{Z} = \mathbf{W} \times \mathbf{g} + \Xi \quad (\text{C.11})$$

onde a matriz ortogonal de regressores, \mathbf{W} e o vetor auxiliar de parâmetros, \mathbf{g} , são definidos como

$$\mathbf{W} = \mathbf{P} \times \mathbf{A}^{-1} \quad (\text{C.12})$$

$$\mathbf{g} = \mathbf{A} \times \Theta \quad (\text{C.13})$$

\mathbf{W} é obtida da decomposição de Cholesky $\mathbf{P} = \mathbf{W} \times \mathbf{A}$, sendo \mathbf{A} uma matriz triangular superior de dimensão $n_\theta \times n_\theta$ de diagonal unitária. Note que agora as colunas de \mathbf{W} constituem os novos regressores ortogonais do problema e \mathbf{g} será o vetor de parâmetros para este conjunto de regressores.

Para o cálculo de \mathbf{g} , o Método de Gram-Schmidt Modificado calcula uma fila por vez e ortogonaliza \mathbf{P} da seguinte forma: no k -ésimo estágio faz que as colunas sub-escritas $k+1, \dots, n_\theta$ de \mathbf{A} sejam ortogonais à k -ésima coluna de \mathbf{P} e repete a operação para $k = 1, \dots, n_\theta - 1$. Usa-se para o estágio inicial a seguinte notação:

$$p_i^0 = p_i \quad \forall i = 1, \dots, n_\theta$$

para realizar os seguintes passos:

$$\left. \begin{aligned} w_k &= p_k^{(k-1)} \\ \alpha_{k,i} &= \frac{\langle w_k, p_i^{(k-1)} \rangle}{\langle w_k, w_k \rangle}, \quad i = k+1, \dots, n_\theta \\ p_i^{(k)} &= p_i^{(k-1)} - \alpha_{k,i} \cdot w_k, \quad i = k+1, \dots, n_\theta \\ w_{n_\theta} &= p_{n_\theta}^{(n_\theta-1)} \end{aligned} \right\} \quad k = 1, \dots, (n_\theta - 1) \quad (\text{C.14})$$

Finalmente os elementos de \mathbf{g} são calculados mediante a transformação de \mathbf{Z} :

$$\left. \begin{aligned} g_k &= \frac{\langle w_k, z^{(k-1)} \rangle}{\langle w_k, w_k \rangle} \\ z^{(k)} &= z^{(k-1)} - g_k \cdot w_k \end{aligned} \right\} \quad k = 1, \dots, n_\theta \quad (\text{C.15})$$

A relação entre o vetor de parâmetros estimados do modelo NARX, $\hat{\Theta}$, com o vetor de parâmetros estimados auxiliar, $\hat{\mathbf{g}}$, é obtida da equação (C.13):

$$\hat{\mathbf{g}} = \mathbf{A} \times \hat{\Theta} \quad (\text{C.16})$$

resultando

$$\hat{\Theta} = \hat{\mathbf{g}} - (\mathbf{A} - \mathbf{I}) \times \hat{\Theta} \quad (\text{C.17})$$

Para o cálculo de $\hat{\Theta}$, temos:

$$\hat{\Theta}_{n_\theta} = \hat{\mathbf{g}}_{n_\theta}, \quad \text{se } i = n_\theta \quad (\text{C.18})$$

Substituindo recursivamente, calcula-se os demais componentes do vetor de parâmetros:

$$\hat{\Theta}_i = \hat{g}_i - \sum_{k=i+1}^{n_\theta} \alpha_{i,k} \Theta_k, \quad i = (n_\theta - 1), \dots, 1 \quad (\text{C.19})$$

C.3 Detecção da estrutura

Como mencionado na Seção C.2, é necessário aplicar múltiplos procedimentos de seleção para obtenção de modelos parcimoniosos, procedimentos estes denominados de detecção de estrutura. Assim, só uns poucos termos poderiam capturar a dinâmica do processo, sendo os restantes não considerados, com uma pequena deterioração na predição (Chen et al., 1989).

C.3.1 Taxa de redução do erro

A taxa de redução do erro (ERR), é definida como o aporte na variância da variável dependente \mathbf{Z} , devido a inclusão de um novo termo na matriz de regressores auxiliares:

$$\text{ERR}_i = \frac{g_i^2 \langle w_i \cdot w_i \rangle}{\langle \mathbf{Z} \cdot \mathbf{Z} \rangle}, \quad i = 1, \dots, n_\theta \quad (\text{C.20})$$

Desta forma, termos não significativos (espúrios) podem ser removidos, definindo um valor de ERR abaixo do qual esses termos não têm contribuição significativa no erro quadrático médio.

C.3.2 Critério de parada

Os seguintes critérios de parada correspondem a um compromisso entre desempenho e complexidade do modelo, considerando o termo de variância dos resíduos (erros de predição) e um termo relacionado à ordem do modelo:

1. Critério de Informação de Akaike (AIC):

$$\text{AIC}(\phi) = N \log C(\hat{\Theta}_s) + M_s \phi \quad (\text{C.21})$$

onde

$$C(\hat{\Theta}_s) = \frac{1}{N} \langle \hat{\mathbf{\Xi}} \cdot \hat{\mathbf{\Xi}} \rangle \quad (\text{C.22})$$

é a variância dos resíduos e ϕ é o valor crítico da distribuição *chi*-quadrado com um grau de liberdade para um grau de significância desejado. Se um valor mínimo de AIC(4) é atingido, o procedimento de detecção de estrutura termina.

2. Critério de Informação de Bayes (BIC):

$$\text{BIC} = N \log C(\hat{\Theta}_s) + (\phi + 1) \log(N) \quad (\text{C.23})$$

é usado no caso em que AIC(4) sobre-estima o vetor de parâmetros verdadeiro.

Ambos critérios auxiliam à ERR no procedimento de detecção de estrutura.

C.4 Algoritmo combinado de estimação de parâmetros e detecção de estrutura

O algoritmo mostrado a seguir, reúne os detalhes do procedimento explicado nas seções anteriores e é baseado no algoritmo de ortogonalização Gram-Schmidt Modificado (MGS), adaptado do trabalho original de (Pröll, 1993).

1° Estágio: $k = 1$.

1. Inicializar $k = 1$, $w_1 = p_1$.
2. Copiar todas as colunas de \mathbf{P} , (p_j) até $p_j^{(k-1)}$; $\forall j = k, \dots, n_\theta$
3. Copiar o vetor observação Z até $Z^{(k-1)}$.
4. Calcular $\hat{g}_k^j = \frac{\langle p_j^{(k-1)}, Z^{(k-1)} \rangle}{\langle p_j^{(k-1)}, p_j^{(k-1)} \rangle}$; $\forall j = k, \dots, n_\theta$.
5. Calcular $[ERR]_k^j = \frac{(\hat{g}_k^j)^2 \langle p_j^{(k-1)}, p_j^{(k-1)} \rangle}{\langle Z, Z \rangle}$; $\forall j = k, \dots, n_\theta$.
6. Obter $[ERR]_k^L = \max\{ERR_k^j, k \leq j \leq n_\theta\}$.
7. Permutar a L -ésima coluna $\mathbf{P}^{(k-1)}$ com a k -ésima coluna de $p_k^{(k-1)}$.
8. Copiar a k -ésima coluna de $\mathbf{P}^{(k-1)}$ ($p_k^{(k-1)}$) sobre w_k .
9. Calcular $\alpha_{k,j} = \frac{\langle w_k, p_j^{(k-1)} \rangle}{\langle w_k, w_k \rangle}$; $\forall j = k + 1, \dots, n_\theta$.
10. Calcular $p_j^{(k)} = p_j^{(k-1)} - \alpha_{k,j} w_k$; $\forall j = k + 1, \dots, n_\theta$.
11. Calcular $g_k = \frac{\langle w_k, Z^{(k-1)} \rangle}{\langle w_k, w_k \rangle}$.
12. Calcular $Z^{(k)} = Z^{(k-1)} - g_k w_k$.
13. Calcular $\hat{\Xi} = Z - \sum_{j=1}^k g_j w_j$.
14. Calcular $C(\hat{\Theta}_s) = \frac{1}{N} \langle \hat{\Xi}, \hat{\Xi} \rangle$.
15. Calcular $AIC(4)_k = N \log C(\hat{\Theta}_s) + 4(k + 1)$ e $BIC_k = N \log C(\hat{\Theta}_s) + (k + 1) \log(N)$.
16. Iniciar para $k = 2$.

2° Estágio: $\forall k \geq 2, \dots, n_\theta - 1$

1. Repetir os passos 4, 5, 6, 7, 8, 9 e 10 do 1° estágio.
2. Permutar os elementos da L -ésima coluna, $\alpha_{l,L}$, com os elementos da k -ésima coluna, $\alpha_{l,k}$, $\forall l = 1, \dots, (k - 2)$.
3. Repetir os passos 11, 12, 13, 14 e 15 do 1° estágio.
4. Se $AIC(4)_k \geq AIC(4)_{(k-1)}$ ou $BIC_k \geq BIC_{(k-1)}$ então executar o 3° estágio.

5. Se $AIC(4)_k < AIC(4)_{(k-1)}$ ou $BIC_k < BIC_{(k-1)}$ então executar o 2º estágio e iniciar para $k = k + 1$.

3º Estágio: Para $AIC(4)_k \geq AIC(4)_{(k-1)}$ ou $BIC_k \geq BIC_{(k-1)}$

1. Iniciar para $M_s = k - 1$.
2. Calcular $\hat{\theta}_{M_s} = g_{M_s}$.
3. Calcular $\hat{\theta}_j = g_j - \sum_{l=j+1}^{M_s} \alpha_{j,l} \hat{\theta}_l; \quad \forall j = M_s - 1, \dots, 1$.

□

A seguir são dados alguns comentários importantes para a implementação do algoritmo.

Comentário C.4.1. Sobre o 1º estágio

- Os passos 1 e 3 inicializam o procedimento mediante a inicialização do contador em $k = 1$.
- Os passos 4 e 6 selecionam a coluna L da matriz de regressores $\mathbf{P}^{(k-1)}$, que calcula a soma dos quadrados dos resíduos máximos.
- O passo 8 copia a coluna selecionada dentro da matriz $\mathbf{W} = [w_1, \dots, w_{n_\theta}]$.
- Os passos 9 e 10 ortogonalizam as colunas restantes de $\mathbf{P}^{(k-1)}$. O passo 9 calcula os elementos da matriz triangular superior \mathbf{A}

$$\mathbf{A} = \begin{bmatrix} 1 & \alpha_{1,1} & \cdots & \cdots & \alpha_{2,n_\theta} \\ & 1 & \alpha_{2,2} & \cdots & \alpha_{2,n_\theta} \\ & & \vdots & \vdots & \vdots \\ & & & 1 & \alpha_{n_\theta-1,n_\theta} \\ & & & & 1 \end{bmatrix} \quad (\text{C.24})$$

Isto corresponde à decomposição de Cholesky dos regressores $\mathbf{P} = \mathbf{W} \times \mathbf{A}$.

- Os passos 11 e 14 calculam a variância dos resíduos.
- O passo 15 calcula o valor do critério de informação de Akaike.

Comentário C.4.2. Sobre o 2º estágio

- Este estágio é repetido enquanto for necessário e novos termos serão selecionados até que o $AIC(4)$ ou BIC seja minimizado.
- O passo 2 terá que ser adicionado ao procedimento dado no estágio 1, com a finalidade de permutar as colunas de \mathbf{A} .
- Os passos 4 e 5 verificam se o mínimo de $AIC(4)$ ou BIC é atingido para o número particular de termos selecionados.

Comentário C.4.3. sobre o 3° estágio

- Depois que o mínimo $AIC(4)$ ou BIC for atingido, M_s termos são selecionados para cada saída. Os termos selecionados são identificados para diferentes valores de L a cada k -ésima iteração e armazenados no vetor $\mathbf{L} = [L_{k=1}, \dots, L_{k=M_s}]^T$. Os elementos de \mathbf{L} são os indicadores para os termos selecionados (colunas) de \mathbf{P} .
- Os passos 2 e 3 mostram como estimar o correspondente parâmetro de forma recursiva.

Com respeito ao uso do modelo reduzido para propósito de controle, duas adições são propostas:

1. Para cada saída, colocar termo bias, coluna p_1 , no modelo reduzido.
2. Para cada saída, colocar a entrada mais recente, $u(t - d - 1)$, no modelo reduzido.

A proposta 1 pode ser introduzida no algoritmo fazendo com que ($L_{k=1} = 1$) no passo 6 do primeiro estágio. A proposta 2, fazendo com que ($L_{k=2} = m.N_y + 2, \dots, (L_{k=r} = m.N_y + (k - 1).N_u + 2)$), em lugar de usar o passo 6, isto é, até $k = r$. Para iterações $k > r$, o algoritmo é executado como apresentado inicialmente.

Isto conclui a parte da redução do modelo, que está baseada em identificação *off-line*, a partir de um conjunto inicial de dados. O sistema é representado agora por i -ésimos sub-sistemas mediante um subconjunto de termos candidatos e representados pelos seguintes vetores:

- O vetor indicador $\mathbf{L}^i = [L_{k=1}, \dots, L_{k=M_s}]^T$, que mostra os termos de \mathbf{P} que foram selecionados para cada saída. Com esta informação, o subconjunto \mathbf{P}_s^i pode ser obtido.
- O correspondente vetor de parâmetros estimados $\hat{\Theta}_s^i = [\hat{\theta}_1, \dots, \hat{\theta}_{M_s}]^T$.

Note que o super-índice $(\cdot)^{(k)}$ denota a interação k -ésima e não a potência e que $\langle \cdot \rangle$ denota o produto interno. É importante ressaltar também que a matriz de regressores completa é considerada no início do algoritmo para cada canal de saída, matriz que é a mesma durante todo o procedimento.