

Universidade Federal de Minas Gerais
Programa de Pós-Graduação em Engenharia Elétrica
Centro de Pesquisa e Desenvolvimento em
Engenharia Elétrica

Modelagem Paramétrica De Sistemas Dinâmicos Não-estacionários

Marcelo Machado Duarte

Dissertação submetida à banca examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Minas Gerais, como parte dos requisitos necessários à obtenção do grau de Mestre em Engenharia Elétrica.

Orientadores: Prof. Eduardo Mazoni Andrade Marçal Mendes, Ph.D.
Prof. Luis Antonio Aguirre, Ph.D.

Belo Horizonte, Dezembro de 2006

Dedico este trabalho:

À minha amada esposa Jaqueline, pela paciência, amor, dedicação e confiança. Sem dúvida alguma esse trabalho não seria possível sem seu suporte.

À minha doce filha Ana Luisa que, tendo a mesma idade deste trabalho, por muitas vezes esteve em meu colo enquanto dedilhava algumas destas letras.

A todos os professores e pesquisadores sérios da UFMG, vários outros desse Brasil gigante e ainda outros pelo mundo afora.

AGRADECIMENTOS

Confesso que quase deixei este trabalho sem a seção de agradecimentos por achar pouco provável que eu conseguisse cobrir a todos que, direta ou indiretamente, contribuíram para seu epílogo. Mas ao fim compreendi que esta seção impedirá, num futuro próximo, que eu mesmo me esqueça de todos os colaboradores. Decidi então aceitar o risco de ser parcial!

Peço a paciência dos céticos, mas creio na soberania de meu salvador Cristo Jesus e que certamente Ele, e somente Ele, permitiu-me finalizar este trabalho.

Agradeço também aos meus pais e a todos meus irmãos em Cristo cujas orações e suporte foram fundamentais para mim, para minha esposa Jaqueline e minha filha Ana Luisa.

Agradeço ao colegiado do PPGEE da UFMG pela compreensão e flexibilidade.

Agradeço ao CPH da UFMG pelo espaço e infraestrutura usados por muito tempo.

Agradeço muito ao Professor Eduardo M. A. M. Mendes pela paciente orientação dispensada a mim, pela ajuda inestimável de algumas de suas rotinas computacionais, pelas dicas e sugestões que se tornaram chaves importantes.

Agradeço muito ao Professor Luis A. Aguirre pela inapreciável ajuda de suas dicas, sugestões e toda a crucial atenção dedicada ao meu trabalho.

Agradeço, pelo apoio e ajuda essenciais, aos meus colegas Mara Cristina da Silveira Coelho, Dair José de Oliveira, Bruno Henrique, Bruno Otávio, Gladstone e Edgar Furtado.

Enfim, agradeço a todos os brasileiros pagadores de seus impostos pela oportunidade de concluir mais esta etapa na UFMG.

“e o encheu do espírito de Deus, no tocante à sabedoria, ao entendimento, à ciência e a todo ofício, para inventar obras artísticas, para trabalhar em ouro, em prata e em bronze, em lavramento de pedras para engastar, em entalhadura de madeira, enfim, para trabalhar em toda obra fina. Também lhe dispôs o coração para ensinar a outros; a ele e a Aoliabe, filho de Aisamaque, da tribo de Dã, a estes encheu de sabedoria do coração para exercerem todo ofício, seja de gravador, de desenhista, de bordador em azul, púrpura, carmesim e linho fino, de tecelão, enfim, dos que exercem qualquer ofício e dos que inventam obras artísticas.”

Êxodo 35:31-35.

CONTEÚDO

Resumo	xv
Abstract	xvii
Lista de Figuras	xxi
Lista de Símbolos e Abreviações	xxvi
1 Introdução	1
1.1 Motivação	3
1.2 Objetivos	4
1.3 Estrutura da Dissertação	5
2 Análise de Séries Temporais	7
2.1 Séries Temporais: Definição	7
2.1.1 Componentes de uma Série Temporal	9
2.1.1.1 Sazonalidade	10

2.1.1.2	Tendências	11
2.2	Séries Temporais Não-estacionárias	12
2.2.1	Definição de Estacionariedade	13
2.2.2	Métodos de Detecção de Não-estacionariedade	15
2.3	Métodos de Análise de Séries Temporais Não-estacionárias	19
3	Identificação de Sistemas Dinâmicos	23
3.1	Procedimento para a identificação de sistemas	24
3.2	Experimentação	25
3.3	Detecção de não-linearidades e/ou não-estacionariedades	26
3.4	Representações de Sistemas Lineares	26
3.4.1	Modelos ARMA	27
3.4.2	Modelos Não Autônomos: ARX, ARMAX.	28
3.4.3	Modelos ARIMA	30
3.5	Representações de Sistemas Não-Lineares	32
3.5.1	Modelos NARMAX	33
3.6	Detecção de Estrutura	35
3.6.1	Método de Busca Ortogonal - OS / Taxa de Redução de Erro - ERR	36
3.6.2	Método de Busca Ótima de Parâmetros - OPS	38
3.6.3	O Critério de Informação de Akaike - AIC	44

3.7	Estimação de Parâmetros	44
3.7.1	Validação do Modelo	46
4	Metodologias Para Identificação de Sistemas Dinâmicos Variantes no Tempo	47
4.1	Time-Varying Optimal Parameter Search	48
4.1.1	Funções Base	50
4.1.2	Descrição da Metodologia TV-OPS.	55
4.2	Estimador de Mínimos Quadrados em Sistemas Variantes no Tempo	63
4.2.1	Estimador Recursivo de Mínimos Quadrados com Fator de Esquecimento Variável.	64
4.2.2	Estimador em Batelada de Mínimos Quadrados com Janelas Deslizantes de Tamanho Variável.	71
5	Análise da Não-Estacionariedade	79
5.1	Evidências de Não-Estacionariedade	82
5.1.1	Parâmetros Constantes	83
5.1.2	Parâmetros Aleatórios	84
5.1.3	Parâmetros Pseudo-Aleatórios	84
5.1.4	Parâmetros com Pequenas Variações Bruscas e Suaves	86
5.1.5	Discussão	87
5.2	Dificuldades Intrínsecas às Metodologias no Rastreamento de Parâmetros	88

5.2.1	Caso I: Ruído na Equação	89
5.2.2	Caso II: Escolha das Variáveis de Projeto	91
5.2.3	Caso III: Interferência Mútua entre Parâmetros	92
5.2.4	Discussão	95
5.3	Não-Estacionariedade \times Não-Linearidade	95
5.3.1	Caso I: Sinal de Entrada Pseudo-Aleatório	96
5.3.2	Caso II: Sinal de Entrada Suave e Determinístico	97
5.3.3	Discussão	98
5.4	Conclusões	99
6	Detecção de Mudança na Dinâmica do Processo	101
6.1	Mudança Brusca da Média	102
6.1.1	Caso1: Circuito de Chua	103
6.1.2	Caso2: Circuito de Chua	106
6.2	Tendência	109
6.2.1	Caso: Produção Mensal de Energia Elétrica	110
6.3	Conclusões	113
7	Discussão e Conclusões Finais	115
7.1	Discussão	116
7.2	Conclusões Finais	120

Bibliografia

RESUMO

Neste trabalho, a identificação e análise de sistemas variantes no tempo são investigadas. O problema de não-estacionariedade é visto no contexto de sistemas dinâmicos paramétricos, discretos e variantes no tempo.

Os fundamentos teóricos necessários para o entendimento de estacionariedade e identificação de sistemas são providos. Três metodologias diferentes, usadas para análise de séries temporais observadas de sistemas dinâmicos variantes no tempo, foram investigadas. O primeiro método, TV-OPS, lida com o problema de seleção de estrutura e rastreia as mudanças nos coeficientes de um modelo polinomial expandindo-os em duas funções base diferentes: Walsh e Legendre. Com respeito aos outros dois métodos, RLSVFF e VLSEWBLs, as mudanças na dinâmica são rastreadas por meio da inspeção dos coeficientes do modelo, os quais são calculados pelos métodos de mínimos quadrados regressivo e em batelada, respectivamente. A fim de identificar sistemas lineares e não-lineares, modelos polinomiais foram usados neste trabalho.

Por meio de séries temporais simuladas, a hipótese de que as variações temporais dos parâmetros do modelo possam explicar, de fato, mudanças no regime dinâmico do sistema foi investigada. Como consequência, o instante de observação em que essas mudanças ocorrem foi identificado. Outras questões, tais como sensibilidade ao ruído de medição, escolha de variáveis de projeto e interferência mútua entre parâmetros, todas essas relacionadas aos métodos supracitados, foram também discutidas.

Além de várias conclusões, o presente trabalho procurou propor e discutir diversas recomendações do uso dos métodos estudados. Dentre essas, cita-se a conclusão de que a estrutura correta do modelo é essencial para todos os métodos. Dados reais foram usados para aplicação das principais idéias e métodos.

ABSTRACT

In this work, the identification and analysis of time-varying systems are investigated. The problem of non-stationarity is addressed in the context of time-varying, parametric and discrete dynamical system identification.

The necessary theoretical background for the understanding of system identification and non-stationarity is given. Three different methods used to analyze time series generated by dynamical time-varying systems were studied. The first method, TV-OPS, deals with the problem of structure selection and tracks the changes in the coefficients of a polynomial model by expanding them into two different basis functions: Walsh and Legendre. In the case of the two other methods, RLSVFF and VLSWBLs, the change in dynamics is tracked by inspecting the model coefficients calculated using the recursive and ordinary least squares methods respectively. Linear and nonlinear systems, polynomial models were used in this work as a basis for system identification.

Using simulated time series, the hypothesis that changes in the model coefficients could explain the dynamical changes in the system behavior was investigated. As a by-product, the location in time where these changes occurred was found. Issues such as sensitivity to measurement noise, choice of free-parameters in the aforementioned methods and the effect of the parameters in each other were also addressed.

The caveats of the above methods were addressed. Several conclusions on how to analyze the results of applying such methods were drawn. One of them is that the correct model structure is essential to all methods. The main ideas and methods were also applied to real data.

LISTA DE FIGURAS

2.1	Exemplos de Séries Temporais	8
3.1	Exemplo de projeção ortogonal num espaço vetorial euclidiano	39
4.1	Exemplo de funções polinomiais de Legendre	51
4.2	Exemplos de funções de Walsh	55
4.3	Gráficos de projeção de distância referente ao modelo da equação (4.26) . .	61
4.4	Gráficos do critério de Informação de Akaike Modificado, com relação ao modelo da equação (4.26)	62
4.5	Parâmetros variantes no tempo estimados para o modelo da equação (4.26), usando-se funções de Walsh e Legendre	63
4.6	Parâmetros variantes no tempo estimados para o modelo da equação (4.26), usando-se somente funções de Walsh ou Legendre	63
4.7	Exemplo de estimação de parâmetros por meio da metodologia RLSVFF. .	71
4.8	Parâmetros do sistema representado na equação (4.40), estimados por meio da metodologia VLSEWBLIS	78
5.1	Gráficos de simulação de modelos ARX cujos parâmetros foram estimados a partir da série temporal simulada do processo da equação (5.1).	81
5.2	Gráficos do critério de informação de Akaike modificado, com relação ao modelo da equação (5.3) com parâmetros constantes.	83

5.3	Gráficos do critério de informação de Akaike modificado, com relação ao modelo da equação (5.3) com parâmetros aleatórios conforme equação (5.5).	85
5.4	Gráficos do critério de informação de Akaike modificado, com relação ao modelo da equação (5.3) com parâmetros aleatórios conforme equação (5.6).	86
5.5	Gráficos do critério de informação de Akaike modificado, com relação ao modelo da equação (5.3) com parâmetros com pequena variação.	87
5.6	Parâmetros estimados para o modelo ARX definido nas equações (5.11) e (5.10), por meio do método RLSVFF.	90
5.7	Parâmetros estimados para o modelo ARX definido na equação (5.11) por meio do método RLSVFF.	92
5.8	Parâmetros estimados para o modelo ARX definido na equação (5.11) por meio do método RLSVFF.	93
5.9	Parâmetros estimados para o modelo ARX definido na equação (5.11) por meio do método RLSVFF.	93
5.10	Exemplo de interferência entre parâmetros, usando o método RLSVFF. . .	94
5.11	Gráfico que mostra os parâmetros originais e estimados, $a(1,k)$ e $b(1,k)$, do modelo ARX definido na equação (5.15), por meio do método TV-OPS. . .	97
5.12	Gráficos que mostram as variações do erro de predição estendido, $Q(k)$, e do fator de esquecimento variável, $\lambda(k)$, na estimação dos parâmetros do modelo ARX definido na equação (5.15), por meio do método RLSVFF. . .	98
5.13	Gráfico que mostra os parâmetros originais e estimados, $a(1,k)$ e $b(1,k)$, do modelo ARX definido na equação (5.15), por meio do método TV-OPS. . .	99
6.1	Modelagem do processo definido pelas equações (6.1) e (6.2). Parâmetros do modelo 1 da Tabela 6.1 estimados via RLSVFF.	104
6.2	Modelagem do processo definido pelas equações (6.1) e (6.2). Parâmetros do modelo 2 da Tabela 6.1 estimados via RLSVFF.	105
6.3	Série temporal obtida por meio da observação do sinal de tensão v_{C_1} do circuito de Chua.	106
6.4	Modelagem da série de Chua com mudança de média.	107

6.5	Série temporal obtida por meio da observação do sinal de tensão v_{C_1} do circuito de Chua.	108
6.6	Análise da série temporal obtida por meio da observação do sinal de tensão v_{C_1} do circuito de Chua via RLSVFF.	109
6.7	Modelagem do processo ARX não-estacionário definido pelas equações (6.1) e (6.4). Os parâmetros do modelo 2 da Tabela 6.1 foram estimados via TV-OPS.	111
6.8	Série de produção mensal de energia elétrica na Austrália observada entre janeiro de 1956 e agosto de 1995	111
6.9	Análise da série de produção mensal de energia elétrica na Austrália, via RLSVFF.	112
6.10	Análise da série de produção mensal de energia elétrica na Austrália, via RLSVFF, com tendência removida.	112

SIMBOLOGIA E ABREVIACOES

Consideraes Gerais

Neste trabalho, matrizes so indicadas por letras latinas maisculas em itlico, por exemplo W , e por letras gregas maisculas, por exemplo Ψ . Vetores so indicados usando-se letras (gregas e latinas) minsculas e em negrito, por exemplo \mathbf{u} e $\boldsymbol{\theta}$. Escalares so representados por letras gregas minsculas com ou sem subndice, por exemplo β , λ_k , e por letras minsculas do alfabeto latino em itlico, com ou sem argumento, por exemplo t e $a(i,k)$. Excees a essas regras so definidas *in loco*.

Abreviaes

Listam-se a seguir as principais abreviaes empregadas nesta dissertao. No caso de siglas consagradas na literatura internacional, optou-se por manter as mesmas em ingls.

AIC	Critrio de informao de Akaike (<i>Akaike's Information Criterion</i>)
AICm	Critrio de informao de Akaike modificado
ERR	Razo de reduo do erro (<i>Error Reduction Ratio</i>)
EMQ	Estendido de Mnimos Quadrados
MQR	Mnimos Quadrados Regressivo
NARMA	Auto-regressivo de mdia mvel no-linear (<i>Nonlinear Autoregressive Moving Average</i>)
OPS	Busca de Parmetro tima (<i>Optimal Parameter Search</i>)
PRBS	Sinal Pseudo-Aleatrio (<i>Pseudo Random Binary Signal</i>)

RLS	Recursive Least Square
RMSE	Raiz quadrada do erro quadrático médio (<i>Root Mean Square Error</i>)
SNR	Relação Sinal Ruído (<i>Signal-to-Noise Ratio</i>)
VFF	Fator de esquecimento variável (<i>Variable Forgetting Factor</i>)

Símbolos

Listam-se a seguir os principais símbolos usados de forma geral. O uso específico de símbolos será definido *in loco*.

$E[\cdot]$	Esperança matemática;
$y(k - i)$	Regressor de saída;
$u(k - j)$	Regressor de entrada;
$e(k)$	Ruído no instante k ;
n_y	Máximo valor do atraso de um regressor de saída de um modelo;
n_u	Máximo atraso do regressor de entrada;
n_e	Máximo atraso do regressor de ruído;
$n_{y\max}$	Máximo valor de n_y na procura heurística de modelos, tal que $n_y = (1, 2, \dots, n_{y\max})$;
k	Instante de observação (tempo discreto);
$\boldsymbol{\theta}$	Vetor de parâmetros;
$\xi(k)$	Resíduo no instante k ;
g_i	Parâmetros ortogonais;
\mathbf{g}	Vetor de parâmetros ortogonais;
\mathbf{w}_i	Vetor de regressores ortogonais;
W	Matriz de regressores ortogonais;
J	Função custo genérica;
n_t	Número de termos de processo;

Ψ	Matriz de regressores;
$\hat{}$	Valor estimado;
ν	Ruído branco;
$\nu(k-i)$	Regressor de ruído em um modelo NARMAX;
\mathbf{I}	Matriz Identidade;
t	Tempo contínuo;
T_s	Tempo de amostragem;
\ln	Logaritmo na base e ;
\mathbb{N}	espaço dos números naturais: $1, 2, \dots$;
\mathbb{R}	espaço dos números reais;
\mathbb{Z}	espaço dos números inteiros: $\{\dots, -1, 0, 1, \dots\}$;
\mathbb{Z}^+	espaço dos números inteiros não negativos: $\{0, 1, 2, \dots\}$;
(a, b)	Intervalo aberto;
$[a, b]$	Intervalo fechado;
T	transposição de vetores ou matrizes;
$\langle \cdot, \cdot \rangle$	produto interno vetorial;
$ \cdot $	módulo ou valor absoluto;
$\ \cdot\ $	norma euclidiana ou comprimento vetorial.

Introdução

Descrever o comportamento de sistemas físicos naturais – como os sistemas biológicos e do meio ambiente –, de sistemas sociais e econômicos – como os que descrevem relações de consumo e produção – e de sistemas industrializados, por meio de modelos matemáticos tem sido alvo de pesquisadores há tempos. Até hoje essa busca tem sido um interessante desafio.

Um conjunto de símbolos e relações matemáticas que traduz, de alguma forma, um fenômeno em questão ou um problema de situação real é denominado *modelo matemático*. Deseja-se que essas relações matemáticas representem adequadamente sistemas reais, ainda que não se consiga mais do que uma representação aproximada de no máximo algumas características dos sistemas em questão.

Quando se observa algum fenômeno ou sistema e, por meio de equações que descrevem a física do processo, consegue-se explicar essas observações, a esse procedimento se dá o nome de modelagem fenomenológica, ou modelagem baseada na física do processo. A busca por métodos de obtenção de modelos matemáticos a partir de dados observados, e não exclusivamente partindo-se das equações que descrevem a física do processo, tem se intensificado nos últimos anos. Esse último procedimento é conhecido como modelagem empírica ou *identificação de sistemas*. Pode-se citar ainda um conjunto de técnicas chamado de identificação *caixa cinza*, que poderia ser classificado como estando entre a modelagem pela física do processo e a modelagem empírica.

Na última década tem-se verificado mais uma tendência geral que tornará o uso de técnicas de identificação e análise de sistemas desejável e até mesmo necessário em praticamente todas as áreas do conhecimento humano. Essa tendência é a inegável capacidade que hoje se tem de *coletar dados* com informação sobre a dinâmica do sistema que está sendo observado (Aguirre, 2004). Em face do avanço da tecnologia de computadores digitais, considerando-se o aumento regular do poder de processamento e os custos em queda, tem se tornado muito comum o emprego de sistemas de aquisição de dados capazes de

monitorar variáveis e processos reais com taxas de amostragem suficientes para representação do comportamento dinâmico do processo por meio dos dados coletados. Esses dados, quando coletados em instantes de tempo bem definidos, são chamados de *série temporal*.

Métodos clássicos de identificação de sistemas fazem algumas considerações com vistas à simplificação da análise. Ljung (1987) e Aguirre (2004) citam duas importantes considerações: *i*) de linearidade, que supõe que o sistema em processo de modelagem comporta-se de forma aproximadamente linear; *ii*) invariância temporal ou estacionariedade¹, relacionado ao pressuposto de que a lei que governa o regime dinâmico do processo não varia no tempo.

Entretanto, o mundo real se mostra cheio de características não-lineares e/ou não-estacionárias. Nos anos mais recentes, tem se desenvolvido técnicas de identificação de sistemas não-lineares com o objetivo, por exemplo, de se estudar características dinâmicas como oscilações e bifurcações (Aguirre, 2004).

A questão de identificação e análise de sistemas em que se observa uma variação temporal no seu padrão dinâmico constitui-se no interesse central deste trabalho. Doravante esses sistemas serão referidos como *sistemas variantes no tempo*.

Em muitas aplicações reais, a questão da variação temporal do padrão dinâmico (não-estacionariedade) é freqüentemente abordada de maneira um tanto quanto subjetiva (Gomes, 2001; Manuca & Savit, 1996). Em alguns casos, exatamente essa variação temporal se torna o objetivo principal de estudo. Noutros, os instantes de tempo em que essas variações ocorrem é que são perseguidos. Ainda em outros, as variações na dinâmica não são de interesse na análise e, ao contrário, trazem danos à análise desejada.

George Box and Gwilym Jenkins propuseram, em 1970, uma forma estendida do modelo ARMA, que lida com o comportamento não-estacionário da série temporal, denominada modelo ARIMA (*AutoRegressive Integrated Moving Average*) (Box & Jenkins, 1970). Tendo causado grande influência² desde que foi proposto, o método Box-Jenkins para análise de séries temporais é muito usado em econometria com foco na predição. Essa metodologia requer pré-processamento dos dados por meio de transformações e diferenciações até que não haja um padrão óbvio de tendência e sazonalidade. A idéia chave desse método é, portanto, descartar a não-estacionariedade antes de prosseguir com a análise.

¹Definições mais formais para *estacionariedade* de uma série temporal serão dadas no capítulo 2.

²X-12-ARIMA, por exemplo, é um pacote de software para ajustamento sazonal desenvolvido e disponibilizado pela agência de estatística do Estados Unidos (U.S. Census Bureau). <http://www.census.gov/srd/www/x12a/>

Algoritmos de mínimos quadrados regressivos (MQR) foram adaptados na tentativa de rastreamento da dinâmica. O conceito de fator de esquecimento variável (VFF) em algoritmos de MQR foi introduzido por Fortescue e colegas em 1981 num contexto de controle auto-sintonizável, para evitar o estouro da matriz de covariância e conseqüente instabilidade no controle. Posteriormente outros trabalhos sugeriram novas abordagens em torno da mesma idéia, como por exemplo os trabalhos de Cho, Kim & Powers (1991) e Cao & Schartz (1999), na linha linha de fator de esquecimento variável; Fortescue, Kershenbaum & Ydstie (1981) e Ljung & Gunnarsson (1990), com o propósito de modificação da matriz de covariância; e (Belge & Miller, 2000), com algoritmos adaptativos do tipo MQR com janelas deslizantes.

Considerando o contexto da inteligência artificial, citam-se técnicas de Redes Neurais Artificiais (RNA) e Redes Neurais *fuzzy* (Zhang, Patuwo & Hu, 1998; Padmakumari, Mohandas & Thiruvengadam, 1999; Liao & Tsao, 2004).

A análise e o processamento de séries não-estacionárias inclui também métodos de representação tempo-frequência (Pola, Macerata, Emdin & Marchesi, 1996; Gamero, Risk, Sobh, Ramirez & Saul, 1996; Potamianos & Maragos, 2001) e técnicas de análise espectral (Bianchi, Mainardi, Petrucci, Signorini, Mainardi & Cerutti, 1993; Pola et al., 1996; Cesarelli, Bifulco & Bracale, 1997; Steenis, Martens & Tulen, 2003). A análise espectral é fundamental em áreas onde o interesse é principalmente a procura de periodicidade nos dados, como em Meteorologia e Oceanografia, Medicina, Biologia, economia e outras (Morettin & Toloí, 2004).

Dentre as várias abordagens disponíveis para análise de séries temporais não-estacionárias, optou-se pelo uso de técnicas de identificação de sistemas dinâmicos variantes no tempo que usam modelos paramétricos. A variação da dinâmica seria então observada pela variação temporal dos parâmetros. Pode-se entender essa variação nos parâmetros por uma série de razões, tais como operação do sistema em regiões diferentes em que um único modelo linear não seria suficiente, envelhecimento de componentes do sistema, ocorrência de falhas e patologias, entre outros (Aguirre, 2004).

1.1 Motivação

A motivação para o estudo de técnicas de identificação de sistemas variantes no tempo surge do fato de que freqüentemente dois caminhos são adotados: *i*) a hipótese de estacionariedade da série temporal em estudo é assumida, como ocorre em muitas técnicas de análise lineares e não-lineares. Observa-se, entretanto, que sistemas reais sofrem vari-

ações em seu padrão dinâmico; *ii*) os dados coletados (série temporal) são divididos em janelas de tempo, de maneira que essas sub-séries possam ser consideradas estacionárias. Várias técnicas são empregadas na tentativa de se determinar a estacionariedade dessas sub-séries. A questão da relevância estatística está em pauta quando essas técnicas são usadas.

É relevante, portanto, o aprimoramento de métodos que possam lidar diretamente com a variação temporal do regime dinâmico de um sistema. Acredita-se poder contribuir com esse aprimoramento ao avaliar-se vários aspectos abordados por análise de sistemas variantes no tempo feitas por técnicas de modelagem paramétrica.

1.2 Objetivos

Este trabalho abordará o problema da modelagem e análise de séries temporais não-estacionárias por meio das seguintes metodologias:

1. *Time-Varying Optimal Parameter Search, TV-OPS.*
2. *Estimador Recursivo de Mínimos Quadrados com Fator de Esquecimento Variável, RLSVFF.*
3. *Estimador em Batelada de Mínimos Quadrados com Janelas Deslizantes de Tamanho Variável.*

Por meio de séries temporais obtidas via simulação e séries temporais observadas de sistemas reais, essas metodologias serão exercitadas nos seguintes aspectos: *i*) capacidade de rastreamento de parâmetros; *ii*) capacidade de detecção do instante de mudança no padrão dinâmico; *iii*) distinção entre não-estacionariedade e não-linearidade; *iv*) problemas intrínsecos às metodologias e/ou à questão da modelagem de sistemas variantes no tempo propriamente dita.

Deseja-se investigar até que ponto essas metodologias de fato mostram variações em padrões dinâmicos, ou, em outras palavras, até que ponto as não-estacionariedades podem ser caracterizadas por elas.

1.3 Estrutura da Dissertação

O *Capítulo 2* apresenta conceitos básicos a respeito de séries temporais com ênfase na questão da não-estacionariedade.

Fundamentos sobre modelagem e indentificação de sistemas são lembrados no *Capítulo 3*. Ressalta-se a abordagem de técnicas de seleção de estrutura como OPS (Optimal Parameter Search) e ERR (Error Reduction Ratio).

No *Capítulo 4* são apresentadas três metodologias consideradas para estudo de modelagem e identificação de sistemas variantes no tempo. São elas: i) *Time-Varying Optimal Parameter Search*, TVOPS; ii) Estimador Recursivo de Mínimos Quadrados com Fator de Esquecimento Variável, RLSVFF; iii) Estimador em Batelada de Mínimos Quadrados com Janelas Deslizantes de Tamanho Variável, VLSWBLS.

A partir de modelos simulados de sistemas variantes no tempo, o *Capítulo 5* aborda a questão da estacionariedade e não-estacionariedade à luz das metodologias supracitadas.

O *Capítulo 6* traz resultados de análises de séries temporais reais com ênfase na detecção dos instantes de tempo em que ocorrem alterações na dinâmica do processo em estudo.

Uma breve discussão dos resultados encontrados, impressões gerais e conclusões estão disponíveis no *Capítulo 7*.

Análise de Séries Temporais

2.1 Séries Temporais: Definição

Uma série temporal é um conjunto de observações $y(k)$, cada uma registrada num instante de observação específico k (Brockwell & Davis, 1996). Pode-se considerar uma série temporal como uma realização de uma determinada variável de um processo. Todas as séries temporais usadas no presente trabalho são do tipo *discreta* pois as observações foram registradas em intervalos de tempo discretos e iguais. Citam-se como exemplo as seguintes séries temporais: *i*) número mensal de manchas solares (Zuerich) entre 1749 e 1983 (Hyndman, 2004). *ii*) preço médio mensal da carne de boi no Estado de São Paulo entre 1954 e 1995 (Aguirre & Aguirre, 2000). *iii*) séries de variabilidade da frequência cardíaca (Oliveira, 2002). *iv*) produção mensal de energia elétrica na Austrália entre 1956 e 1995 (Hyndman, 2004). Alguns desses exemplos podem ser vistos na Figura 2.1.

Uma característica intrínseca de uma série temporal é que, tipicamente, as observações adjacentes são *dependentes*. A natureza dessa dependência entre as observações da série temporal tem sido objeto de interesse de diversas áreas científicas. Pode-se dizer que a *análise de séries temporais* diz respeito às técnicas, sejam elas matemáticas ou não, para a análise dessa dependência (Box, Jenkins & Reinsel, 1994).

Um relevante propósito da análise de séries temporais é, baseado numa possível dependência entre as observações, aprender ou conhecer características intrínsecas do sistema gerador dessas observações. Uma maneira de se fazer a análise é construir um modelo matemático (ou uma família de modelos) que em alguma medida descreve algumas características desse sistema, como por exemplo seu comportamento dinâmico (Aguirre, 2004). Uma vez que um modelo satisfatório é obtido, ele pode ser usado de várias maneiras dependendo do campo de aplicação.

Dentre os campos científicos que têm feito uso da análise de séries temporais cita-

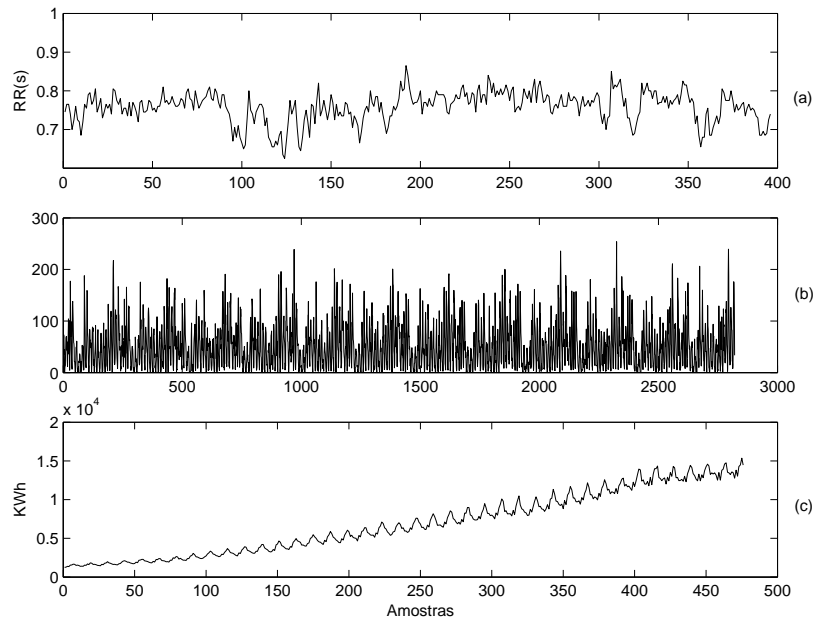


Figura 2.1: (a) Registro de 5 minutos de uma série de variabilidade da frequência cardíaca de um humano com doença de chagas, porém ainda sem os sintomas (Oliveira, 2002). (b) Número mensal de manchas solares entre 1749 e 1983 (Hyndman, 2004). (c) Produção mensal de energia elétrica na Austrália entre janeiro de 1956 a agosto 1995 (Hyndman, 2004).

se: Engenharia, Física, Estatística, Biologia, Sociologia e Economia (Brockwell & Davis, 1996). Modelos matemáticos têm sido utilizados ao longo da história para os mais diversos fins, como por exemplo: entender e explicar fenômenos observados tanto na natureza quanto em sistemas sociais, econômicos, biológicos, equipamentos; projetos de monitorização e controle; predição; estimação de estados; simulação e treinamento (Aguirre, 2004).

Outro objetivo da análise de séries temporais é a predição de valores futuros da série por meio dos modelos matemáticos. A predição é feita baseada na extrapolação de observações passadas da série (Kandil, El-Debeiky & Hasanien, 2001; Morettin & Tolo, 2004). Essas predições podem ser de curto prazo, como para séries de vendas, produção ou estoque, ou longo prazo, como para séries populacionais e de produtividade (Morettin & Tolo, 2004).

Morettin & Tolo (2004) citam ainda outros interesses na análise de série temporal, tais como: *i*) a descrição apenas do comportamento da série. Nesse caso a construção do

gráfico da série, a verificação da existência de tendências, ciclos e variações sazonais, a construção de histogramas e diagramas de dispersão, podem ser ferramentas úteis para tomada de decisão e estudo de caso; *ii*) a busca de periodicidades relevantes nos dados. Nesse caso técnicas de análise espectral podem ser de grande utilidade.

2.1.1 Componentes de uma Série Temporal

Muitas séries temporais exibem variações do tipo tendência e sazonalidade. Baseado em modelos clássicos de decomposição, séries temporais podem ser decompostas nas seguintes componentes não observáveis: *i*) sazonal, *ii*) de tendência, *iii*) cíclica e *iv*) de resíduo (Brockwell & Davis, 1996; Zhang & Qi, 2003).

As relações funcionais específicas entre essas componentes podem assumir formas diferentes. Entretanto, duas possibilidades diretas são consideradas na literatura (Brockwell & Davis, 1996), tais como os modelos com sazonalidade aditiva e multiplicativa:

$$y(k) = T(k) \times C(k) + S(k) + e(k), \quad (2.1)$$

$$y(k) = T(k) \times C(k) \times S(k) \times e(k), \quad (2.2)$$

em que o índice k denota um instante de tempo, y uma série temporal, T a componente de tendência, C a componente cíclica, S a componente sazonal e e representa o resíduo, que está relacionado aos distúrbios que podem ser estocásticos, erros e irregularidades.

Os modelos (2.1) e (2.2) são interessantes quando consideramos a possibilidade de estimar as componentes de tendência, T , e sazonal, S , e construir uma série livre das influências dessas componentes. Há várias razões para empregar-se procedimentos de ajuste de tendência e sazonalidade (Morettin & Tolo, 2004; Zhang & Qi, 2003; Virili & Freisleben, 2000). Faz-se a seguir considerações sobre algumas dessas razões.

Uma das suposições mais freqüentes que se faz a respeito de uma série temporal é a de *estacionariedade* (Morettin & Tolo, 2004), cuja definição formal será feita na seção 2.2. Muitos métodos de modelagem assumem a estacionariedade das séries temporais como pré-requisito de seus procedimentos. Uma das soluções para esse problema é a remoção das componentes de tendência e sazonalidade para se obter uma série de resíduos $e(k)$. Para muitas séries temporais, como séries econômicas e financeiras, o resíduo resultante pode ser considerado estacionário. Prossegue-se então com a escolha de um modelo a partir da série residual. As predições efetuadas por meio desse modelo devem ser atualizadas pela transformação inversa da série, ou seja, acrescentando-se as componentes de tendência e

sazonalidade.

Pierce (1980) faz o seguinte comentário acerca do interesse no ajuste sazonal: “*pode estar relacionado com a idéia de que nossa habilidade em reconhecer, interpretar ou reagir a movimentos importantes não-sazonais numa série (tais como: pontos de mudança e outros eventos cíclicos, novos padrões emergentes, ocorrências não esperadas cujas possíveis causas são investigadas) é perturbada pela presença de movimentos sazonais*”.

Neste trabalho são analisados métodos que, em princípio, são capazes de lidar com séries não-estacionárias com vistas à modelagem e análise. Ainda assim, a literatura tem indicado que a remoção das componentes de tendência e sazonalidade podem melhorar os resultados.

Zhang & Qi (2003) examinam a capacidade das redes neurais no que diz respeito à modelagem e predição usando-se séries temporais com tendência e sazonalidade. Seus resultados indicam que redes neurais não são capazes de modelar a sazonalidade diretamente. Além do ajuste de sazonalidade nos dados, eles também concluíram que a remoção da componente de tendência é de grande relevância para melhorar as predições. Virili & Freisleben (2000) também observaram que a presença de tendência em séries temporais econômicas podem ser um grande obstáculo à geração de predições satisfatórias com redes neurais.

2.1.1.1 Sazonalidade

A sazonalidade é um padrão recorrente e muitas vezes periódico que, para séries econômicas por exemplo, pode ser causada por fatores climáticos, feriados, promoções, comportamento de agentes econômicos, dentre outros. A diferença entre uma componente *cíclica* e uma *sazonal* é que a última ocorre em intervalos regulares, enquanto os fatores cíclicos normalmente têm uma duração maior e varia de ciclo para ciclo (Zhang & Qi, 2003).

Considerando o modelo (2.1), por meio de um procedimento de ajuste sazonal deseja-se estimar $S(k)$ de maneira que a série ajustada seja obtida do seguinte modo:

$$y^{\text{as}}(k) = y(k) - \hat{S}(k), \quad (2.3)$$

em que y^{as} é a série após o ajustamento sazonal e \hat{S} é a estimativa da componente sazonal da série original y .

Existem vários procedimentos para se estimar a componente sazonal de uma série.

Citam-se (Morettin & Toloí, 2004; Zhang & Qi, 2003; Brockwell & Davis, 1996): *i) método de regressão*, que é indicado para séries que apresentam sazonalidade *determinística*, ou seja, uma componente sazonal que não varie com o tempo; *ii) método de médias móveis*, que é indicado para séries cuja componente sazonal varie com o tempo. Essa sazonalidade é chamada *estocástica* pela impossibilidade de predição; *iii) método X-12-ARIMA*, que é o último da família de métodos de ajustes sazonais desenvolvidos pela agência de estatística dos Estados Unidos e do Canadá¹.

2.1.1.2 Tendências

Embora as variações sazonais sejam uma componente muito importante numa série temporal (Zhang & Qi, 2003), a presença de *tendência* nas séries as tornam necessariamente não-estacionárias e causa uma relevante degradação em exercícios de predição (Virili & Freisleben, 2000).

As componentes de tendência e sazonal são, em geral, bastante relacionadas e a influência da tendência sobre a componente sazonal pode ser muito forte, por duas razões Morettin & Toloí (2004): *i) métodos de estimação da componente sazonal podem ser bastante afetados se não levarmos em conta a tendência; ii) a especificação da componente sazonal depende da especificação da componente de tendência*. Essas observações sugerem que não se pode isolar uma das componentes sem tentar isolar a outra.

Partindo da equação (2.3), supondo-se que a componente sazonal tenha sido removida da série original y , temos:

$$y^{\text{at}}(k) = y^{\text{as}}(k) - \hat{T}(k), \quad (2.4)$$

em que y^{at} é a série após a remoção da tendência e \hat{T} é a estimativa da componente de tendência da série y^{as} .

Ao considerar-se uma série com tendência *determinística*², pode-se estimar a componente \hat{T} por meio de ajuste de uma curva polinomial aos valores observados (Zhang & Qi, 2003). Morettin & Toloí (2004) comentam que um grande problema com essa metodologia é que, embora o polinômio possa ajustar-se bem às observações (interpolação), extrapolações (projeções futuras) podem ser bem ruins. Como alternativa, pode-se usar

¹US Census Bureau e Statistics Canada.

²Tendência *determinística* caracteriza uma série cujas flutuações ocorrem ao redor de uma linha fixa de tempo, de maneira que essa linha possa ser aproximada por um polinômio de grau n .

métodos de suavização, tais como: *i*) filtro linear média móvel; *ii*) método *lowess*³, que consiste em se fazer ajustes sucessivos de retas (via mínimos quadrados) a sub-conjuntos de dados.

No caso em que a série exiba uma tendência *estocástica*⁴, pode-se usar o método das *diferenças sucessivas*. A primeira diferença é definida como:

$$\nabla y^{\text{at}}(k) = y^{\text{as}}(k) - y^{\text{as}}(k - 1). \quad (2.5)$$

De modo geral, a n -ésima diferença de y^{at} é (Brockwell & Davis, 1996; Morettin & Tolo, 2004):

$$\nabla^n y^{\text{at}}(k) = \nabla[\nabla^{n-1} y^{\text{as}}(k)]. \quad (2.6)$$

2.2 Séries Temporais Não-estacionárias

Como foi brevemente discutido na seção 2.1, muitas técnicas de modelagem e análise de séries temporais devem assumir algum tipo de *estacionariedade* do sistema gerador dessas séries, ou seja, esses sistemas são considerados invariantes no tempo (Manuca & Savit, 1996; Schreiber, 1997). Mas a maioria dos sistemas reais varia ao longo do tempo, ou seja, a dinâmica que regula a evolução temporal sofre alguma variação. Essa variação pode ser consequência de flutuações de variáveis não medidas que afetam a operação do sistema (Aguirre, 2004).

A variação no tempo das variáveis e/ou parâmetros do sistema leva ao problema de *não-estacionariedade* que é um dos grandes desafios na análise de séries temporais. Considerar o sistema invariante no tempo, ou que ele seja estacionário, em muito simplifica o problema de modelagem, uma vez que a relação causa e efeito não varia significativamente com o tempo e apenas um modelo é requerido para representá-lo (Aguirre, 2004). Embora muitas técnicas de análise de séries temporais lineares e não-lineares assumirem que a série sob investigação é estacionária, muitas dessas séries geradas por processos geofísicos, fisiológicos, econômicos, dentre outros, são de fato não-estacionárias (Gao, 2001).

A não-estacionariedade é uma propriedade quase sempre considerada um efeito indesejável, pois sua presença dificulta a análise do sistema em estudo (Gomes, Souza, Guimarães & Aguirre, 2000). Entretanto, existem situações em que exatamente as não-

³LOcally WEighted regression Scatter plot Smoothing.

⁴Tendência cuja variação é aleatória, impossibilitando seu rastreamento. O processo estocástico *random walk* é um exemplo clássico de um sistema gerador de série com tendência estocástica.

estacionariedades constituem o principal foco de análise por representarem um aspecto relevante da dinâmica e com possíveis significados físicos. Cita-se, como exemplo, os registros de eletro-encefalograma, que freqüentemente são feitos com o propósito de se identificar mudanças no estado dinâmico do cérebro. Essas mudanças podem acontecer, por exemplo, entre diferentes estágios do sono ou entre ataques epiléticos e uma atividade cerebral normal (Schreiber, 1997; Rieke, Mormann, Andrzejak, Kreuz, David, Elger & Lehnertz, 2003). Gomes (2001) considera também que o estudo da variação de indicadores da variabilidade da freqüência cardíaca (VFC) pode ser realizado ao longo de alguma manobra fisiológica ou farmacológica para se observar a mudança de dinâmica (não-estacionariedade) e compreender melhor como os sistemas envolvidos evoluem.

2.2.1 Definição de Estacionariedade

De uma maneira geral e sem rigor matemático, uma série temporal é dita estacionária se suas propriedades estatísticas essenciais não dependem do tempo (Witt, Kurths & Pikovsky, 1998). Do ponto de vista da matemática e estatística, podemos distinguir dois tipos de estacionariedade (Leon-Garcia, 1994): *i*) estacionariedade estrita (forte). *ii*) estacionariedade no sentido amplo⁵ (fraca). Pode-se defini-las da maneira que se segue:

Um processo estocástico $\{X_t\}$, sendo $t \in \mathbb{R}$, é chamado *estritamente estacionário*, se para qualquer seqüência temporal $t_1, t_2, t_3, \dots, t_n$ e para qualquer inteiro τ , a distribuição de probabilidade conjunta de $\{X_{t_1}, X_{t_2}, \dots, X_{t_n}\}$ e de $\{X_{t_1+\tau}, X_{t_2+\tau}, \dots, X_{t_n+\tau}\}$ coincidem. Na linguagem de sistemas dinâmicos isso significa que as propriedades estatísticas do fluxo no espaço de fases obtido a partir da imersão de janelas distintas da série temporal são as mesmas (Witt et al., 1998).

De uma maneira menos estrita, diz-se que um processo estocástico $\{X_t\}$ é *estacionário no sentido amplo* se o mesmo possui média e autocovariância (ou de maneira equivalente, a autocorrelação) que satisfazem ao seguinte conjunto de equações:

$$\begin{aligned} m_X(t) &= m, & \forall t \\ C_X(t_1, t_2) &= C_X(t_1 - t_2), & \forall t_1, t_2, \end{aligned} \quad (2.7)$$

⁵Wide-Sense Stationarity (WSS).

ou seja, possui média⁶ constante e autocovariância⁷, C_X , dependente somente do atraso relativo de tempo $t_1 - t_2$ e não dos tempos t_1 e t_2 propriamente dito (Leon-Garcia, 1994).

Obviamente, um processo *estritamente estacionário* é também *estacionário no sentido amplo* (WSS).

Considera-se ainda os processos chamados *Cicloestacionários*. Um processo estocástico $\{X_t\}$, sendo $t \in \mathbb{R}$, é chamado *cicloestacionário*, se para qualquer seqüência temporal $t_1, t_2, t_3, \dots, t_n$ e para múltiplos inteiros m de um período T , a distribuição de probabilidade conjunta de $\{X_{t_1}, X_{t_2}, \dots, X_{t_n}\}$ e de $\{X_{t_1+mT}, X_{t_2+mT}, \dots, X_{t_n+mT}\}$ coincidem (Leon-Garcia, 1994).

De maneira semelhante aos processos estacionários no sentido amplo, define-se um processo estocástico *cicloestacionário no sentido amplo* se suas funções de média e autocovariância atendem às seguintes equações:

$$\begin{aligned} m_X(t + mT) &= m_X(t), & \forall t, m \\ C_X(t_1 + mT, t_2 + mT) &= C_X(t_1, t_2), & \forall t_1, t_2, m \end{aligned} \quad (2.8)$$

em que m_X é a função de média, C_X a função de autocovariância, $m \in \mathbb{Z}^+$ é um inteiro qualquer e $T \in \mathbb{R}$ é um período de tempo. Ressalta-se ainda que, se um processo é cicloestacionário, então esse é também cicloestacionário no sentido amplo (Leon-Garcia, 1994).

Do ponto de vista físico, sistemas são chamados estacionários se suas propriedades físicas principais não mudam com o tempo. Na maioria dos casos essas propriedades são, entretanto, desconhecidas e a estacionariedade do sistema poderá somente ser analisada por meio de sua observação temporal (Rieke, Sternickel, Andrzejak, David, Elger & Lehnertz, 2002). O caráter determinístico de muitos sistemas físicos não conflita com a interpretação de seus sinais como realizações de processos estocásticos. Se um conjunto de observações é analisado, não se sabe, *a priori*, se o processo sob investigação tem um caráter determinístico ou estocástico ou uma mistura de ambos. Na linguagem de sistemas dinâmicos, o equivalente para estacionariedade (estrita) é a existência de uma medida ergódica invariante (Witt et al., 1998), embora um processo estacionário (estacionariedade estrita ou no sentido amplo) não necessariamente precise ser ergódico⁸

⁶ $m_X(t) = E[X_t]$.

⁷ $C_X(t_1, t_2) = E[\{X(t_1) - m_X(t_1)\}\{X(t_2) - m_X(t_2)\}]$

⁸Um processo é considerado *ergódico* em alguma medida de quantidade se essa quantidade converge à proporção que o tempo de observação aumenta. Cita-se por exemplo um processo cuja média temporal

(Leon-Garcia, 1994). Portanto, existe uma relação muito próxima entre as visões física e estatística da estacionariedade.

2.2.2 Métodos de Detecção de Não-estacionariedade

Embora se tenha uma definição matematicamente precisa de estacionariedade, não existe um método claro e não ambíguo para se determinar a estacionariedade de uma série temporal real. Na prática, por exemplo, é extremamente difícil, senão impossível, verificar-se *estacionariedade estrita* (Manuca & Savit, 1996; Leon-Garcia, 1994), embora essa questão não seja necessariamente um problema. Na análise de série temporal linear, muitas aplicações verificam mudanças em momentos até segunda ordem (média, variância ou potência espectral) para testes de estacionariedade no sentido amplo (Schreiber, 1997). No contexto de sistemas não-lineares, estacionariedade no sentido amplo pode não ser suficiente, principalmente se forem aplicados algoritmos da teoria de caos determinístico nas séries temporais. É desejável então, nesse contexto, usar-se quantificadores não-lineares (correlações de alta ordem, expoentes de Lyapunov, e outros) para checar não-estacionariedade (Gomes, 2001; Schreiber, 1997).

Em muitas aplicações reais, a questão de estacionariedade é freqüentemente abordada de maneira bastante subjetiva (Gomes, 2001; Manuca & Savit, 1996). Além da subjetividade encontrada na diversidade de métodos para análise de estacionariedade, é importante considerar-se que a determinação (ou detecção) da estacionariedade requer que o tempo de observação⁹ seja grande o suficiente para cobrir as escalas de tempo essenciais¹⁰ do sistema observado (Witt et al., 1998). Deve-se, portanto, considerar o tempo de observação um parâmetro de importância crucial na análise de estacionariedade. Até mesmo um sistema estacionário pode levar a uma detecção espúria de não-estacionariedade se o tempo de observação for menor que as escalas de tempo características do sistema. Esse tipo de detecção espúria de não-estacionariedade foi chamado *não-estacionariedade de medição* por Rieke, Andrzejak, Mormann & Lehnertz (2004).

Uma outra forma de detecção espúria de não-estacionariedade, chamada *não-estacionariedade estatística*, pode ser atribuída à subestimação de flutuações estatísticas usadas por vários métodos estatísticos clássicos na determinação da não estacionariedade (Rieke et al., 2004).

converge para a média real ao longo do tempo de observação. Nesse caso diz-se que o processo é *ergódico na média*.

⁹Tempo de coleta da série temporal.

¹⁰Período de tempo mínimo necessário para observar-se os padrões dinâmicos do processo.

Apesar das considerações feitas nos parágrafos anteriores, é muito importante ter-se um procedimento que possibilite verificar se uma série temporal é estacionária ou não e ainda possa detectar regiões de estacionariedade em um registro de observação.

Pode-se encontrar na literatura vários métodos estatísticos para teste de estacionariedade. A maioria desses métodos são baseados em uma mesma idéia fundamental: estimar um certo parâmetro usando diferentes segmentos da série. Se as variações observadas forem consideradas significativas, ou seja, estiverem além das flutuações estatísticas esperadas, a série temporal é declarada não-estacionária (Schreiber, 1997). Outros métodos, ao invés de focarem nas flutuações estatísticas dos segmentos das séries, voltam sua atenção para as características dinâmicas intrínsecas das séries temporais (Witt et al., 1998).

Bendat & Piersol (1986) apresentam um método clássico que busca por flutuações estatísticas de propriedades de primeira e segunda ordem em segmentos não sobrepostos. Esse método, conhecido como *teste de arranjos reversos*¹¹, divide a série em vários segmentos e calcula a média e desvio padrão de cada segmento. A partir dessas propriedades estatísticas o algoritmo segue para decidir sobre a estacionariedade ou não da série. O uso apenas dessas propriedades pode levar a uma detecção espúria de estacionariedade pelo fato das mesmas conterem pouca informação a respeito do processo. Isliker & Kurths (1993) citam, por exemplo, o trabalho de Mandelbrot (1982), no qual é mostrado que o *movimento Browniano fracional* possui variância¹² que depende somente de sua duração t . Partes de uma série temporal gerada por esse processo, que é não-estacionário, terão, portanto, a mesma variância se tiverem o mesmo tamanho. Essa condição levaria a uma detecção espúria de estacionariedade.

Isliker & Kurths (1993) propuseram um teste de estacionariedade que supostamente supera estratégias clássicas como a citada anteriormente. Uma medida invariante é apresentada como uma descrição estatística de um sistema no espaço de estado, ou seja, uma densidade de probabilidade, medindo quão frequentemente as diferentes partes do espaço de estado são *visitadas*. De acordo com os autores, essa medida reflete a dinâmica de um sistema por levar em consideração que algumas partes são mais *visitadas* que outras. Essa densidade contém a informação de todos os momentos estatísticos, os quais são calculados a partir dela. Esse método foi usado por Braun, Kowallik, Freking, Hader, Kniffki & Meesmann (1998) no estudo de não-linearidades presentes no sinal de variabilidade da frequência cardíaca.

¹¹Do inglês: Reverse Arrangements Test.

¹²A variância é proporcional a $t^{\frac{1}{D_H}}$, em que D_H é constante (ver detalhes em (Mandelbrot, 1982)).

Manuca & Savit (1996) analisam, nesse importante trabalho, as propriedades de estacionariedade de uma série temporal considerando fortemente os padrões dinâmicos do sistema gerador dessa série. Essa metodologia assume que o sistema a ser estudado pode ser descrito como um sistema contendo alguns conjuntos de padrões internos de dinâmica e imersos em um *ambiente* possivelmente variante no tempo. O teste de não-estacionariedade consiste em dividir-se, inicialmente, a série temporal em algumas janelas. Por meio de uma medida estatística obtida por integrais de correlação (Grassberger & Procaccia, 1983), quantifica-se uma medida de distância entre regimes dinâmicos subjacentes de pares das janelas da série. O objetivo dos autores é estimar quão próximos estão os padrões dinâmicos entre duas janelas da série.

Schreiber (1997) propôs um teste de estacionariedade que verifica a compatibilidade de aproximações não-lineares dos padrões dinâmicos em diferentes segmentos da série. Define-se o índice *erro de predição cruzada* (Cross-prediction error) usado na comparação dos segmentos. A idéia geral desse índice é obter o erro de predição cruzada ao se tomar predições usando-se um segmento i como base de dados para predizer valores dentro de outro segmento j . Schreiber (1997) ainda pondera que a ampla noção de que um sistema deve permanecer invariante durante o tempo de medição, ou não é necessária ou não é uma condição suficiente para estacionariedade. A razão é que não há uma distinção *a posteriori* entre um parâmetro do sistema (a permanecer constante) e uma variável (que pode evoluir no tempo). Portanto, com esse método o autor considera um sinal estacionário se qualquer coisa que varie no tempo (não importando se é chamada variável ou parâmetro), o faça numa escala de tempo tal que as mudanças tendam a uma média num tempo muito menor que a duração da medição.

Kennel (1997) propõe um teste de não-estacionariedade extraindo informação da distribuição temporal de pontos num espaço de estado reconstruído a partir de dados observados. Um teste de hipótese estatístico é sugerido, podendo discernir se algumas mudanças lentas subjacentes ocorrem. A não-estacionariedade é quantificada usando-se propriedades dos *vizinhos mais próximos* (Nearest neighbors) no espaço de estado. A idéia básica é que o valor esperado do número de pontos no espaço de fase reconstruído, que tem sua vizinhança mais próxima na mesma metade da seqüência, é mínimo para uma seqüência estacionária. Quando se pensa em geometria no espaço de fase, não-estacionariedade traz uma tendência de que pontos próximos no espaço estão também próximos no tempo (Schreiber, 1999). De acordo com o autor, esse método parece poderoso e depende de informação dinâmica não trivial, embora não forneça um teste estatístico óbvio e requeira dados de baixa dimensão.

Casdagli (1997) apresentou em detalhes a técnica de *mapas de recorrência*¹³ e algumas variantes, introduzida primeiramente por (Eckmann et al., 1987), para análise de séries temporais não-estacionárias. Dentre os objetivos da análise proposta, o teste de hipótese de estacionariedade foi considerado. O mapa de recorrência da série em estudo é construído e a partir desse mapa prossegue-se com a detecção de pontos de mudança de padrões dinâmicos. O autor usa dados subrogados para teste de significância estatística dos pontos de mudança encontrados. Schreiber (1997, 1999) afirma que essa técnica é de difícil análise e não possui muito boa resolução temporal.

Witt et al. (1998) propuseram também um teste estatístico de séries temporais, combinando um teste para independência temporal da densidade de probabilidade de uma dimensão (1D probability density) com um outro baseado na independência temporal do espectro de potência. O tamanho dos segmentos da série em estudo é um fator limitante desse procedimento. O autor apresenta um estudo comparativo na aplicação desse método em vários tipos de processos.

Mohanty (2000) propôs um teste de detecção de não-estacionariedade no contexto da análise de dados de ondas gravitacionais. A grande quantidade de dados produzidos pelos detectores de ondas gravitacionais são principalmente ruído. Espera-se que, imersos no ruído, se encontre o sinal desejado. Na análise dos dados para detecção desses sinais gravitacionais, requer-se um conhecimento preciso do ruído, em outras palavras, uma caracterização estatística do processo estocástico. A detecção de não-estacionariedade é relevante nesse procedimento. A idéia central do teste de não-estacionariedade é a detecção de mudanças estatisticamente significativas na densidade espectral de potência (Power Spectral Density). Para comparação da PSD de diferentes segmentos da série, o teste *t-student* é usado.

Em Gao (1999, 2001), o autor sustenta o estudo de recorrência de estados, ou seja, quão freqüentemente uma região pequena (vizinhança mais próxima) no espaço de fase é visitada, como uma forma de observar mudanças no padrão dinâmico de sistemas. Gao (1999) fez uma análise de quantificação de recorrência e propôs dois algoritmos baseados na estatística de tempo de recorrência para detectar não-estacionariedades e transições de estado (Gao, 2001).

¹³A recorrência de estados, no sentido de que estados estão arbitrariamente próximos depois de algum tempo, é uma propriedade fundamental de sistemas dinâmicos determinísticos e típicos para sistemas caóticos e não-lineares. Mapas de recorrência (recurrence plots) podem ser visualizados no espaço de fases. A recorrência de um estado no tempo i em um outro tempo j é construída graficamente numa matriz quadrada (duas dimensões) com pontos pretos e brancos, em que os pontos pretos marcam a recorrência e ambas as coordenadas são de tempo (Eckmann, Kamphorst & Ruelle, 1987).

Rieke et al. (2002) também propuseram uma medida de não estacionariedade baseada na análise de distribuição de distâncias temporais de vetores de vizinhança no espaço de estado. O desvio da distribuição dessas distâncias de uma distribuição de distâncias que é esperada sob condições de estacionariedade, foi usado para medir a não-estacionariedade¹⁴. Quase todos esses métodos dividem as séries temporais em segmentos menores, o que frequentemente reduz a capacidade estatística das mesmas. O diferencial proposto por Rieke et al. (2002) é não dividir a série temporal para detectar não-estacionariedades. Rieke et al. (2004) propuseram uma modificação desse método, visando uma estimação correta da significância estatística do valor medido para não-estacionariedade.

Cao, wen Tung, Gao, Protopopescu & Hively (2004) argumentaram que a maioria dos métodos baseados numa quantificação de algum aspecto dos vizinhos mais próximos no espaço de fases, possuem alto custo computacional. Eles propuseram o uso do conceito de *entropia de permutação*¹⁵ na detecção de mudanças no regime dinâmico de séries temporais complexas. Pondera-se que a medida de entropia de permutação é conceitualmente simples e computacionalmente eficiente. Observa-se, entretanto, que é necessário uma grande quantidade de observações da série temporal em estudo, para que a mesma seja dividida em janelas de tamanho suficientemente longo por causa de seu conteúdo estatístico.

Fica claro que a questão da estacionariedade ainda está aberta e a análise e o entendimento de sinais não-estacionários é um tópico de pesquisa atual em muitas áreas da ciência. Muito provavelmente não existe uma metodologia única para determinar-se não-estacionariedades cuja aplicabilidade seja universal e de uso indiscriminado. Cao et al. (2004) ainda ponderaram que várias medidas de estacionariedade poderiam ser usadas de maneira complementar, a fim de tirar vantagens de seus respectivos méritos dentro do âmbito de suas aplicações.

2.3 Métodos de Análise de Séries Temporais Não-estacionárias

A análise e processamento de séries temporais não-estacionárias tem despertado o interesse científico nos últimos anos. Dentre as motivações para o estudo dessas séries,

¹⁴Rieke et al. (2002, 2004) chama esse processo de *perda de recorrência*, quando uma não-estacionariedade é observada.

¹⁵Entropia de permutação (permutation entropy) é uma medida de complexidade de séries temporais introduzida por Bandt & Pompe (2002). Medidas de complexidade como, entropia, dimensão fractal e expoentes de Lyapunov, foram desenvolvidas para comparar séries temporais e distinguir comportamentos regulares, caóticos e estocásticos.

embora já tenham sido consideradas nas seções anteriores, citam-se as mais importantes: *i)* observações feitas em muitos sistemas reais são não-estacionárias; *ii)* a hipótese fundamental assumida pela maioria das técnicas, lineares e não-lineares, de análise de séries temporais é a estacionariedade da série em estudo.

Segundo Casdagli (1997), pode-se considerar quatro objetivos principais na análise de séries temporais não-estacionárias:

1. *Caracterização.* Suponha que os padrões dinâmicos subjacentes de certas séries temporais sejam similares no início e no fim dessas séries, mas diferentes na porção intermediária. O objetivo de uma técnica que caracteriza a não-estacionariedade é extrair essa informação de uma maneira conveniente. A metodologia mais comum para caracterização de não-estacionariedades é a divisão da série temporal em segmentos e verificando possíveis mudanças, ao longo dos segmentos, de uma quantidade de interesse. Esse tipo de abordagem é muito dependente da quantidade escolhida e é, até certo ponto, arbitrário. Citam-se os trabalhos de Manuca & Savit (1996), Schreiber (1997), Casdagli (1997).
2. *Predição.* Com o propósito de obter-se predições acuradas para séries temporais não-estacionárias, pode ser necessário modificar os algoritmos de predição desenvolvidos para séries temporais estacionárias. Esse objetivo está proximamente relacionado com a caracterização e modelagem da não-estacionariedade. Citam-se os trabalhos de Casdagli (1997) e Stark (1993).
3. *Detecção de ponto de mudança.* Em alguns casos pode haver um ponto único a partir do qual a dinâmica subjacente da série temporal muda. O objetivo é identificar esse ponto de mudança. Citam-se os trabalhos de Kennel (1997) Casdagli (1997).
4. *Teste de hipótese.* O objetivo é desenvolver testes de hipótese cuja hipótese nula é a estacionariedade da série temporal em estudo. Se a hipótese nula for aceita, pode-se então, com mais confiança em princípio, usar técnicas de análise de séries temporais estacionárias. Esse problema pode ser abordado sob uma perspectiva não-linear ou linear. Citam-se, como exemplo, os trabalhos de Isliker & Kurths (1993), Kennel (1997), Rieke et al. (2002) e Casdagli (1997).

Os objetivos 3 e 4 acima estão relacionados entre si e de uma maneira geral foram abordados na seção 2.2.2. No contexto da identificação de sistemas variantes no tempo vê-se a abordagem, não exclusiva, dos objetivos 1 e 2 mencionados acima.

Além dos trabalhos já mencionados, pode-se ainda citar outros que envolvem modelagem paramétrica e estimação de parâmetros de sistemas variantes no tempo, tais como: Box & Jenkins (1970), Grenier (1983), (Martin, 1986), Cho et al. (1991), Tsatsanis & Giannakis (1993), Penny & Roberts (1999), Cao & Schartz (1999), Bouzeghoub, Ellacott, Easdown & Brown (2000), Wei & Billings (2002), Zou, Wan & Chon (2003), Jiang & Zhang (2004a) e Zou & Chon (2004). Algumas das metodologias abordadas nos trabalhos ora citados serão consideradas nos capítulos seguintes.

Os modelos para a análise e previsão de séries temporais podem ser divididos em duas classes principais: os modelos baseados em técnicas convencionais e os baseados em inteligência artificial. O método mais proeminente dentre as técnicas clássicas é a abordagem proposta por Box & Jenkins (1970). Dentre as derivações lineares do modelo de Box & Jenkins vale destacar: *i*) modelos ARIMA (*AutoRegressive Integrated Moving Average*), que são apropriados para séries não-estacionárias; *ii*) modelos AR (*AutoRegressive*) e ARMA (*AutoRegressive Moving Average*), apropriados para séries estacionárias.

Ainda dentro dos modelos de regressão encontram-se os modelos NARMAX (*Nonlinear AutoRegressive Moving Average with eXogenous inputs*) polinomiais/racionais que utilizam técnicas de identificação não-lineares (Leontaritis & Billings, 1985b,a). Especialmente apropriados para séries temporais financeiras não-lineares, citam-se os modelos ARCH (*AutoRegressive Conditional Heteroscedasticity*) e suas derivações (Moretlin & Toli, 2004), embora trabalhos recentes ponderem que esses modelos não são adequados para qualquer série econométrica. Como exemplo, Brooks & Hinich (1998) afirmam que modelos ARCH falham em capturar a natureza variante no tempo de taxas de retorno em mercados de câmbio.

Considerando as metodologias de modelagem baseadas em inteligência artificial, podem-se citar técnicas de *Redes Neurais Artificiais* (RNA). Inspirada em sistemas biológicos, particularmente em pesquisas no cérebro humano, as redes neurais artificiais são capazes de aprender e generalizar a partir da experiência. Dentre outras características importantes, ressalta-se o fato de serem consideradas aproximadores universais e não-lineares (Zhang et al., 1998). As redes neurais artificiais possuem derivações como as redes neurais *fuzzy*. RNAs têm sido usadas em conjunto com outras técnicas visando uma melhora no desempenho do horizonte de previsão e/ou otimização do tamanho das redes (Padmakumari et al., 1999; Liao & Tsao, 2004).

A análise e processamento de séries não-estacionárias inclui também métodos de representação tempo-frequência (Pola et al., 1996; Gamero et al., 1996; Potamianos & Maragos, 2001) e técnicas de análise espectral (Bianchi et al., 1993; Pola et al., 1996; Cesarelli et al.,

1997; Steenis et al., 2003). A análise espectral é fundamental em áreas onde o interesse básico é a procura de periodicidade nos dados, como em Meteorologia e Oceanografia, Medicina, Biologia, economia e outras (Morettin & Tolo, 2004).

Identificação de Sistemas Dinâmicos

A identificação de sistemas trata do problema de construção de modelos baseados em dados medidos do sistema.

Um modelo tem como objetivo descrever a relação entre as variáveis do sistema. Quando tal relação for expressa em termos de equações matemáticas, o modelo é dito matemático. Dentre as motivações para a construção de modelos pode-se citar que: *i)* algumas técnicas de controle avançado requerem um modelo do sistema; *ii)* modelos podem ser usados para análise, simulação e previsão.

A complexidade de um modelo matemático dependerá da aplicação visada, seja na pesquisa, no projeto ou no controle (Eykhoff, 1974).

O levantamento e a formulação matemática de todos os fenômenos que afetam o comportamento de um dado sistema é uma tarefa extremamente complexa. Com isso, é impossível o modelo reproduzir exatamente o comportamento do sistema original. O modelo deve então, ser capaz de reproduzir o comportamento original da melhor maneira possível, pois caso contrário, todos os esforços posteriores para a análise do modelo e controle do sistema serão pouco eficientes. Dois grandes grupos de técnicas usadas para a obtenção de modelos matemáticos, são:

- modelagem pela física do processo ou modelagem caixa-branca (Garcia, 1997);
- modelagem a partir de dados ou identificação de sistemas (Aguirre, 2004).

A modelagem pela física do processo utiliza as leis físicas que descrevem os fenômenos envolvidos. Esta é feita somente quando se tem disponível uma visão global do comportamento do processo. Quando o sistema é grande e complexo e há a necessidade de atuação de controle em tempo real, por exemplo, torna-se inviável a utilização de técnicas de modelagem caixa-branca (Jácome, 1996; Coelho, 2002). Uma alternativa é a utilização de

técnicas de identificação de sistemas, por meio da qual o modelo é obtido a partir de dados de entrada e saída e distúrbios (quando disponível) do sistema. Identificação caixa-cinza utiliza informação *a priori* para auxiliar na determinação da estrutura e/ou estimação dos parâmetros, resultando em modelos mais concisos (Corrêa, 2001).

Vários algoritmos têm sido desenvolvidos para auxiliar a identificação de sistemas usando técnicas caixa-preta (Billings & Fakhouri, 1978; Korenberg, Billings, Liu & McIlroy, 1988; Billings, Chen & Korenberg, 1989).

No que diz respeito à *análise de séries temporais e identificação de sistemas*, Korenberg & Paarmann (1991) observam que existe uma considerável sobreposição entre ambas abordagens. Uma notável diferença é que na análise de séries temporais não há acesso ao modelo da “entrada”. Assume-se, nesse trabalho, “identificação de sistemas” como modelagem de séries temporais ainda que seja incluída uma entrada que não seja mensurável. A seção 3.2 faz considerações a cerca de sinais de entrada para modelagem de sistemas.

3.1 Procedimento para a identificação de sistemas

O procedimento para a identificação de sistemas exige o constante envolvimento do modelador, e consiste em:

- projeto e execução de testes para a obtenção dos dados e determinação da taxa de amostragem;
- detecção de não-linearidades e/ou não-estacionariedades no sistema;
- escolha da representação;
- detecção da estrutura do modelo;
- estimação dos parâmetros do modelo;
- validação do modelo.

A obtenção dos dados depende do sistema a ser investigado, bem como das condições experimentais.

A aquisição de dados experimentais deve ser realizada de tal forma que se minimize o efeito do ruído. A modelagem do ruído pode comprometer a validade do modelo encontrado, e uma forma de aliviar este problema é através de cuidados na escolha da estrutura

do modelo (Mendes & Billings, 1998). É possível fazer a identificação de um sistema sem a segunda etapa, por exemplo, identificação linear.

3.2 Experimentação

A experimentação consiste na determinação do sinal de entrada, na escolha da taxa de amostragem e na coleta dos dados para a estimação dos parâmetros e ajuste do modelo (Söderström & Stoica, 1989). Esta etapa no processo de identificação de sistemas é a mais importante, pois os resultados posteriores dependerão da qualidade e da quantidade de informações contidas nos dados coletados.

Os dados coletados durante o processo de operação normal do sistema podem, em alguns casos, ser usados na identificação do sistema. Porém, é preferível, sempre que possível, injetar sinais externos no sistema, capazes de excitar toda a dinâmica do sistema durante o processo de identificação. Todas as características estáticas e dinâmicas que não forem excitadas não aparecerão nos dados e o que não estiver nos dados não poderá ser modelado. Dentre os sinais usados na identificação de sistemas podem-se destacar os sinais binários pseudo-aleatórios (PRBS) e o ruído “branco”, uma vez que esses podem excitar toda a dinâmica do sistema¹ (Aguirre, 2004).

A escolha da taxa de amostragem também é importante, pois taxas de amostragem diferentes podem resultar em modelos diferentes. Pode-se classificar os dados com relação à taxa de amostragem em três categorias: subamostrados, bem amostrados e superamostrados. Dados subamostrados tipicamente não contêm informação dinâmica suficiente, impossibilitando assim, a determinação de um modelo para o processo. Um sinal superamostrado, em que sucessivas amostras tendem a estar fortemente correlacionadas, pode proporcionar mau condicionamento numérico da matriz de regressores e dificultar a determinação da estrutura do modelo (Billings & Aguirre, 1995), porém, essas dificuldades podem ser contornadas se esses sinais forem dizimados, possibilitando assim, a utilização desses na estimação e validação do modelo.

Para auxiliar na escolha da taxa de amostragem, pode-se empregar um procedimento que consiste em utilizar as funções de autocorrelação linear, $\phi_{y'y'}$, e não-linear, $\phi_{y'2y'2}$, da saída do sistema, definidas em (3.1) e (3.2), respectivamente:

$$\phi_{y'y'} = E[(y(k) - \overline{y(k)})(y(k - \tau) - \overline{y(k)})], \quad (3.1)$$

¹Sinais PRBS e ruído branco são exemplos de sinais persistentemente excitantes.

$$\phi_{y'2y'2} = E[(y^2(k) - \overline{y^2(k)})(y^2(k - \tau) - \overline{y^2(k)})], \quad (3.2)$$

sendo que $\overline{y(k)}$ e $\overline{y^2(k)}$ representam os valores médios, E a esperança matemática e o apóstrofe indica que o sinal teve sua média removida.

A partir das funções de autocorrelação (3.1) e (3.2), pode-se determinar a seguinte constante:

$$\tau_m = \min\{\tau_{y'}, \tau_{y'2}\}. \quad (3.3)$$

sendo $\tau_{y'}$ o instante do primeiro mínimo de $\phi_{y'y'}$ e $\tau_{y'2}$ o instante do primeiro mínimo de $\phi_{y'2y'2}$. Desta forma, pode-se determinar a taxa de amostragem, T_s , através da seguinte relação (Billings & Aguirre, 1995; Aguirre, 2004):

$$\frac{\tau_m}{20} \leq T_s \leq \frac{\tau_m}{10}. \quad (3.4)$$

Em alguns casos, este intervalo pode ser relaxado para:

$$\frac{\tau_m}{25} \leq T_s \leq \frac{\tau_m}{5}. \quad (3.5)$$

3.3 Detecção de não-linearidades e/ou não-estacionariedades

A existência de não-linearidades no sistema pode ser verificada a partir de testes não paramétricos, utilizando apenas os sinais coletados, ou seja, estes testes são realizados independentes do tipo de representação e da estrutura. Dentre os testes para detectar não-linearidade pode-se citar: testes em regime permanente, testes do valor médio da saída, testes no domínio da frequência, testes no domínio do tempo e testes de correlação (Coelho, 2002), sendo que os dois últimos destacam-se pela simplicidade e eficiência.

A questão de detecção de não-estacionariedades foi abordada no capítulo 2 e pode ser revisada na Seção 2.2.2.

3.4 Representações de Sistemas Lineares

Pode-se dizer que a modelagem de séries temporais teve início com o trabalho de Yule (1927), que utilizou modelos lineares auto-regressivos AR (*AutoRegressive*) no problema de previsão da série de números de manchas solares de Wolfer, sendo essa uma série bastante conhecida devido à sua alta variabilidade e dificuldade de predição (Mingoti

& Neves, 1999). O modelo autônomo AR, o qual utiliza informações passadas da série temporal, para estimar a saída atual, $y(k)$, é dado por:

$$y(k) = \sum_{j=1}^{n_y} a_j y(k-j) + e(k), \quad (3.6)$$

sendo $y(k)$ a série temporal, $a_0 = 1$ e a_j os parâmetros da parte auto-regressiva do modelo ($a_{n_y} \neq 0$), n_y é um valor inteiro que define o máximo atraso de y , conseqüentemente, a ordem dinâmica do modelo AR. $e(k)$ é o resíduo de modelagem, ou seja, o erro um passo à frente entre a série observada e a série obtida via modelo AR.

Em (Wold, 1954) foi mostrado que qualquer série temporal discreta poderia ser representada por modelos AR e modelos de média móvel, MA (*Moving Average*). Porém, a implementação desses métodos só foi possível a partir da década de 60 com o advento dos computadores (Oliveira, 2002; Aguirre, 2004).

Nas seções seguintes serão considerados alguns modelos de sistemas invariantes no tempo, com exceção do modelo ARIMA. No capítulo 4 abordar-se-á a modelagem de sistemas variantes no tempo.

3.4.1 Modelos ARMA

Uma classe mais geral de representações lineares autônomas, denominada modelos ARMA (*AutoRegressive Moving Average*), utiliza o resíduo de estimação no processo de estimação, sendo dado por:

$$y(k) = \sum_{j=1}^{n_y} a_j y(k-j) + \sum_{i=1}^{n_e} c_i e(k-i) + e(k) \quad (3.7)$$

Destaca-se que nos modelos ARMA as raízes dos polinômios:

$$\begin{aligned} A(q) &= 1 - \sum_{j=1}^{n_y} a_{n_y} q^{-j} \\ C(q) &= \sum_{j=1}^{n_e} c_j q^{-j}, \end{aligned} \quad (3.8)$$

devem ter módulo menor que a unidade (Tong, 1990). $A(q)$ e $C(q)$ são denominados funções geradoras auto-regressivas e de médias móveis, respectivamente, e q o operador de atraso. Para assegurar estacionariedade $|A| < 1$ (Box & Jenkins, 1970).

Algumas vantagens e limitações dos modelos ARMA podem ser encontradas em Tong (1990), tais como:

- Matematicamente, os modelos lineares baseados em equações de diferenças lineares são considerados o tipo mais simples de representações discretas e possuem uma teoria bem “fundamentada” já disponível (Box et al., 1994).
- O tempo computacional requerido para obter um modelo parsimonioso ARMA é pequeno.
- Salienta-se que ao longo dos anos, estes modelos têm sido amplamente utilizados como ferramentas práticas em análise, previsão e controle de séries temporais e têm apresentado resultados considerados satisfatórios (Box & Jenkins, 1970; Tong, 1990).
- Modelos ARMA estacionários tem distribuição simétrica, logo, não são apropriados para aproximar os dados que exibam forte assimetria, como por exemplo, dados hidrológicos.
- Pode ser usado para explicar relações não-lineares desde que essas linearizadas.

3.4.2 Modelos Não Autônomos: ARX, ARMAX.

De maneira geral, os modelos discretos podem ser extraídos a partir do seguinte modelo geral (Aguirre, 2004):

$$A(q)y(k) = \frac{B(q)}{F(q)}u(k) + \frac{C(q)}{D(q)}\nu(k) \quad (3.9)$$

sendo q o operador de atraso, $\nu(k)$ o ruído branco e $A(q)$, $B(q)$, $C(q)$, $D(q)$ e $F(q)$ polinômios definidos como:

$$\begin{aligned} A(q) &= 1 + a_1q^{-1} + \dots + a_{n_y}q^{-n_y}; \\ B(q) &= b_1q^{-1} + \dots + b_{n_u}q^{-n_u}; \\ C(q) &= 1 + c_1q^{-1} + \dots + c_{n_\epsilon}q^{-n_\epsilon}; \\ D(q) &= 1 + d_1q^{-1} + \dots + d_{n_d}q^{-n_d}; \\ F(q) &= 1 + f_1q^{-1} + \dots + f_{n_f}q^{-n_f}; \end{aligned} \quad (3.10)$$

De acordo com a forma de cada um dos polinômios (3.10), pode-se obter modelos AR, ARMA, além dos sub-tipos não-autônômos, ARX, ARMAX e outras (Aguirre, 2004).

A derivação não-autônôma do modelo AR, modelo ARX (*AutoRegressive with eXogenous inputs*), pode ser obtida a partir do modelo geral (3.9), tomando-se $C(q) = D(q) = F(q) = 1$ sendo $A(q)$ e $B(q)$ polinômios arbitrários, resultando em:

$$\begin{aligned} y(k) &= \frac{B(q)}{A(q)}u(k) + \frac{1}{A(q)}\nu(k), \\ y(k) &= a_0 + \sum_{j=1}^{n_y} a_j y(k-j) + \sum_{i=1}^{n_u} b_i u(k-i) + e(k), \end{aligned} \quad (3.11)$$

em que $e(k) = \nu(k)/A(q)$ representa o ruído modelado como um processo branco filtrado por um filtro auto-regressivo, com pólos idênticos aos do processo, que são as raízes do polinômio $A(q)$.

Pode-se obter também o modelo ARMAX (*AutoRegressive Moving Average with eXogenous inputs*) por meio do modelo geral (3.9), tomando-se $D(q) = F(q) = 1$ sendo $A(q)$, $B(q)$ e $C(q)$ polinômios arbitrários, cujo resultado é:

$$\begin{aligned} y(k) &= \frac{B(q)}{A(q)}u(k) + \frac{C(q)}{A(q)}\nu(k), \\ y(k) &= a_0 + \sum_{j=1}^{n_y} a_j y(k-j) + \sum_{i=1}^{n_u} b_i u(k-i) + e(k), \end{aligned} \quad (3.12)$$

em que $e(k) = \frac{C(q)}{A(q)}\nu(k)$ representa o ruído modelado como um processo branco filtrado por um filtro ARMA. Pode-se observar que o modelo ARMA, discutido na seção 3.4.1, é um caso particular do modelo ARMAX, quando não há sinais exógenos, ou seja, quando $u(k) = 0$.

O modelo ARX e o ARMAX são considerados pertencentes à classe de modelos de *erro na equação*. Modelos desse tipo são aqueles em que as funções de transferência do processo e do ruído têm o polinômio $A(q)$ como fator comum. Um outro grupo de modelos não descritos aqui, é o do tipo *erro na saída*, que podem ser escritos na forma da equação (3.9), mas com o polinômio $A(q) = 1$.

3.4.3 Modelos ARIMA

Muitas séries temporais reais apresentam comportamento *não-estacionário*, como fora discutido no Capítulo 2. Box & Jenkins (1970) propuseram uma forma estendida do modelo ARMA, que lida com o comportamento não estacionário da série temporal, denominado modelo ARIMA (*AutoRegressive Integred Moving Average*) de ordem (n_y, d, n_e) .

A série temporal não-estacionária, $y(k)$ é diferenciada de uma ordem d (normalmente, $d = 0, 1$ ou 2), tornando-se, em princípio, uma série estacionária². Uma vez que a série foi transformada em uma série estacionária, $z(k)$, esta é então modelada usando a representação ARMA (p e $q \neq 0$), ou AR ($q = 0$) ou MA ($p = 0$). De maneira geral, o modelo ARIMA é descrito por:

$$z(k) = \sum_{i=1}^{n_y} a_i z(k-i) + \sum_{j=1}^{n_e} c_j e(k-j) + e(k),$$

$$A(q)z(k) = C(q)e(k), \quad (3.13)$$

sendo que:

$$z(k) = y(k) - y(k-d), \quad (3.14)$$

$$z_k = \nabla^d y_k, \quad (3.15)$$

e d é a d -ésima diferença da série temporal de modo a reduzir ou eliminar a característica não-estacionária da série. O significado do termo “integral”, na sigla ARIMA, é devido à seguinte relação inversa à $z_k = \nabla^d y_k$, dada por:

$$y_k = \nabla^{-d} z_k$$

$$y_k = S^d z_k, \quad (3.16)$$

²Deve-se levar em consideração que a eliminação da não-estacionariedade de uma dada série temporal, só é possível se a mesma permitir esta transformação. Às vezes, a diferenciação pode somente reduzir a não-estacionariedade.

em que,

$$\begin{aligned} y(k) &= Sz(k) + y(1) \\ &= \sum_{j=0}^{\infty} z(k-j) = z(k) + z(k-1) + z(k-2) + \dots + z(1) + y(1). \end{aligned}$$

Por exemplo:

$$\begin{aligned} z(1) = y(2) - y(1) &\Rightarrow y(2) = z(1) + y(1) \\ z(2) = y(3) - y(2) &\Rightarrow y(3) = z(2) + z(1) + y(1) \\ z(3) = y(4) - y(3) &\Rightarrow y(4) = z(3) + z(2) + z(1) + y(1) \\ &\quad \vdots \quad \vdots \quad \vdots \\ z(k-1) = y(k) - y(k-1) &\Rightarrow y(k) = z(k-1) + z(k-2) + \dots + z(1) + y(1) \end{aligned}$$

A metodologia de Box & Jenkins consiste nas seguintes etapas (Box et al., 1994):

- escolha da estrutura do modelo;
- estimação dos parâmetros do modelo;
- diagnóstico (ou validação);
- predição.

Na etapa de identificação é verificada o valor do parâmetro d , ou seja, quantas diferenciações devem ser realizadas na série para retirar sua não estacionariedade. Nesta etapa é feita, também, a escolha de qual tipo de representação será utilizada, visto que o modelo ARIMA engloba todas as outras representações lineares auto-regressivas.

A ordem do modelo da parte auto-regressiva (AR) pode ser obtida via análise das funções de autocorrelação e autocorrelação parcial (Aguirre, 2004). Segundo Oliveira (2002) este é um dos aspectos mais difíceis na utilização da metodologia de Box & Jenkins, visto que para a análise das funções de autocorrelação requer intervenção do analista/modelador, o que implica que este deve ter conhecimento e experiência com tal metodologia.

A técnica para a estimação dos parâmetros usada por Box & Jenkins foi a Máxima Verossimilhança. Entretanto, pode-se utilizar outras técnicas, como a de mínimos quadrados e a Bayesiana.

A validação do modelo obtido, chamada de diagnóstico, analisa os resíduos de modelagem. Se o resíduo assemelhar-se ao ruído branco, o modelo estimado é considerado apropriado para a série temporal em questão. Caso contrário, o modelo deve ser reestimado (Brockwell & Davis, 1996).

A última etapa é a da predição, que consiste em gerar valores futuros da série através do modelo estimado.

Além da não-estacionariedade, séries temporais advindas de diversas áreas apresentam um comportamento sazonal (ver seção 2.1.1.1). A utilização de modelos ARIMA em séries sazonais é extremamente difícil (Coelho, 2002). Frequentemente, é necessário uma quantidade de dados maior do que a usada em séries não sazonais.

3.5 Representações de Sistemas Não-Lineares

Há um vasto número de representações para sistemas não-lineares e a escolha de qual usar e em que circunstância, parecem ser questões que estão longe de serem resolvidas. A escolha do tipo de representação depende principalmente da finalidade do modelo e das ferramentas disponíveis para a sua obtenção, bem como das informações disponíveis sobre o sistema. Sabe-se que para algumas representações não há uma forma sistemática para se detectar a melhor estrutura do modelo (Cassini, 1999).

Dentre as representações não-lineares pode-se destacar: modelos baseados em redes neurais (Zhang et al., 1998; Braga, Carvalho & Ludemir, 2000), modelos NARMAX polinomiais (Leontaritis & Billings, 1985a; Aguirre, 2004), modelos contínuos (Freitas, 2001), Wavelets (Billings & Coca, 1999) e outros.

A representação de sistemas não-lineares teve início por volta de 1930, quando Volterra mostrou que para um sistema não-linear invariante no tempo, o qual gera uma saída contínua e limitada, $y(t)$, quando excitado por uma entrada também contínua e limitada, $u(t)$, a relação entre a entrada e a saída pode ser expressa como (Volterra, 1930):

$$y(t) = h_0 + \int_{-\infty}^{\infty} h_1(\tau_1)u(t - \tau_1)d\tau_1 + \int_{-\infty}^{+\infty} \int_{-\infty}^{\infty} h_2(\tau_1, \tau_2)u(t - \tau_1)u(t - \tau_2)d\tau_1 d\tau_2 \\ + \dots + \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} h_n(\tau_1, \dots, \tau_n)u(t - \tau_1) \dots u(t - \tau_n)d\tau_1 \dots d\tau_n. \quad (3.17)$$

A equação (3.17) é denominada Série de Volterra, e a função $h_n(\tau_1, \dots, \tau_n)$ é chamada *kernel* de Volterra.

A série de Volterra, apesar de sua ampla aplicabilidade na representação de sistemas não-lineares, possui algumas limitações, tal como o grande número de parâmetros requerido para explicar um simples sistema não-linear, o que acarreta em um grande esforço computacional. Este fato é consequência da série de Volterra mapear as entradas passadas para a saída atual $y(t)$. Uma forma de reduzir o número de parâmetros é utilizar valores da saída e da entrada para determinar $y(t)$, ou seja, usar recorrência ou auto-regressão da saída. Os modelos NARMAX têm essas características (Aguirre, 2004).

3.5.1 Modelos NARMAX

O modelo NARMAX (Nonlinear AutoRegressive Moving Average with eXogenous variables) é uma derivação não-linear da representação linear ARMAX (AutoRegressive Moving Average with eXogenous variables). Os modelos NARMAX foram introduzidos em (Leontaritis & Billings, 1985a,b), e realizam o mapeamento não-linear das entradas e saídas passadas para a saída atual, $y(k)$, dada por:

$$y(k) = F[(y(k-1), \dots, y(k-n_y), u(k-1), \dots, u(k-n_u+1), e(k-1), \dots, e(k-n_e))] + e(k), \quad (3.18)$$

sendo que os sinais de entrada, saída e o ruído aditivo são representados, respectivamente por $u(k)$, $y(k)$ e $e(k)$, com atrasos máximos de n_u , n_y e n_e . $F(\cdot)$ é uma função não-linear, que pode assumir uma variedade de formas, tais como, racional e polinomial. Esta última é satisfatória em muitos casos. A justificativa teórica do uso de modelos NARMAX polinomiais pode ser encontrada em (Chen & Billings, 1989), onde é mostrado que existe uma função polinomial de grau ℓ que aproxima o sistema original com uma boa exatidão.

O modelo NARMAX polinomial é obtido expandindo-se a função F como um polinômio de grau ℓ , tal que:

$$y(k) = \theta_0 + \sum_{i_1=1}^n \theta_{i_1} x_{i_1}(k) + \sum_{i_1=1}^n \sum_{i_2=i_1}^n \theta_{i_1 i_2} x_{i_1}(k) x_{i_2}(k) + \dots + \sum_{i_1=1}^n \dots \sum_{i_\ell=i_{\ell-1}}^n \theta_{i_1 \dots i_\ell} x_{i_1}(k) \dots x_{i_\ell}(k) + e(k), \quad (3.19)$$

sendo que $n = n_y + n_u + n_e$, θ_i são os parâmetros a serem estimados e

O número total de termos candidatos, n_θ , em um modelo NARMAX polinomial mo-

$$x_1(k) = y(k-1), x_2(k) = y(k-2), \dots, x_{n_y+1}(k) = u(k-1), \dots, \\ x_{n_y+n_u+1}(k) = e(k-1), \dots, x_n(k) = e(k-n_e).$$

novariável é dado pela seguinte relação (Korenberg et al., 1988):

$$n_\theta = M + 1, \\ M = \sum_{i=1}^{\ell} n_i, \quad n_i = \frac{n_{i-1}(n_y + n_u + n_e + i - 1)}{i}, \quad n_0 = 1. \quad (3.20)$$

Verifica-se que o número de termos candidatos cresce rapidamente de acordo com o grau de não-linearidade ℓ e com os atrasos máximos da entrada, saída e do ruído, n_u , n_y e n_e , respectivamente. Quando o modelo é multivariável, n_θ , também dependerá do número de entradas e saídas.

Se todos os termos candidatos forem incluídos no modelo, como demonstra a equação (3.20), este poderá conter muitos termos, proporcionando, na maioria das vezes, mau condicionamento da matriz de regressores. Uma maneira de contornar este problema é determinar qual a estrutura que melhor se ajusta às informações disponíveis sobre o sistema em questão. Este procedimento é chamado de detecção ou escolha da estrutura. No contexto deste trabalho, objetiva-se obter a melhor estrutura que descreva aspectos dinâmicos e estáticos do sistema.

A generalização do modelo NARMAX (3.18) para o caso multivariável, é dada por (Billings et al., 1989):

$$y(k) = F[\mathbf{y}(k-1), \dots, \mathbf{y}(k-n_y), \mathbf{u}(k-1), \dots, \mathbf{u}(k-n_u), \\ \mathbf{e}(k-1), \dots, \mathbf{e}(k-n_e)] + \mathbf{e}(k), \quad (3.21)$$

sendo, $F(\cdot)$ uma função vetorial não-linear, $\mathbf{y}(k)$ o vetor de saída com atraso máximo n_y , $\mathbf{u}(k)$ o vetor de entrada com atraso máximo n_u e $\mathbf{e}(k)$ o vetor de ruído com atraso máximo n_e , tal que,

$$\mathbf{y}(k) = \begin{bmatrix} y_1(k) \\ y_2(k) \\ \vdots \\ y_m(k) \end{bmatrix}, \quad \mathbf{u}(k) = \begin{bmatrix} u_1(k) \\ u_2(k) \\ \vdots \\ u_r(k) \end{bmatrix}, \quad \mathbf{e}(k) = \begin{bmatrix} e_1(k) \\ e_2(k) \\ \vdots \\ e_m(k) \end{bmatrix}.$$

Um caso particular do modelo NARMAX polinomial monovariável, que não tem modelo de ruído, é conhecido como modelo NARX polinomial e é dado por:

$$y(k) = F^\ell[(y(k-1), \dots, y(k-n_y), u(k-1), \dots, u(k-n_u+1))] + e(k), \quad (3.22)$$

sendo F uma função polinomial de grau ℓ .

Técnicas para detecção da estrutura, validação dos modelos e métodos para a estimação dos parâmetros baseados no algoritmo dos mínimos quadrados têm sido desenvolvidos nas últimas décadas para ajustar os modelos NARX polinomiais a sistema não-lineares.

3.6 Detecção de Estrutura

A detecção da estrutura é uma etapa extremamente importante e difícil no processo de identificação de sistemas, visto que as características dinâmicas e estáticas que serão extraídas estão ligadas a estrutura do modelo estimado (Aguirre, 1994; Aguirre & Billings, 1995).

A presença de termos desnecessários no modelo pode proporcionar instabilidade numérica e ocasionar comportamentos dinâmicos espúrios ao sistema. Além do número excessivo de termos, efeitos espúrios podem ocorrer devido a sobreparametrização do modelo (Aguirre & Billings, 1995; Mendes & Billings, 1998).

No caso de modelos lineares, a escolha da sua estrutura se restringe, basicamente, à escolha do número de pólos e de zeros, bem como à determinação do atraso puro de tempo. Em representações não-lineares há inúmeras técnicas para cada tipo de representação. Neste trabalho as considerações a respeito da determinação de estrutura limitar-se-ão às representações lineares nos parâmetros, como os modelos polinomiais.

No contexto de identificação caixa-preta, alguns métodos podem ser usados para auxiliar na determinação da estrutura, tais como: o critério de informação de Akaike, AIC, e derivações (Akaike, 1974; Aguirre, 2004; Zou & Chon, 2004), agrupamento de termos (Aguirre, 2004), a busca ortogonal, OS, (ou busca ortogonal rápida - FOS)³ com taxa de redução de erro, ERR⁴, (Korenberg, 1985, 1989a,b; Billings et al., 1989; Chen, Cowan & Grant, 1991; Paarmann & Korenberg, 1992) e a busca ótima de parâmetros com índice de projeção de distância, OPS⁵ (Lu, Ju & Chon, 2001; Zou & Chon, 2004).

³Orthogonal Search e Fast Orthogonal Search.

⁴Error Reduction Ratio.

⁵Optimal Parameter Search.

3.6.1 Método de Busca Ortogonal - OS / Taxa de Redução de Erro - ERR

O método de *busca ortogonal* foi desenvolvido por Korenberg (1985) com o propósito de sistematicamente construir modelos polinomiais do tipo (N)ARMA(X) de sistemas dinâmicos com estrutura desconhecida. Um algoritmo muito semelhante foi usado anteriormente por Desrochers (1981) na construção de modelos estáticos de sistemas não-lineares. Korenberg et al. (1988) introduziram o uso do método de busca ortogonal numa simples e eficiente combinação de seleção de estrutura e estimação de parâmetros de sistemas estocásticos SISO⁶, sendo esta metodologia referida na literatura em geral como *Mínimos Quadrados Ortogonal* (OLS⁷). Billings, Korenberg & Chen (1988) apresentaram uma versão complementar do OLS, denominada *mínimos quadrados ortogonal com regressão direta*. Posteriormente, Billings et al. (1989) estenderam a metodologia para sistemas estocásticos MIMO⁸. Korenberg (1989a,b) ainda apresentou uma versão do método de busca ortogonal computacionalmente mais eficiente que é denominada *busca ortogonal rápida* (Fast Orthogonal Search, FOS).

A fim de descrever sucintamente a metodologia de busca ortogonal e o OLS, será considerado o seguinte modelo NARMAX geral na forma matricial:

$$\mathbf{y} = \Psi\boldsymbol{\theta}^T + \boldsymbol{\xi}, \quad (3.23)$$

em que $\boldsymbol{\xi} \in \mathbb{R}^{N \times 1}$ o vetor de erros de predição de um passo à frente cometidos ao se tentar explicar $\mathbf{y} \in \mathbb{R}^{N \times 1}$, vetor de saída do modelo, como $\Psi\boldsymbol{\theta}^T$. $\Psi \in \mathbb{R}^{N \times n_\theta}$ é uma matriz cujas colunas, $\boldsymbol{\psi}_i$, são os vetores de regressão (ou variáveis independentes) e $\boldsymbol{\theta} \in \mathbb{R}^{N \times 1}$ é o vetor de parâmetros a serem estimados.

Considerando que os diferentes regressores em Ψ são geralmente correlacionados, não fica claro como um regressor individual contribui na explicação da variável dependente \mathbf{y} . Portanto, faz-se necessário uma transformação dos regressores em Ψ em uma base de regressores ortogonais e posteriormente poder-se-á calcular a importância de cada regressor da base ortogonal. Cita-se, dentre outros, os procedimentos matemáticos de *Gram-Schmidt*, *Gram-Schmidt* modificado e a transformação de *Householder*⁹, os quais são usados na construção da base de regressores ortogonais¹⁰. A equação (3.23) poderá,

⁶Single Input, Single Output.

⁷Orthogonal Least Square.

⁸Multi-Input, Multi-Output.

⁹Veja detalhes desses e outros procedimentos em Aguirre (2004).

¹⁰(Korenberg, 1989a; Korenberg & Paarmann, 1991) ponderam que no algoritmo de busca ortogonal, a construção da base de regressores ortogonais, por meio do procedimento de Gram-Schmidt ou outros,

portanto, ser reescrita para um modelo ortogonalizado:

$$\mathbf{y} = W\mathbf{g}^T + \boldsymbol{\xi}, \quad (3.24)$$

em que $W \in \mathbb{R}^{N \times n_\theta}$ é a matriz de regressores mutuamente ortogonais com colunas \mathbf{w}_i (termos candidatos), e $\mathbf{g} \in \mathbb{R}^{N \times 1}$ é o vetor de coeficientes da expansão ortogonal a serem estimados. Chen et al. (1991) ressalta que o espaço formado pelo conjunto de regressores de base ortogonal, \mathbf{w}_i , é o mesmo espaço formado pelo conjunto dos regressores originais, $\boldsymbol{\psi}_i$, da equação (3.23), embora essa afirmação seja verdadeira se os regressores originais formam de fato uma *base*¹¹ (Chen, 1999).

A solução de mínimos quadrados ortogonal (OLS) de $\hat{\mathbf{g}}$ é dada por (Chen et al., 1991):

$$\hat{g}_i = \frac{\mathbf{w}_i^T \mathbf{y}}{\mathbf{w}_i^T \mathbf{w}_i}, \quad i \in [1, n_\theta], \quad (3.25)$$

sendo n_θ o número de regressores ortogonais. Maiores detalhes a cerca do estimador de mínimos quadrados podem ser encontrados na seção 3.7.

Os coeficientes \hat{g}_i são selecionados para minimizar o *erro quadrático médio* (mse) do vetor de saída \mathbf{y} , conforme equação abaixo:

$$N^{-1} \boldsymbol{\xi}^T \boldsymbol{\xi} = N^{-1} \mathbf{y}^T \mathbf{y} - N^{-1} \sum_{i=1}^{n_\theta} \hat{g}_i^2 \mathbf{w}_i^T \mathbf{w}_i, \quad (3.26)$$

em que N é a quantidade de dados de observação, $\boldsymbol{\xi}$ é o erro de predição 1 passo a frente, \hat{g}_i indica os elementos do vetor de parâmetros $\hat{\mathbf{g}}$, \mathbf{w}_i indica os regressores ortogonais (termos do modelo), n_θ é o total de regressores.

Pode-se mostrar facilmente a partir da equação (3.26), que a adição de termos $g_i \mathbf{w}_i$ no modelo (3.24) reduz o erro quadrático médio pela seguinte quantidade (Korenberg, 1985; Korenberg & Paarmann, 1991):

$$Q_i = \hat{g}_i^2 \mathbf{w}_i^T \mathbf{w}_i, \quad (3.27)$$

é considerada computacionalmente intensiva. O algoritmo de *busca ortogonal rápida* (FOS) evita o problema anterior explorando a fatorização de *Cholesky* para identificação do modelo. Ambos os métodos OS e FOS são equivalentes no que diz respeito ao resultado, e, de uma maneira explícita ou não, ambos ortogonalizam os regressores candidatos.

¹¹Um conjunto de vetores num determinado espaço vetorial, pode ser considerado uma base se forem linearmente independentes entre si (Chen, 1999).

em que \mathbf{w}_i indica o regressor incluído e \hat{g}_i o seu respectivo parâmetro. A idéia é avaliar Q_i para cada regressor candidato a ser adicionado ao modelo (3.24). Escolhe-se o candidato \mathbf{w}_i para o qual Q_i é maior, considerando que a adição desse termo resultará numa maior redução do erro quadrático médio (mse). Ao continuar-se a selecionar termos dessa maneira, pode-se construir modelos concisos, embora precisos, de sistemas reais, particularmente se um valor de limiar (*threshold*) é usado para rejeitar termos inadequados (Korenberg & Paarmann, 1991).

A quantidade Q_i pode ser normalizada com relação aos dados de saída (Billings et al., 1989), definindo-se, enfim, a *taxa de redução de erro* (ERR):

$$[ERR]_i = \frac{(\hat{g}_i^2 \mathbf{w}_i^T \mathbf{w}_i)}{\mathbf{y}^T \mathbf{y}}, \quad i \in [1, n_\theta]. \quad (3.28)$$

A taxa de redução de erro (*error reduction ratio*) ou ERR (Billings et al., 1989) atribui a cada termo candidato um valor correspondente à contribuição deste na explicação do valor quadrático médio dos dados de saída, ou seja, quantifica a redução no erro da saída do modelo devido à introdução de cada termo.

O ERR pode ser utilizado na determinação de estrutura de modelos polinomiais e RBFs. Escolhe-se o número de termos desejado através de uma ferramenta auxiliar, o critério de informação de Akaike por exemplo, e consideram-se aqueles que possuem os maiores valores de ERR.

3.6.2 Método de Busca Ótima de Parâmetros - OPS

O método de *busca ótima de parâmetros* (OPS) foi introduzido por Lu et al. (2001) como uma metodologia de seleção de estrutura e de estimação de parâmetros de modelos polinomiais, originalmente, do tipo ARX invariante no tempo. De acordo com Lu et al. (2001) e Zou & Chon (2004), métodos como FOS e GMDH¹² não fariam uma estimação tão precisa em condições em que os dados estejam livres de ruído e de seleção incorreta, *a priori*, da ordem do modelo. Esse problema seria atribuído ao fato do critério de busca dos métodos mencionados não perseguir o erro mínimo através de todos os subconjuntos possíveis de funções (termos) candidatas do modelo dentro do espaço funcional.

OPS é um critério de busca ótima baseado no princípio de *geometria afim*, que possibilita a estimação de parâmetros a despeito da seleção, *a priori*, incorreta da ordem

¹²Group Method of Data Handling.

do modelo. A geometria afim, um subconjunto da geometria Euclidiana, lida principalmente com *pontos*, *retas* e *incidências*¹³, e portanto não considera ângulos. O operador chamado *produto interno*¹⁴, que mapeia um par de vetores em um escalar, é próprio da geometria euclidiana e, conseqüentemente, as definições de ângulo e projeção ortogonal o são também.

Como foi visto na Seção 3.6.1, o algoritmo de busca ortogonal é baseado em conceitos da geometria euclidiana porque a ortogonalidade dos regressores candidatos do modelo a ser construído é necessária. OPS utiliza um critério de busca não-ortogonal de regressores candidatos do modelo. Lu et al. (2001) apontam uma desvantagem do método de busca ortogonal que pode ser melhor entendida por meio do seguinte exemplo.

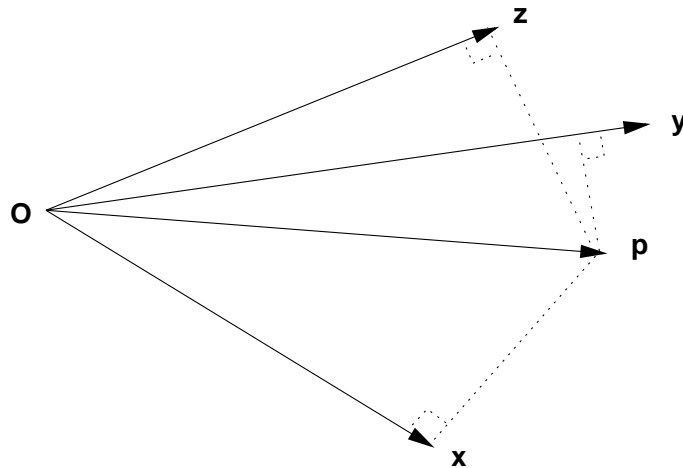


Figura 3.1: Projeção ortogonal encontra o ponto mais próximo do vetor \mathbf{p} no espaço vetorial formado pelos vetores base.

Um vetor \mathbf{p} está no espaço construído pelos vetores \mathbf{x} e \mathbf{y} , como mostrado na Figura 3.1. Se o ângulo formado por \mathbf{zop} for menor que os ângulos \mathbf{yop} e \mathbf{xop} , então o vetor \mathbf{z} seria escolhido num procedimento de busca ortogonal, embora ele não esteja no espaço construído pelos vetores \mathbf{x} e \mathbf{y} .

Em outras palavras, a projeção ortogonal encontra o ponto mais próximo a \mathbf{p} no espaço formado pelos vetores base, sem considerar se o vetor pertence ou não ao espaço vetorial construído por \mathbf{x} e \mathbf{y} . De fato, para ortogonalizar um conjunto de vetores, por meio do método *Gram-Schmidt* ou derivações, é necessário que estes formem uma *base*¹⁵

¹³Incidência: quando elementos como linhas, planos ou pontos coincidem.

¹⁴O produto interno entre dois vetores $\mathbf{v} \in \mathbb{R}^m$ e $\mathbf{w} \in \mathbb{R}^n$ é definido como: $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^T \mathbf{w}$.

¹⁵Um conjunto de vetores *linearmente independentes* em \mathbb{R}^n é chamado uma base se cada vetor em \mathbb{R}^n

(Boldrini, Costa, Figueiredo & Wetzler, 1980; Chen, 1999).

De acordo com Lu et al. (2001), o cenário discutido acima produziria um parâmetro estimado incorretamente com os métodos OS/FOS, quando a ordem do modelo fosse escolhida *a priori*. Além disso, Chen et al. (1991) afirmam que os regressores de um modelo escolhido *a priori*, são geralmente correlacionados entre si. Em outras palavras, eles seriam geralmente *linearmente dependentes* entre si. No algoritmo OPS, por ser baseado em uma busca não-ortogonal, esse tipo de cenário não ocorreria. Ressalta-se que se os vetores são todos perpendiculares entre si, então não há diferença entre os métodos de busca ortogonal e não-ortogonal.

Para descrever-se o algoritmo de busca não-ortogonal OPS, considera-se o seguinte modelo ARX invariante no tempo:

$$y(k) = \sum_{j=1}^{n_y} a_j y(k-j) + \sum_{i=0}^{n_u} b_i u(k-i) + \xi(k), \quad (3.29)$$

em que n_y e n_u são, respectivamente, as ordens máximas dos termos (ou regressores) de saída AR, $y(k-j)$, e das entradas (ou regressores) exógenas, $u(k-i)$. O erro de predição é $\xi(k)$ e os parâmetros que se deseja estimar é a_j e b_i representam. Esse modelo pode ser reescrito na seguinte forma matricial:

$$\mathbf{y} = \Psi \boldsymbol{\theta}^T + \boldsymbol{\xi}, \quad (3.30)$$

sendo $\Psi \in \mathbb{R}^{N \times n_\theta}$ a matriz de regressores e n_θ é a quantidade de regressores e ordem máxima do modelo.

A matriz Ψ ¹⁶ pode ser construída da seguinte maneira:

$$\Psi = [\boldsymbol{\psi}_1 \ \boldsymbol{\psi}_2 \ \cdots \ \boldsymbol{\psi}_{n_\theta}], \quad \text{em que} \quad n_\theta = n_y + n_u + 1.$$

$$\Psi = [\mathbf{y}_{(k-1)} \ \mathbf{u}_{(k)} \ \mathbf{y}_{(k-2)} \ \mathbf{x}_{(k-1)} \ \cdots \ \mathbf{y}_{(k-n_y)} \ \mathbf{x}_{(k-n_u)}],$$

pode ser expresso como uma combinação linear única do conjunto (Chen, 1999).

¹⁶É importante, neste ponto, perceber as diferenças entre notações como $y(k-1)$, \mathbf{y} e $\mathbf{y}_{(k-1)}$. Neste exemplo, \mathbf{y} é formado tomando-se várias observações *escalares* no tempo, $y(k-1)$ é um valor escalar tomado do vetor de observações \mathbf{y} no instante $k-1$ e $\mathbf{y}_{(k-1)}$ é o vetor das p observações correspondentes às p saídas, *todas tomadas no mesmo instante* $k-1$.

$$\Psi = \begin{bmatrix} y(0) & u(1) & y(-1) & u(0) & \cdots & y(1 - n_y) & u(1 - n_u) \\ y(1) & u(2) & y(0) & u(1) & \cdots & y(2 - n_y) & u(2 - n_u) \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ y(k-1) & u(k) & y(k-2) & u(k-1) & \cdots & y(k - n_y) & u(k - n_u) \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ y(N-1) & u(N) & y(N-2) & u(N-1) & \cdots & y(N - n_y) & u(N - n_u) \end{bmatrix}. \quad (3.31)$$

Para modelos Narmax, a matriz da equação (3.31) pode ser expandida para incluir produtos entre regressores. O capítulo 4 trata do uso dessa metodologia por modelos variantes no tempo (TV-OPS).

O primeiro passo para a estimação dos coeficientes do modelo da equação (3.30) é a seleção de regressores *linearmente independentes* do conjunto de regressores candidatos, Ψ .

Boldrini et al. (1980) definem *independência linear* de um conjunto de vetores como se segue. Seja V um espaço vetorial e $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \in V$ um conjunto de vetores. Diz-se que o conjunto $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ é linearmente independente, ou que os vetores $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ são linearmente independentes, se a condição abaixo for satisfeita:

$$\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \cdots + \lambda_n \mathbf{v}_n = 0 \text{ implique } \lambda_1 = \lambda_2 = \cdots = \lambda_n = 0,$$

dado que $\lambda_1, \lambda_2, \dots, \lambda_n$ são escalares. No caso em que exista algum $\lambda_i \neq 0$, diz-se que o conjunto $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ é linearmente *dependente* ou que os vetores o são.

Pode-se expressar a mesma verdade a cerca da independência linear de um conjunto de vetores por meio do seguinte teorema (Boldrini et al., 1980):

Teorema 3.6.1 *O conjunto de vetores $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\} \in V$ é linearmente dependente se, e somente se um destes vetores for uma combinação linear dos outros.*

Baseado no Teorema 3.6.1 e conforme sugerido por Lu et al. (2001) e Zou et al. (2003), a seleção de regressores linearmente independentes da equação (3.31) é feita conforme a seguinte estratégia: toma-se, inicialmente, o primeiro regressor candidato, $\boldsymbol{\psi}_1$, e o segundo regressor, $\boldsymbol{\psi}_2$, para verificar se eles são linearmente independentes. Para isso, faz-se um ajuste linear entre esses regressores por meio do estimador de mínimos quadrados¹⁷, como pode ser visto na equação abaixo:

¹⁷Ver detalhes sobre o estimador de mínimos quadrados na seção 3.7.

$$\begin{aligned}\boldsymbol{\psi}_2 &= \boldsymbol{\psi}_1 \mathbf{g}^T + \boldsymbol{\xi}, \\ \hat{\mathbf{g}} &= [\boldsymbol{\psi}_1^T \boldsymbol{\psi}_1]^{-1} \boldsymbol{\psi}_1^T \boldsymbol{\psi}_2,\end{aligned}\tag{3.32}$$

em que $\hat{\mathbf{g}} \in \mathbb{R}^{1 \times 1}$ é o vetor de coeficientes estimados para o ajuste linear entre os regressores, embora nesse primeiro passo $\hat{\mathbf{g}}$ seja um vetor de apenas um elemento. $\boldsymbol{\xi} \in \mathbb{R}^{N \times 1}$ é o vetor de erros do ajuste (ou predição). Calcula-se um *resíduo* de ajuste entre os regressores por meio da variância do erro:

$$\phi = \text{var}(\boldsymbol{\xi}) = \text{var}(\boldsymbol{\psi}_2 - \boldsymbol{\psi}_1 \hat{\mathbf{g}}^T),\tag{3.33}$$

sendo o operador, var , a variância da grandeza correspondente. Se o resíduo calculado, ϕ , for maior que o *limiar*¹⁸ pré-estabelecido, então $\boldsymbol{\psi}_2$ é considerado linearmente independente de $\boldsymbol{\psi}_1$. Senão, descarta-se $\boldsymbol{\psi}_2$ e seleciona-se um terceiro candidato $\boldsymbol{\psi}_3$, repetindo-se o procedimento. Porém, uma vez que $\boldsymbol{\psi}_2$ tenha sido considerado linearmente independente de $\boldsymbol{\psi}_1$, esses dois regressores são então usados consecutivamente para estimar a independência linear do próximo regressor $\boldsymbol{\psi}_3$. A equação (3.32) pode, então, ser reescrita como:

$$\begin{aligned}\boldsymbol{\psi}_3 &= W \mathbf{g}^T + \boldsymbol{\xi}, \\ \hat{\mathbf{g}} &= [W^T W]^{-1} W^T \boldsymbol{\psi}_3,\end{aligned}\tag{3.34}$$

em que $\hat{\mathbf{g}} \in \mathbb{R}^{2 \times 1}$ é o vetor de coeficientes de ajuste linear estimado entre o regressor $\boldsymbol{\psi}_3 \in \mathbb{R}^{N \times 1}$ e a matriz $W \in \mathbb{R}^{N \times 2}$ possui colunas formadas pelos regressores $\boldsymbol{\psi}_1$ e $\boldsymbol{\psi}_2$. Como foi feito na etapa inicial, calcula-se o resíduo do erro da equação (3.34):

$$\phi = \text{var}(\boldsymbol{\xi}) = \text{var}(\boldsymbol{\psi}_3 - W \hat{\mathbf{g}}^T).\tag{3.35}$$

Se o resíduo ϕ dessa etapa for menor que o valor de *limiar*, o regressor $\boldsymbol{\psi}_3$ deverá ser descartado e um próximo regressor $\boldsymbol{\psi}_{i+1}$ deverá ser escolhido. Caso o resíduo seja maior, o regressor $\boldsymbol{\psi}_3$ é considerado linearmente independente dos regressores $\boldsymbol{\psi}_1$ e $\boldsymbol{\psi}_2$, e esses três regressores passam a integrar a matriz W a fim de que o próximo regressor $\boldsymbol{\psi}_{i+1}$ seja então testado.

Este procedimento deve continuar até que todos os regressores mutuamente linear-

¹⁸Detalhes práticos a cerca do valor do limiar ϕ (threshold) será visto no capítulo 4.

mente independentes da matriz Ψ , sejam identificados para formar uma nova matriz de regressores de base linearmente independente:

$$W = [\mathbf{w}_1 \quad \mathbf{w}_2 \quad \mathbf{w}_3 \quad \cdots \quad \mathbf{w}_{ni}], \quad (3.36)$$

em que $\mathbf{w}_m \in \mathbb{R}^{N \times 1}$ são os vetores regressores que formam as colunas da matriz $W \in \mathbb{R}^{N \times ni}$ e ni é o número de regressores linearmente independentes e a ordem máxima do modelo.

Feito, então, a redução da estrutura do modelo da equação (3.30), esse mesmo poderá ser reescrito com a nova matriz de regressores linearmente independentes:

$$\mathbf{y} = W\mathbf{g}^T + \boldsymbol{\xi}, \quad (3.37)$$

em que $\mathbf{g} \in \mathbb{R}^{ni \times 1}$ é o vetor de parâmetros, do modelo, a ser estimado. A estimação é feita por meio do estimador de mínimos quadrados:

$$\hat{\mathbf{g}} = [W^T W]^{-1} W^T \mathbf{y}, \quad (3.38)$$

em que $\hat{\mathbf{g}} = [\hat{g}_1 \quad \hat{g}_2 \quad \hat{g}_3 \quad \cdots \quad \hat{g}_{ni}]^T$.

O segundo passo do algoritmo OPS é a identificação dos regressores mais relevantes para um modelo mais preciso da equação (3.29). Para determinar-se quais os termos candidatos são mais significativos, calcula-se o *índice de distância de projeção* (Lu et al., 2001; Zou et al., 2003):

$$c_m = \frac{1}{N} \sum_{k=1}^N \hat{g}_m^2 w_m(k)^2, \quad m = 1, 2, \dots, ni; \quad (3.39)$$

visto que o escalar c_m é o índice de distância de projeção, \hat{g}_m é o parâmetro estimado em (3.38) e refere-se ao termo regressor \mathbf{w}_m da matriz W da equação (3.36). N é o número de observações.

São escolhidos os regressores \mathbf{w}_m que reduzem significativamente o erro de predição, $\boldsymbol{\xi}$, da equação (3.37). Caso ocorra uma redução ou aumento desprezível do erro devido à inclusão de um regressor \mathbf{w}_m , pode-se excluir esse regressor do modelo. Essa última avaliação é feita por meio do índice de distância de projeção, sendo que os termos que possuem os maiores índices são mantidos e os demais excluídos.

O último procedimento do algoritmo OPS é a estimação final dos parâmetros da estrutura escolhida do modelo. Como no passo anterior, essa estimação é feita também por meio do estimador de mínimos quadrados.

3.6.3 O Critério de Informação de Akaike - AIC

O critério de informação de Akaike (AIC) é um critério estatístico, que verifica a redução na variância dos resíduos à medida que termos são acrescentados ao modelo, tal que (Akaike, 1974):

$$AIC(n_\theta) = N \ln[\phi_{\text{erro}}^2(n_\theta)] + 2n_\theta, \quad (3.40)$$

em que N é o número de observações, $\phi_{\text{erro}}^2(n_\theta)$ a variância dos resíduos (erro de modelagem) e n_θ o número de termos no modelo.

A primeira parcela da equação 3.40 quantifica a diminuição na variância dos resíduos resultante da inclusão de um termo, ao passo que a segunda parcela de (3.40) penaliza a inclusão de cada termo (Aguirre, 2004).

Há outros critérios de informação semelhantes a $AIC(n_\theta)$, sendo que eles se diferem basicamente em como ponderam as duas parcelas em (3.40) (Akaike, 1974). No capítulo 4 será considerado um critério de informação de Akaike modificado.

Ressalta-se que a taxa de redução de erro (ERR) ou a busca de parâmetros ótima (OPS), e os critérios de informação são complementares entre si. Uma possibilidade, por exemplo, seria o uso da ERR ou OPS para ordenar hierarquicamente os termos candidatos do modelo. Portanto, tendo-se ordenado os termos, poder-se-ia empregar um critério de informação para efetuar o *corte* necessário para a escolha da ordem do modelo representativo.

3.7 Estimação de Parâmetros

Como visto na seção 3.6, após determinar qual o tipo de representação e qual a sua estrutura, o próximo passo no processo de identificação de sistemas, consistiu em estimar os parâmetros do modelo de modo a minimizar a diferença entre a predição e a saída estimada pelo modelo.

Deseja-se nesta seção abordar sucintamente a metodologia de mínimos quadrados, por ser esta muito usada em estimação de parâmetros de modelos polinomiais.

Inicialmente, tem-se que o modelo ARX dado pela equação (3.19), pode ser representado como:

$$y(k) = \sum_{i=1}^{n_\theta} \psi_i(k-1)\theta_i + e(k) \quad (3.41)$$

sendo $y(k)$ a saída atual do modelo, $\psi(k-1)$ são os termos que constituem a matriz de regressores Ψ tomadas até o instante $k-1$, θ os parâmetros a serem estimados e $e(k)$ o ruído. O número de termos no modelo é representado por n_θ .

Desprezando o termo de ruído, pode-se reescrever (3.41) em sua forma matricial:

$$\mathbf{y} = \Psi\boldsymbol{\theta}, \quad (3.42)$$

sendo que,

$$\mathbf{y} = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix}, \quad \Psi = \begin{bmatrix} \Psi^T(0) \\ \Psi^T(1) \\ \vdots \\ \Psi^T(N-1) \end{bmatrix}.$$

e $\boldsymbol{\theta}$ o vetor de parâmetros.

A saída do modelo, predita um passo à frente é dada por¹⁹:

$$\begin{aligned} \hat{y}(k) &= \sum_{i=1}^{n_\theta} \psi_i(k-1)\hat{\theta}_i \\ &= \Psi^T(k-1)\hat{\boldsymbol{\theta}}. \end{aligned} \quad (3.43)$$

O erro entre o valor estimado e o valor medido da saída é denominado resíduo e é representado por:

$$\boldsymbol{\xi}(k) = \mathbf{y}(k) - \hat{\mathbf{y}}(k) = \mathbf{y}(k) - \Psi^T(k-1)\hat{\boldsymbol{\theta}}. \quad (3.44)$$

Utilizando-se um número maior de amostras, N , do que o número de regressores n_θ , tem-se um sistema de equações sobre determinado, resultando em uma matriz Ψ não quadrada, o que impossibilita a estimação de $\boldsymbol{\theta}$ a partir da inversão de Ψ em (3.42). Dessa forma, define-se um algoritmo para a estimação de $\hat{\boldsymbol{\theta}}$ que minimize a soma dos quadrados

¹⁹O símbolo $\hat{}$ significa que são valores a serem estimados.

do erro dado pela seguinte função custo:

$$J_{MQ} = \sum_{k=1}^N \xi(k|k-1, \hat{\boldsymbol{\theta}})^2 = \boldsymbol{\xi}^T \boldsymbol{\xi} = \|\boldsymbol{\xi}\|^2, \quad (3.45)$$

sendo que $\xi(k|k-1, \hat{\boldsymbol{\theta}})$ é o erro de predição (ou resíduo) cometido no instante k , ao fazer a predição com informação apenas até o instante $k-1$, usando o vetor estimado $\hat{\boldsymbol{\theta}}$.

Substituindo-se (3.44) em (3.45), tem-se:

$$\begin{aligned} J &= (\mathbf{y} - \Psi \hat{\boldsymbol{\theta}})^T (\mathbf{y} - \Psi \hat{\boldsymbol{\theta}}) \\ &= \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \Psi \hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^T \Psi^T \mathbf{y} - \hat{\boldsymbol{\theta}}^T \Psi^T \Psi \hat{\boldsymbol{\theta}}. \end{aligned} \quad (3.46)$$

A minimização da função custo J_{MQ} em relação ao vetor de parâmetros $\hat{\boldsymbol{\theta}}$ é feita resolvendo-se $\partial J_{MQ} / \partial \hat{\boldsymbol{\theta}} = 0$, cuja solução é dada por:

$$\hat{\boldsymbol{\theta}}_{MQ} = (\Psi^T \Psi)^{-1} \Psi^T \mathbf{y}. \quad (3.47)$$

A equação (3.47) é denominada estimador dos mínimos quadrados clássico.

Se a matriz de covariância dos regressores for singular ou mal condicionada, esta estará sujeita a problemas numéricos que podem afetar a estabilidade do algoritmo dos mínimos quadrados²⁰. Para contornar este problema, pode-se utilizar métodos numéricos que ortogonalizam a matriz Ψ , evitando então a correlação entre regressores.

3.7.1 Validação do Modelo

A validação dinâmica tem por finalidade verificar se o modelo estimado é capaz de representar a dinâmica do sistema em questão. A capacidade de generalização de um modelo deve ser analisada usando uma massa de dados de validação diferente da massa de dados de identificação. Dessa forma, é possível verificar se o modelo consegue explicar uma outra massa de dados do mesmo sistema.

Para validar um modelo dinamicamente, pode-se usar a simulação livre do mesmo, que consiste em usar um conjunto de dados do sistema, os de validação, e as predições passadas, na matriz de regressores.

²⁰O mau condicionamento é devido à alta correlação entre as colunas dos regressores da matriz Ψ .

Metodologias Para Identificação de Sistemas Dinâmicos Variantes no Tempo

Aborda-se neste capítulo o problema da modelagem paramétrica de sistemas dinâmicos variantes no tempo. Por consequência, a série temporal gerada por tais sistemas é não-estacionária, conforme visto no Capítulo 2. Apresentam-se três metodologias para análise de sistemas variantes no tempo. Investiga-se o contexto teórico e algumas peculiaridades de cada método. Exemplos numéricos simulados são usados nesta abordagem.

São estes o métodos:

1. *Time-Varying Optimal Parameter Search, TVOPS*, a seleção de estrutura de modelos polinomiais é feita por meio de um algoritmo variante do método OPS (ver Seção 3.6.2), e a estimação *off-line* dos coeficientes do modelo é feita por meio da expansão deles em um conjunto finito de múltiplas funções base;
2. *Estimador Recursivo de Mínimos Quadrados com Fator de Esquecimento Variável, RLSVFF*, em que os parâmetros de um modelo polinomial pré-escolhido são estimados recursivamente no tempo de observação e o fator de esquecimento variável é empregado como resposta adaptativa às variações da dinâmica;
3. *Estimador em Batelada de Mínimos Quadrados com Janelas Deslizantes de Tamanho Variável, VLSWBLs*, em que os parâmetros são estimados por meio do estimador de mínimos quadrados em blocos ou janelas de observações cujos tamanhos variam para adaptar os parâmetros às variações na dinâmica subjacente.

4.1 Time-Varying Optimal Parameter Search

Como visto na Seção 3.6.2, Lu et al. (2001) propuseram um algoritmo de identificação de modelos polinomiais invariantes no tempo (TIV-Time InVariant). Esse algoritmo, denominado *optimal parameter search (OPS)*, é baseado em conceitos da geometria *afim*, sobre os quais a característica saliente do algoritmo é a remoção de regressores linearmente dependentes do *pool* de regressores candidatos ou da matriz de regressores. OPS estabelece, portanto, um critério de determinação de estrutura ou seleção da ordem do modelo proposto.

Considerando a possibilidade de caracterização da dinâmica de um sistema variante no tempo por meio da variação de seus parâmetros, Zou et al. (2003) adaptaram o algoritmo OPS para ser usado em sistemas com essas características. A estratégia é fazer uma expansão de cada parâmetro em um conjunto finito de seqüências de uma função base¹ ortogonal, de maneira que os parâmetros se tornem invariantes no tempo. Complementando essa metodologia, Zou & Chon (2004) apresentaram uma extensão da técnica de expansão de parâmetros, sugerindo o uso de múltiplas funções base. Essa última estratégia, denominada *Time-Varying Optimal Parameter Search (TV-OPS)*, visa o rastreamento de diferentes comportamentos dinâmicos do sistema em observação.

Essa manipulação dos parâmetros leva a dois benefícios significativos: *i)* os parâmetros *variantes* no tempo se tornam *invariantes* no tempo, o que torna a tarefa de estimação analiticamente viável; *ii)* uma redução no número de parâmetros necessários para rastrear cada coeficiente pode ser obtida (Zou & Chon, 2004; Zou et al., 2003; Ralston, Zoubir & Boashash, 1996).

A literatura aponta para dois problemas básicos com esse tipo de abordagem. (Tzatsanis & Giannakis, 1993; Wei & Billings, 2002):

- a escolha dos tipos de função base;
- a seleção dos termos significantes dentre as famílias de função base.

Coifman & Wickerhauser (1992) utilizaram funções base trigonométricas (senos e cossenos) e pacotes wavelet² para a expansão dos coeficientes variantes no tempo. Para

¹Em Análise Funcional, ramo da matemática que estuda os espaços de funções e suas aplicações, um espaço funcional pode ser visto como um espaço vetorial de dimensão infinita cujos vetores base são as *funções base*. Isso significa que cada função no espaço funcional pode ser composta por uma combinação linear dessas *funções base*.

²Wavelet Packets.

seleção das melhores bases, foi usado um algoritmo baseado na entropia.

Tsatsanis & Giannakis (1993) propuseram o uso de funções base do tipo wavelet para a expansão dos coeficientes e um teste estatístico (Teste F de Åström) para determinar os termos mais relevantes.

Observa-se em Ralston et al. (1996) o uso de funções Walsh e Slepian (ou *discrete prolate spheroidal*) na estimação de parâmetros de séries produzidas por modelos *Hammerstein* variantes no tempo. Nesse caso foi feito um truncamento empírico no número de termos das funções.

No trabalho de Wei & Billings (2002) pode-se também encontrar o uso de funções wavelets de multiresolução na expansão dos coeficientes variantes no tempo e dos algoritmos OLS (Orthogonal Least Square) e ERR (Error Reduction Ratio) para definição do número de termos relevantes da função.

Pode-se encontrar na literatura exemplos de outras funções base como polinômios de *Legendre*, de *Chebyshev* e *Laguerre* dentre outros, na identificação de sistemas variantes no tempo (Hwang & Shih, 1982; Grenier, 1983; Mohan & Datta, 1988; Zou et al., 2003).

Na metodologia TV-OPS, tenta-se minimizar os problemas básicos conhecidos com o uso de funções base. Pode-se aliviar o problema inicial de escolha do tipo de função base, por meio do emprego de múltiplos tipos de função base. Além disso, emprega-se um critério denominado *critério de informação de Akaike modificado* para determinar quais os termos mais relevantes das funções base escolhidas.

Reassalta-se que o uso de funções base na modelagem e identificação de sistemas não é recente. Citam-se, por exemplo, os trabalhos clássicos de Lee (1933) e Wiener (1949). A literatura também aponta o emprego de tais funções em modelagem e identificação de sistemas invariantes no tempo. Considerando sistemas não-lineares e invariantes no tempo, cita-se, por exemplo o uso de polinômios discretos de *Kautz*, de *Laguerre* e outros, na aproximação de modelos de Wiener/Volterra (Tanguy, Morvan, Vilbé, & Calvez, 2002; Campello, Fávier & Amaral, 2004). No caso de sistemas lineares e invariante no tempo, cita-se o trabalho de Van Den Hof, Heuberger & Bokor (1995), que estudaram um método de identificação via mínimos quadrados, o qual estima um número finito de coeficientes numa expansão em série da função de transferência, cuja expansão é feita em termos de funções base *generalizadas*.

4.1.1 Funções Base

Dentre as classes de funções base disponíveis, utilizou-se nesse trabalho funções polinomiais de *Legendre* e funções de *Walsh*. Normalmente a escolha dessas funções é feita baseada em algum conhecimento *a priori* da dinâmica do sistema em análise. Este conhecimento nem sempre está disponível. Os polinômios de Legendre têm sido usados para rastrear parâmetros que variam lenta e suavemente (Grenier, 1983; Mohan & Datta, 1988; Zou et al., 2003). Para parâmetros que variam mais abruptamente, funções de Walsh podem ser mais adequadas (Stolen & Hardness, 1994; Ralston et al., 1996; Zou et al., 2003). Definem-se variações abruptas como mudanças em características que ocorrem rapidamente com relação ao período de amostragem das observações, podendo até serem instantâneas (Basseville & Nikiforov, 1993).

Considerando que muitos sinais não-estacionários, como sinais fisiológicos e séries econômicas, exibem dinâmica que se caracteriza por variar lenta e rapidamente em diferentes partes do sinal, parece razoável utilizar conjuntamente as funções de Legendre e Walsh para rastrear parâmetros variantes no tempo com essas características.

Funções Polinomiais de Legendre.

Adrien Marie Legendre (1752-1833) foi um matemático francês que em 1782 determinou, por meio de uma equação, a força de atração para certos sólidos de revolução³ (Abramowitz & Stegun, 1972). Essa equação é conhecida atualmente como *equação diferencial de Legendre* e é apresentada a seguir:

$$(1 - x^2)\ddot{y} - 2x\dot{y} + n(n + 1)y = 0. \quad (4.1)$$

A equação diferencial de Legendre (4.1) pode ser resolvida usando-se o método padrão de séries de potência. A solução é finita (ou seja, as séries convergem) dado que $|x| < 1$. Além disso, para $x = \pm 1$ a solução é finita se n é um número inteiro não negativo. Nesse caso, as soluções formam uma seqüência de polinômios chamados *polinômios de Legendre*. Cada polinômio de Legendre, aqui denotado por $P_n(x)$, é um polinômio de grau n (Abramowitz & Stegun, 1972).

³Em 1784 Legendre publicou a obra *L'attraction des ellipsoïdes*. Ele introduziu as chamadas Funções de Legendre e as séries de potência para determinar a atração de um elipsóide a qualquer ponto exterior.

Os polinômios de Legendre $P_n(x)$ podem ser expressos pela fórmula de *Rodrigues*:

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n, \quad (4.2)$$

em que $x \in \mathbb{R}$ é um valor qualquer no intervalo $-1 \leq x \leq 1$ e $n \in \mathbb{Z}^+$ é o grau do polinômio, ou seja $n = 0, 1, 2, \dots$.

Em seguida, apresenta-se alguns exemplos de polinômios de Legendre:

$$\begin{aligned} P_0(x) &= 1, \\ P_1(x) &= x, \\ P_2(x) &= \frac{1}{2}(3x^2 - 1), \\ P_3(x) &= \frac{1}{2}(5x^3 - 3x), \\ P_4(x) &= \frac{1}{8}(35x^4 - 30x^2 + 3). \end{aligned} \quad (4.3)$$

Pode-se ver também na Figura 4.1 os gráficos das primeiras funções polinomiais de Legendre, conforme a equação (4.3).

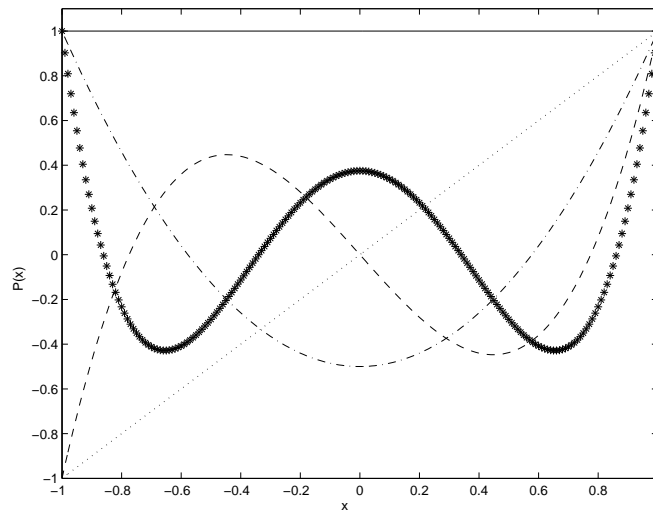


Figura 4.1: Polinômios de Legendre $P_n(x)$, para grau $n = (-) 0, (.) 1, (-.) 2, (- -) 3, (**) 4$.

Na equação a seguir, mostra-se uma relação de recorrência de polinômios, $P_n(x)$, que pode ser usada para gerar um dado polinômio de grau n , sendo $n > 1$:

$$(n + 1)P_{n+1}(x) = (2n + 1)xP_n(x) - nP_{n-1}(x). \quad (4.4)$$

Uma importante propriedade das funções polinomiais de Legendre é que as mesmas são *ortogonais*, com respeito ao produto interno no espaço L^2 , no intervalo $-1 \leq x \leq 1$:

$$\int_{-1}^1 P_m(x)P_n(x) dx = \frac{2}{2n + 1}\delta_{mn}, \quad (4.5)$$

em que δ_{mn} denota o delta de *Kronecker*⁴.

De fato, uma maneira alternativa para gerar-se polinômios de Legendre é por meio do método *Gram-Schmidt* sobre os polinômios $(1, x, x^2, \dots)$ com relação ao produto interno da equação (4.5) (Abramowitz & Stegun, 1972). Neste trabalho foi usada a função *legendre()* do software Matlab⁵, a qual emprega uma variante do método Gram-Schmidt para gerar os polinômios de Legendre.

No contexto do método *TV-OPS*, o intervalo contínuo $x \in [-1,1]$ é dividido em intervalos discretos, os quais são equivalentes aos instantes discretos de observação, k , considerados no processo. Cita-se o seguinte exemplo: dado uma série temporal de 500 observações, deseja-se modelar o processo que gerou essa série fazendo-se a expansão dos coeficientes do modelo escolhido por meio de funções polinomiais de Legendre. Dentre as possibilidades, escolhe-se uma função polinomial de grau 2, $P_2(x)$. Para essa função, são gerados 500 valores de x igualmente espaçados dentro do intervalo $[-1,1]$ e o valor da função polinomial, $P_2(x)$, é avaliada para cada elemento x . O polinômio $P_2(x)$ para os 500 valores de x pode ser verificado na Figura 4.1.

Funções de Walsh.

As funções e séries de Walsh foram apresentadas por J. L. Walsh em 1923 e desde então têm despertado o interesse de diversas áreas científicas tais como a espectroscopia, a sismologia, processamento e reconhecimento de voz, a medicina, a detecção remota (radares e sonares), a comunicação digital e a modelagem e identificação de sistemas (Castro, 1996).

Tais funções podem ser definidas como um conjunto ordenado de pulsos retangulares apresentando apenas duas amplitudes possíveis, $+1$ e -1 , de maneira que as transições

⁴Delta de Kronecker, δ_{mn} , é uma função de duas variáveis, m e n , que é igual a zero quando as variáveis tem valores diferentes e igual a um quando as variáveis são iguais.

⁵MATLAB Version 6.5.0.180913a (R13).

ocorrem somente em instantes fixos e em um intervalo limitado. Nesta seção as funções de Walsh serão denotadas por $wal(g,t)$, sendo t a posição no intervalo de definição e g é um número de ordem que se encontra relacionado com a frequência. Walsh (1923) propôs que, para uma dada função $wal(g,t)$, g fosse igual ao número total de mudanças de sinal algébrico existentes.

Dentre as propriedades das funções de Walsh, enunciam-se em seguida as mais relevantes no que diz respeito às suas aplicações no presente trabalho: *i)* são *ortogonais* entre si; *ii)* quando definidas no intervalo $t \in [0,1]$, são *ortonormais*⁶ entre si. As demonstrações podem ser verificadas em Paley (1932), Chien (1975) e Castro (1996).

São vários os métodos usados na geração das funções de Walsh. Citam-se os mais importantes: *i)* aplicação da fórmula de recorrência de Harmuth (Harmuth, 1972); *ii)* aplicação da fórmula de Chien (Chien, 1975); *iii)* produtos sucessivos de funções de Rademacher (Davies, 1972; Paley, 1932); *iv)* montagem da matriz de Hadamard (Henderson, 1970);

Optou-se neste trabalho pela utilização do método de montagem da matriz de Hadamard para gerar funções de Walsh no intervalo de definição $t \in [0,1]$. A escolha desse método é devido à simplicidade computacional envolvida e a escolha do intervalo de definição está relacionada à propriedade de ortogonalidade supracitada.

A matriz de Hadamard é um tipo de matriz quadrada na qual cada elemento assume apenas dois valores possíveis, $+1$ e -1 , sendo suas linhas (ou colunas) ortogonais entre si. Tendo-se uma matriz de Hadamard, é sempre possível permutar linhas (ou colunas) e/ou trocar o sinal de todos os elementos pertencentes a uma dada linha (ou coluna) sem alterar a ortogonalidade existente. Essas características tornam possível a obtenção de uma matriz simétrica, conhecida como *forma normal da matriz de Hadamard*, na qual a primeira linha e a primeira coluna contêm apenas elementos positivos ($+1$). A matriz de Hadamard de ordem mais baixa é:

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \quad (4.6)$$

Matrizes de ordem superior são obtidas por meio da equação:

⁶Na álgebra linear, um conjunto de vetores $\{\phi_n\}$ é chamado *ortonormal*, se $\int_a^b \phi_m \phi_n = 0$ para $m \neq n$ e $\|\phi_n\| = 1$. A base que forma um conjunto ortonormal é chamada base ortonormal.

$$H_n = H_{\frac{n}{2}} \otimes H_2, \quad (4.7)$$

em que $H_n \in \mathbb{R}^{n \times n}$ é a matriz de Hadamard de ordem n , sendo n uma potência de 2. O símbolo \otimes denota o produto de *Kronecker*. O produto de Kronecker implica a substituição de cada elemento +1 da matriz $H_{\frac{n}{2}}$ pela matriz H_2 e cada elemento -1 pela matriz $-H_2$. A matriz de quarta ordem, por exemplo, é obtida por meio da seguinte equação:

$$H_4 = H_2 \otimes H_2 = \left[\begin{array}{cc|cc} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ \hline 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{array} \right] \quad (4.8)$$

Para gerar uma certa função de Walsh, $wal(g,t)$, considera-se o intervalo de definição das funções de Walsh $t \in [0,1]$, dividido em g subintervalos que terão dimensão igual quando g for múltiplo de 2. Uma matriz H_n é construída de maneira que $n = g$, se g for múltiplo de 2, ou n será o próximo número múltiplo de 2 maior que g . A linha da matriz H_n que possui g mudanças de sinal estabelece então qual o valor (+1 ou -1) que a função toma ao longo daqueles g subintervalos. Vê-se alguns exemplos na equação (4.8). Pode-se observar também na Figura 4.2 o gráfico de tais funções.

É oportuno ressaltar que a natureza matemática das funções de Walsh é tal que não permite a existência de uma única forma de ordenação (Yuen, 1972). Citam-se as principais possibilidades de ordenação (Yuen, 1972; Castro, 1996): *i*) seqüencial; *ii*) diádica; *iii*) natural.

Neste trabalho emprega-se a ordenação seqüencial proposta por Walsh (1923), em que as funções são ordenadas conforme o número de mudanças de sinal verificado em cada uma delas (observar Figura 4.2).

No contexto do método *TV-OPS*, o intervalo contínuo $t \in [0,1]$ é dividido em intervalos discretos, os quais são equivalentes aos instantes de observação k considerados no processo. Cita-se o seguinte exemplo: dado uma série temporal de 1000 amostras, deseja-se modelar o processo que gerou essa série fazendo-se a expansão dos coeficientes do modelo escolhido por meio de funções de Walsh. Dentre as possibilidades, escolhe-se uma função $wal(1,t)$. Para esta função, são geradas 1000 amostras dentro do intervalo $t = [0,1]$, sendo os primeiros 500 valores iguais +1 e os últimos iguais a -1, conforme é indicado na Figura 4.2.

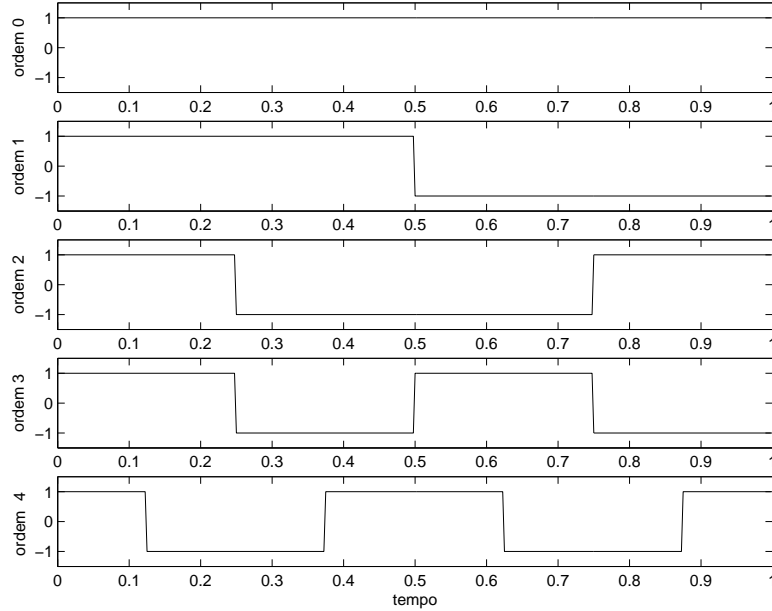


Figura 4.2: Funções de Walsh, $wal(g,t)$, segundo a ordenação seqüencial $g = 0,1,2,3,4$ e $t \in [0,1]$.

4.1.2 Descrição da Metodologia TV-OPS.

Para descrever o método TV-OPS, considere-se inicialmente o seguinte modelo ARX:

$$y(k) = \sum_{i=1}^{n_y} a(i,k)y(k-i) + \sum_{j=1}^{n_u} b(j,k)u(k-j) + e(k), \quad (4.9)$$

em que $a(i,k)$ e $b(j,k)$ são parâmetros variantes no tempo a serem determinados. As constantes n_y e n_u são as ordens máximas do modelo sugerido *a priori* e serão consideradas invariantes no tempo. O ruído do processo, incertezas ou dinâmica não modelada são representados por $e(k)$.

O método propõe expandir os coeficientes $a(i,k)$ e $b(j,k)$ em múltiplos conjuntos de funções base, como:

$$a(i,k) = \sum_{l=1}^L \sum_{r_l=0}^{V_l} \alpha^{(l)}(i,r_l) \pi_{r_l}^{(l)}(k), \quad (4.10)$$

$$b(j,k) = \sum_{l=1}^L \sum_{r_l=0}^{V_l} \beta^{(l)}(j,r_l) \pi_{r_l}^{(l)}(k), \quad (4.11)$$

em que k indica o instante atual de observação, α e β representam os termos de expansão dos parâmetros, π_{r_l} representa um tipo qualquer de função base, V_l é o valor máximo de r_l , que indica grau ou ordem da função base ($r_l = 0, 2, \dots, V_l$). Um conjunto ou tipo específico de função base é indicado pelos valores de $l = 1, 2, \dots, L$. Neste trabalho $L = 2$, pois serão usados dois conjuntos de funções base: *Legendre* e *Walsh*.

Substituindo as equações (4.10) e (4.11) na equação (4.9), obtém-se:

$$y(k) = \sum_{l=1}^L \left[\sum_{i=1}^{n_y} \sum_{r_l=0}^{V_l} \alpha^{(l)}(i,r_l) \pi_{r_l}^{(l)}(k) y(k-i) + \sum_{j=1}^{n_u} \sum_{r_l=0}^{V_l} \beta^{(l)}(j,r_l) \pi_{r_l}^{(l)}(k) u(k-j) \right] + e(k). \quad (4.12)$$

No que se segue, será utilizada a seguinte mudança de variáveis:

$$\begin{aligned} y_{r_l}^{(l)}(k-i) &= \pi_{r_l}^{(l)}(k) y(k-i), \\ u_{r_l}^{(l)}(k-i) &= \pi_{r_l}^{(l)}(k) u(k-i), \end{aligned} \quad (4.13)$$

o que resulta em:

$$y(k) = \sum_{l=1}^L \left[\sum_{i=1}^{n_y} \sum_{r_l=0}^{V_l} \alpha^{(l)}(i,r_l) y_{r_l}^{(l)}(k-i) + \sum_{j=1}^{n_u} \sum_{r_l=0}^{V_l} \beta^{(l)}(j,r_l) u_{r_l}^{(l)}(k-j) \right] + e(k). \quad (4.14)$$

A equação (4.14) mostra que o modelo ARX variante no tempo da equação (4.9), pode agora ser considerado invariante no tempo, sendo que $\alpha^{(l)}(i,r_l)$ e $\beta^{(l)}(j,r_l)$ não são função do tempo (k).

Seguindo um procedimento semelhante ao descrito na Seção 3.6.2, o primeiro passo para estimar os coeficientes α e β do modelo (4.14) é a montagem da seguinte matriz de regressores expandidos⁷:

$$M = \left[Y_{(k-1)}^{(1)} \cdots Y_{(k-1)}^{(L)}, U_{(k)}^{(1)} \cdots U_{(k)}^{(L)}, Y_{(k-i)}^{(1)} \cdots Y_{(k-i)}^{(L)}, U_{(k-j)}^{(1)} \cdots U_{(k-j)}^{(L)} \right], \quad (4.15)$$

⁷Um regressor expandido é formado pelo produto entre coordenadas de uma função base $\pi_{r_l}^{(l)}$ e um regressor original do modelo (4.9), formando os regressores $y_{r_l}^{(l)}$ e/ou $u_{r_l}^{(l)}$. Essa transformação pode ser verificada na equação (4.13).

sendo $M \in \mathbb{R}^{N \times N_{rc}}$ (N é o número de observações e N_{rc} o total de regressores expandidos candidatos) uma matriz cujos elementos $Y_{(k-i)}^{(l)} \in \mathbb{R}^{N \times V_l}$ e $U_{(k-j)}^{(l)} \in \mathbb{R}^{N \times V_l}$ são matrizes construídas conforme a definição abaixo:

$$Y_{(k-i)}^{(l)} = \begin{bmatrix} \mathbf{y}_1^{(l)} & \mathbf{y}_2^{(l)} & \mathbf{y}_3^{(l)} & \cdots & \mathbf{y}_{V_l}^{(l)} \end{bmatrix}, \quad (4.16)$$

$$U_{(k-j)}^{(l)} = \begin{bmatrix} \mathbf{u}_1^{(l)} & \mathbf{u}_2^{(l)} & \mathbf{u}_3^{(l)} & \cdots & \mathbf{u}_{V_l}^{(l)} \end{bmatrix}, \quad (4.17)$$

sendo $\mathbf{y}_{r_l}^{(l)} \in \mathbb{R}^{N \times 1}$ e $\mathbf{u}_{r_l}^{(l)} \in \mathbb{R}^{N \times 1}$ chamados regressores expandidos. Pode-se ver em seguida suas definições:

$$\mathbf{y}_{r_l}^{(l)} = \left[\pi_{r_l}^{(l)}(1)y(1-i), \dots, \pi_{r_l}^{(l)}(N)y(N-i) \right]^T, \quad (4.18)$$

$$\mathbf{u}_{r_l}^{(l)} = \left[\pi_{r_l}^{(l)}(1)u(1-j), \dots, \pi_{r_l}^{(l)}(N)u(N-j) \right]^T. \quad (4.19)$$

O total de regressores expandidos candidatos, N_{rc} , na matriz M é calculado conforme a equação (4.20):

$$N_{rc} = n_y \times \sum_{l=1}^L (V_l) + (n_u + 1) \times \sum_{l=1}^L (V_l). \quad (4.20)$$

Conforme o método OPS descrito na Seção 3.6.2, os regressores candidatos da matriz M linearmente independentes são escolhidos para formar uma nova matriz:

$$W = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{n_i}], \quad (4.21)$$

em que $\mathbf{w}_{n_i} \in \mathbb{R}^{N \times 1}$ são os regressores candidatos expandidos linearmente independentes. Vale ressaltar que o regressor \mathbf{w}_{n_i} é o produto de uma função base com os regressores originais de entrada ou saída:

$$\begin{aligned} \mathbf{w}_{n_i} &= \boldsymbol{\pi}_{r_l}^{(l)}(k)\mathbf{y}(k-i) \\ &\text{ou} \\ \mathbf{w}_{n_i} &= \boldsymbol{\pi}_{r_l}^{(l)}(k)\mathbf{u}(k-j), \end{aligned} \quad (4.22)$$

em que $r_l \in [0, V_l]$, $i \in [1, n_y]$, $j \in [0, n_u]$ e $l \in [1, L]$.

Conforme a equação (3.38), a matriz W de regressores linearmente independentes é usada pelo estimador de mínimos quadrados na estimação de todos os coeficientes α e β

da equação (4.14).

O vetor de parâmetros estimados por mínimos quadrados é :

$$\hat{\mathbf{g}} = [W^T W]^{-1} W^T \mathbf{y}, \quad (4.23)$$

em que $\hat{\mathbf{g}} = [\hat{g}_1 \quad \hat{g}_2 \quad \hat{g}_3 \quad \cdots \quad \hat{g}_{ni}]^T$ e \hat{g}_{ni} são os coeficientes estimados α ou β .

Em seguida deve-se determinar os termos significativos do modelo (4.14), o qual foi sugerido *a priori*. Para se determinar a relevância dos regressores candidatos, calcula-se o *índice de projeção de distância*. Esse índice se refere à distância de projeção de um regressor a outro linearmente independente:

$$c_m = \frac{1}{N} \sum_{k=1}^N \hat{g}_m^2 w_m(k)^2, \quad m = 1, 2, \dots, ni; \quad (4.24)$$

visto que o escalar c_m é o índice de distância de projeção, \hat{g}_m é a estimação do parâmetro (α ou β) do regressor \mathbf{w}_m , conforme equações (4.21) e (4.23). N é o número de observações.

Note que a operação da equação (4.24) envolve a soma dos índices de distância de projeção, c_m , de cada regressor expandido \mathbf{w} obtido a partir de um mesmo regressor original $\mathbf{y}(k-i)$ ou $\mathbf{u}(k-i)$ mas com uma função base $\boldsymbol{\pi}_{r_l}^{(l)}$ diferente (ver equação (4.22)). Os valores obtidos para cada regressor original são arranjados em ordem decrescente. A estratégia é reter somente os termos que reduzem significativamente o resíduo de estimação, determinando-se um modelo *ótimo* para a equação (4.14).

Definida a estrutura do novo modelo a partir dos regressores considerados mais relevantes, novos termos $\alpha^{(l)}(i, r_l)$ e $\beta^{(l)}(j, r_l)$ devem ser estimados por meio do método de mínimos quadrados. Finalmente, os parâmetros variantes no tempo $a(i, k)$ e $b(j, k)$ são calculados por meio das equações (4.10) e (4.11).

Durante o procedimento de seleção de estrutura do modelo para o processo da equação (4.9), definiu-se empiricamente quantos ordens ou graus ($r_l = 1, \dots, V_l$) das funções base de Walsh e Legendre ($L = 1, 2$) a serem usados. Em princípio esse procedimento não acarreta prejuízo na determinação da estrutura (Zou et al., 2003; Zou & Chon, 2004).

Para se determinar o número adequado de funções base, uma versão modificada do

critério de informação de Akaike (AIC) foi empregada (Zou & Chon, 2004):

$$\text{AICm} = N \ln \sigma^2 + 2 \left(np \sum_{l=1}^L r_l + 1 \right), \quad (4.25)$$

em que N é quantidade de dados, r_l denota o grau ou ordem de cada tipo de função base l , L é a quantidade de tipos de função base, np é o número de regressores originais significativos da estrutura do modelo selecionado e σ^2 é a variância do erro de predição.

O algoritmo proposto nesta seção pode ser resumido, como:

1. escolha, *a priori*, um modelo polinomial cuja ordem (n_y, n_u, n_e) seja baseada em algum conhecimento prévio ou restrição computacional. Escolha ainda, empiricamente, os tipos de funções base (L) e o número de termos (r_l) de cada uma delas;
2. construa a matriz M a partir da expansão dos regressores candidatos, conforme a equação (4.15);
3. selecione os regressores linearmente independentes por meio do algoritmo OPS (Seção 3.6.2) e construa uma matriz de regressores reduzida, W ;
4. a partir da matriz de regressores expandidos W , use o estimador de mínimos quadrados para determinar os coeficientes α e β conforme equação (4.23). Calcule os índices de projeção acumulados para cada regressor expandido \mathbf{w}_m . Organize esses índices em ordem decrescente e escolha os termos mais relevantes;
5. selecionada a estrutura do modelo, determine o número adequado de funções base pelo critério modificado de Akaike, conforme a equação (4.25);
6. por meio do método de mínimos quadrados, calcule os novos coeficientes α e β , e, finalmente, calcule os parâmetros variantes no tempo por meio das equações (4.10) e (4.11).

Exemplo Numérico da Metodologia TV-OPS.

Com o objetivo de exemplificar o algoritmo TV-OPS descrito na Seção 4.1.2, será usada uma série temporal produzida pelo seguinte processo ARX variante no tempo (Zou et al., 2003):

$$y(k) = a(3,k)y(k-3) + a(9,k)y(k-9) + b(1,k)u(k-1), \quad (4.26)$$

sendo a entrada $u(k)$ um sinal aleatório de distribuição gaussiana $\mathcal{N}(0,1)$. Os coeficientes da equação (4.26) variam no tempo conforme descrição a seguir:

$$\begin{aligned} a(9,k) &= 0,75 + 0,25 \sin(2\pi f_1 k); & k = 1,2, \dots, 300; & f_1 = \frac{1}{300} \text{Hz}; \\ a(9,k) &= -0,60 + 0,20 \sin(2\pi f_1 k); & k = 301, \dots, 600; & \\ a(3,k) &= 0,25 - 0,25 \sin(2\pi f_2 k); & k = 1,2, \dots, 600; & f_2 = \frac{1}{600} \text{Hz}; \\ b(1,k) &= 1; & k = 1,2, \dots, 600. & \end{aligned} \quad (4.27)$$

Foi escolhido, *a priori*, um modelo ARX(9,12)⁸ propositalmente incorreto, para que a partir desse fosse extraída a estrutura correta do processo da equação (4.26). A matriz de regressores expandidos, M , foi construída usando-se as quatro primeiras funções de Legendre (graus de 0 a 3) e as duas primeiras funções de Walsh (ordens 0 e 1). O número de termos das funções base foi escolhido empiricamente, pois, de acordo com Zou et al. (2003) e Zou & Chon (2004), esses valores não afetariam a escolha da estrutura.

O algoritmo OPS foi então aplicado e os índices de distância de projeção calculados. Ressalta-se que, nesse trabalho, foi usado empiricamente um valor de limiar igual a 10^{-6} , para determinar a independência linear entre os regressores (ver equação (3.33)). Zou et al. (2003) e Zou & Chon (2004) sugeriram um limiar da ordem de 10^{-3} e 10^{-4} respectivamente. Entretanto com esses valores não foi possível selecionar corretamente a estrutura do exemplo da equação 4.26. O limiar possivelmente dependa do nível de ruído.

Pode-se verificar na Figura 4.3 os gráficos com os índices de distância de projeção e um índice de distância de projeção relativo. O índice relativo, definido como:

$$\frac{c_m - c_{m+1}}{c_m} \times 100, \quad (4.28)$$

foi proposto por Zou et al. (2003) a fim de facilitar o discernimento de termos relevantes. A idéia é considerar os primeiros picos. Testes mostraram o índice relativo como sendo de pouca ajuda, embora para esse exemplo ele tenha tido bom resultado com o primeiro pico correspondendo ao terceiro termo mais relevante.

Os termos $u(k-1)$, $y(k-9)$ e $y(k-3)$ têm, nessa ordem, maior relevância após algumas realizações. Verificou-se o mesmo resultado com uma relação sinal-ruído de 18dB.

⁸Modelo composto de regressores de saída com até nove atrasos e regressores de entrada com até 12 atrasos.

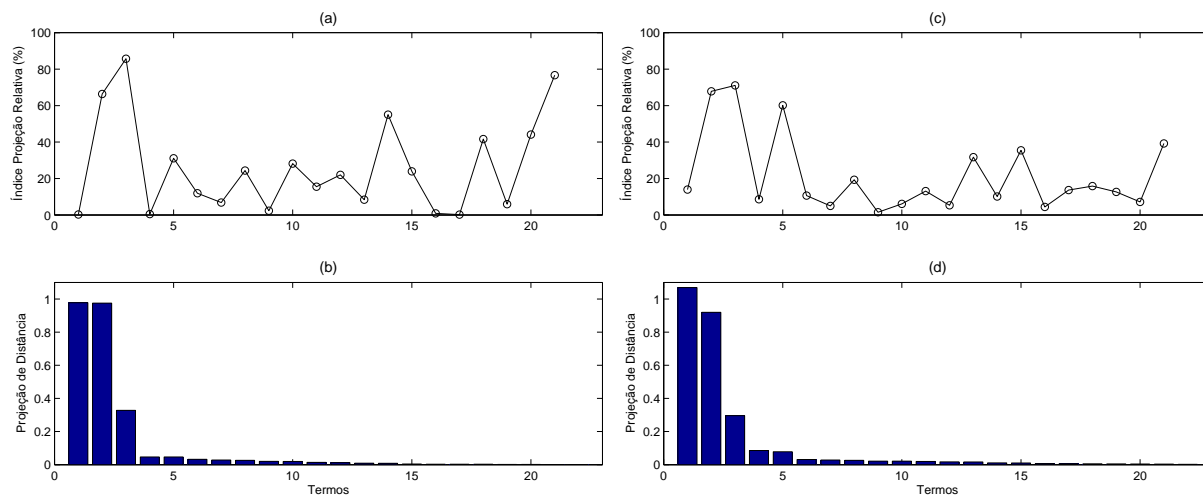


Figura 4.3: Escolha de estrutura para o processo da equação (4.26) a partir do modelo ARX(9,12). Gráficos (a) e (b) mostram o índice de distância de projeção relativo e de distância de projeção respectivamente, obtidos com a série temporal do processo ARX sem ruído. Gráficos (c) e (d) mostram as mesmas grandezas para uma série com ruído aditivo (SNR 18dB). Em ambos os casos, os 3 termos mais relevantes são $u(k-1)$, $y(k-9)$ e $y(k-3)$.

Em tempo, a relação sinal-ruído será doravante definida como:

$$\text{SNR} = 20 \log \left(\frac{\|\mathbf{y}\|}{\|\mathbf{e}\|} \right), \quad (4.29)$$

sendo \mathbf{y} o vetor de observações ou o sinal em que se deseja acrescentar ruído e \mathbf{e} o vetor de ruído. O sinal de ruído possui uma distribuição gaussiana de média zero.

Escolhida então a estrutura do modelo com os três primeiros termos mais relevantes ($np = 3$), calculou-se os valores de AICm, conforme equação (4.25). Foram feitos cálculos em varredura com até 20 funções de Legendre e 20 funções de Walsh. A Figura 4.4a mostra o gráfico com os valores de AICm dependendo da combinação de funções Walsh e Legendre, enquanto a Figura 4.4b mostra o mapa de contorno⁹ referente ao gráfico do critério de Akaike. Baseado nos dados apresentados pelo critério de Akaike modificado, optou-se em expandir os parâmetros do modelo com 8 funções polinomiais de Legendre e 2 de Walsh porque acima desses valores a queda em AICm não foi significativa. É importante ressaltar que esse número de funções base representa o máximo considerado, ou seja, é possível que o procedimento OPS possa ter descartado alguma função de grau ou ordem inferior, como sempre acontece com o grau 0 de Legendre ou ordem 0 de Walsh

⁹Mapa de contorno traçado com 10 níveis.

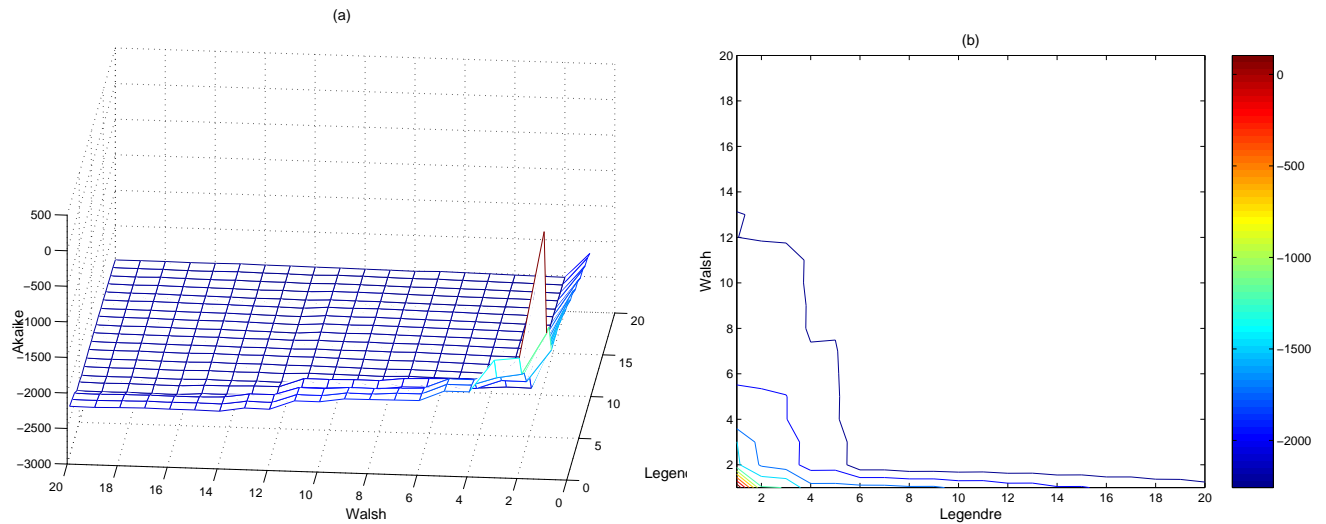


Figura 4.4: Gráficos do critério de Informação de Akaike Modificado, AICm, com relação ao modelo da equação (4.26) (a) AICm. Varredura feita com até 20 funções de Legendre e de Walsh. (b) Mapa de contorno correspondente. A cor azul e suas variações indicam os índices mais baixos.

por serem iguais.

O procedimento TV-OPS é então repetido usando-se a estrutura fixa escolhida, assim como as funções de Legendre e Walsh selecionadas pelo critério de informação de akaike modificado. A Figura 4.5 mostra os parâmetros rastreados pelo algoritmo TV-OPS em uma situação livre de ruído e outra com ruído aditivo (15dB SNR).

Percebe-se que o parâmetro $a(9,k)$ do modelo da equação (4.26) varia suavemente com exceção ao instante $k = 300$, em que o mesmo sofre uma mudança brusca. A fim de se comparar o desempenho do algoritmo no uso de múltiplas funções base ao mesmo tempo, foram estimados os parâmetros usando somente 10 funções de Legendre e outra usando-se somente 10 funções de Walsh. Observa-se na Figura 4.6 que as funções de Walsh modelam bem as variações bruscas, enquanto as funções de Legendre modelam melhor as variações mais suaves. Essas preferências ocorrem também no caso de ruído aditivo. Para sistemas como o da equação (4.26), cuja dinâmica pode variar lenta e bruscamente, o uso de múltiplas funções base, as quais sejam capazes de reproduzir essas variações, é interessante, como pode ser visto na Figura 4.5.

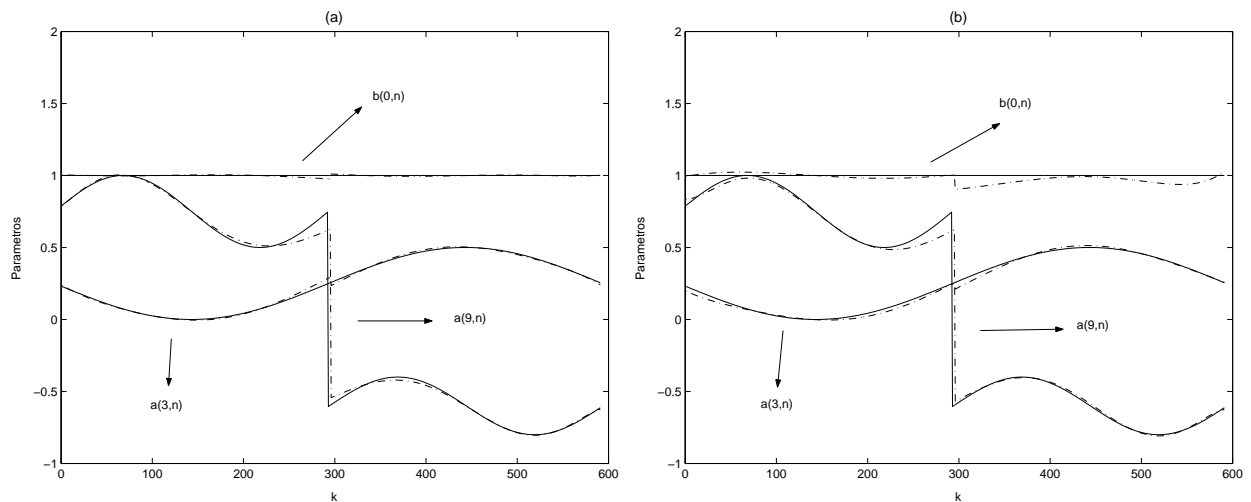


Figura 4.5: Parâmetros TV estimados a partir das observações do sistema da equação (4.26), com 8 funções de Legendre e 2 de Walsh. O gráfico (a) mostra os parâmetros TV estimados (---) em contraste com os parâmetros reais (—), no caso sem ruído. O gráfico (b) mostra as mesmas grandezas para o caso em que foi adicionado ruído na série de observação (15dB SNR).

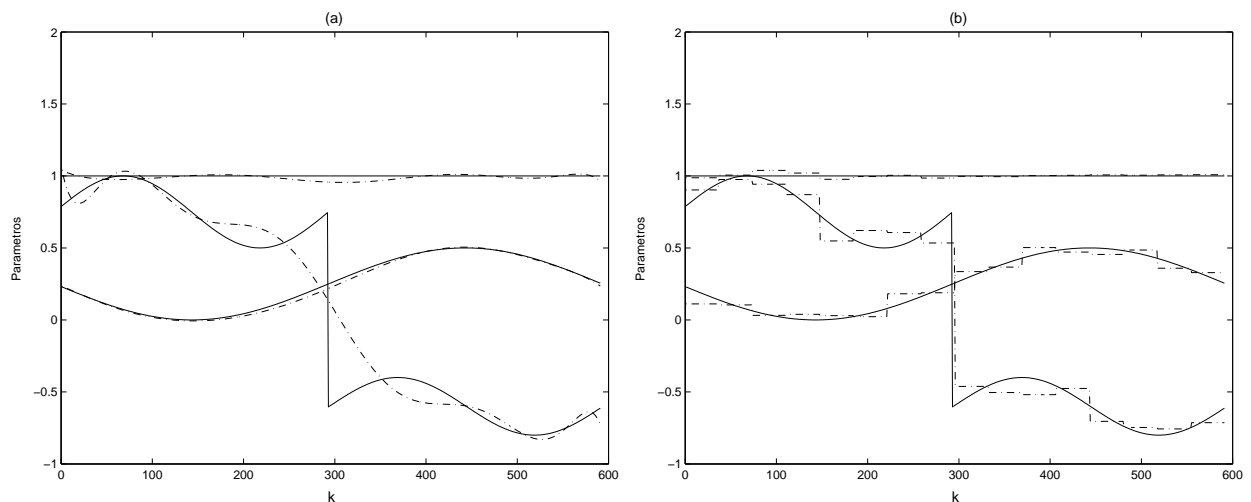


Figura 4.6: Parâmetros TV estimados a partir das observações do sistema da equação (4.26). O gráfico (a) mostra os parâmetros estimados (---) em contraste com os parâmetros reais (—), usando-se 10 funções polinomiais de Legendre. O gráfico (b) mostra os parâmetros estimados usando-se 10 funções de Walsh.

4.2 Estimador de Mínimos Quadrados em Sistemas Variantes no Tempo

O método de mínimos quadrados (MQ) teve seus fundamentos básicos propostos por Karl Gauss¹⁰ (1777-1855) em 1795 a fim de prever o movimento de planetas e cometas

¹⁰Os estudos astronômicos que motivaram a invenção dos mínimos quadrados estão descritos em sua obra *Theoria Motus Corporum Coelestium*, publicado em 1809, na qual referencia o uso do método desde

usando dados de medições telescópicas (Sorenson, 1970). Desde então, o algoritmo de mínimos quadrados e seus variantes tem sido largamente usados em diversas áreas da ciência. Considerando-se definida a estrutura do modelo do processo em observação, a Seção 3.7 apresentou a formulação clássica do estimador de mínimos quadrados. Söderström & Stoica (1989) e Aguirre (2004) abordam outros algoritmos baseados no método clássico de mínimos quadrados, tais como, mínimos quadrados estendido e variáveis instrumentais.

Em todos os métodos de mínimos quadrados supracitados, os sistemas observados foram considerados invariantes no tempo. Sabe-se que o estimador de mínimos quadrados possui um bom desempenho quando os parâmetros do modelo são constantes em sua essência (Ljung, 1987; Söderström & Stoica, 1989). Entretanto, muitas vezes os parâmetros a serem estimados variam no tempo e deseja-se identificar essas variações. Nas seções seguintes, dois métodos baseados no algoritmo de mínimos quadrados serão investigados, tendo em vista a estimação de parâmetros de sistemas variantes no tempo, ou seja, sistemas cuja dinâmica não é constante.

4.2.1 Estimador Recursivo de Mínimos Quadrados com Fator de Esquecimento Variável.

Atribui-se também a Karl Gauss a idéia original do método de mínimos quadrados recursivo. Entrementes, somente em 1950 foi redescoberto por Robin L. Plackett, ainda antes do advento da computação eletrônica *on-line*, e quase passou despercebido. Houve uma segunda redescoberta dos algoritmos recursivos em 1960 no contexto da teoria de controle, a partir da qual desencadeou-se grande interesse (Plackett, 1950; Young, 1984).

Nos métodos de identificação recursivos, também chamados *adaptativos* ou *on-line*¹¹, os parâmetros estimados são calculados recursivamente (ou sequencialmente) no tempo. Isto significa que se há uma estimação de um parâmetro $\hat{\theta}(k-1)$ baseada nos dados até o instante $k-1$, então $\hat{\theta}(k)$ é calculado através de uma modificação *simples* de $\hat{\theta}(k-1)$ (Söderström & Stoica, 1989).

Uma das grandes vantagens desses métodos é que os parâmetros de um determinado modelo podem ser estimados à medida que os dados de observação do processo são disponibilizados. Neste caso, a estimação *on-line* dos parâmetros do modelo deve ser feito de tal forma que o processamento das medições de uma amostra possa ser completada

o ano de 1795. Adrien M. Legendre (1752-1833) independentemente também desenvolveu o método de mínimos quadrados, conforme seu trabalho publicado em 1806 (Sorenson, 1970).

¹¹O processamento *on-line* pode também ser denominado *em tempo real* se o processamento ocorrer suficientemente rápido de maneira que o resultado esteja disponível para influenciar o processo.

durante o intervalo de amostragem. Uma outra vantagem normalmente atribuída aos métodos recursivos é que os mesmos são computacionalmente mais eficientes que os métodos em batelada (Aguirre, 2004; Ljung, 1987). Não obstante, será visto na Seção 4.2.2 que a vantagem supracitada tem sido revista devido aos avanços computacionais da atualidade.

Embora os algoritmos RLS sejam adequados para estimação *on-line* e possuam propriedades estatísticas semelhantes aos algoritmos não-recursivos de mínimos quadrados, os mesmos sofrem de alguns problemas, tendo algumas possíveis fontes de erro (Cho et al., 1991; Söderström & Stoica, 1989; Jiang & Zhang, 2004b):

1. erro devido ao ruído, como por exemplo, o ruído de medição;
2. erro do modelo, que pode estar contido no erro de predição;
3. erro de estimação que ocorre devido à quantidade finita de dados de observação. Esse erro possui polarização zero e variância que diminui com o aumento do tamanho da janela de dados. A escolha das condições iniciais também o influencia;
4. erros numéricos devido aos arredondamentos como resultado da matemática computacional finita;
5. *erro devido às não-estacionariedades*¹², que tem uma variância que aumenta com o tamanho da janela de dados. Esse está ligado com o fato do ganho do algoritmo RLS convergir para zero, o que leva à perda da capacidade de rastreamento de parâmetros variantes no tempo.

O erro causado por não-estacionariedades pode ser minimizado descontando-se o passado e baseando-se a estimação em informações mais recentes. Nesse caso, os parâmetros $\hat{\theta}(k)$ não convergirão à medida que k tender ao infinito, mesmo para sistemas invariantes no tempo, pois o algoritmo descarta informações mais antigas a fim de responder à uma mudança no processo (Söderström & Stoica, 1989). Não obstante, essa estratégia pode ser facilmente incorporada em algoritmos do tipo RLS, o que, aliado à sua eficiência computacional, torna os algoritmos recursivos a categoria mais comum dentre as técnicas que lidam com a estimação de parâmetros de sistemas variantes no tempo.

Pode-se classificar os algoritmos de mínimos quadrados recursivos para sistemas dinâmicos variantes no tempo em três grupos principais (Jiang & Zhang, 2004a): *i*) fator de esquecimento variável (Cho et al., 1991; Cao & Schartz, 1999); *ii*) modificação da matriz

¹²Também conhecido como *lag effect*.

de covariância (Fortescue et al., 1981; Ljung & Gunnarsson, 1990); *iii*) janelas deslizantes (Belge & Miller, 2000).

O objetivo desta seção, é descrever uma das variantes de algoritmos recursivos idealizada para ambientes não-estacionários, chamada *método recursivo de mínimos quadrados com fator de esquecimento variável* (RLSVFF)¹³ (Cho et al., 1991; Coelho & Coelho, 2004).

O estimador recursivo de mínimos quadrados com fator de esquecimento variável tem como ponto de partida o seguinte modelo dinâmico (pode-se ver o desenvolvimento matemático detalhado desse estimador em Aguirre (2004)):

$$y(k) = \boldsymbol{\psi}^T(k-1)\boldsymbol{\theta} + \xi(k), \quad (4.30)$$

sendo que k indica o instante considerado e $\boldsymbol{\psi}$ é um vetor de $n_\theta = \dim[\boldsymbol{\theta}]$ variáveis regressoras tomadas até o instante $k-1$.

O objetivo é estimar $\boldsymbol{\theta}$, via mínimos quadrados, de forma recursiva e de maneira que pondere diferenciadamente os valores observados. O estimador de mínimos quadrados ponderado (MQP) pode ser descrito por meio da seguinte equação:

$$\hat{\boldsymbol{\theta}}_{(k)} = \left[\sum_{i=1}^k w_i(k) \boldsymbol{\psi}(i-1) \boldsymbol{\psi}^T(i-1) \right]^{-1} \sum_{i=1}^k w_i(k) \boldsymbol{\psi}(i-1) y(i), \quad (4.31)$$

sendo que $w_i(k)$ é o valor do i -ésimo peso na k -ésima iteração ($k \leq i$). A seqüência de pesos deverá satisfazer as seguintes restrições:

$$\begin{aligned} w_k(k) &= 1 \\ w_i(k) &= \lambda_k w_i(k-1), \quad i < k, \end{aligned} \quad (4.32)$$

ou seja, o maior peso sempre corresponde ao último valor recebido e é igual a um. Os pesos são multiplicados por um fator λ_k , que pode variar em cada iteração k , sempre que um novo dado é recebido. O fator λ_k é conhecido como *fator de esquecimento*.

Partindo-se da representação do estimador MQP na equação 4.31, pode-se desenvolver o estimador recursivo de mínimos quadrados com fator de esquecimento, cuja formulação é mostrada abaixo:

¹³Do inglês: Recursive Least-Squares with Variable Forgetting Factor algorithm.

$$K_k = \frac{P_{k-1}\boldsymbol{\psi}_k}{\boldsymbol{\psi}_k^T P_{k-1} \boldsymbol{\psi}_k + \lambda_k}, \quad (4.33)$$

$$P_k = \frac{1}{\lambda_k} \left(P_{k-1} - \frac{P_{k-1}\boldsymbol{\psi}_k\boldsymbol{\psi}_k^T P_{k-1}}{\boldsymbol{\psi}_k^T P_{k-1} \boldsymbol{\psi}_k + \lambda_k} \right), \quad (4.34)$$

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_{k-1} + K_k \left[y(k) - \boldsymbol{\psi}_k^T \hat{\boldsymbol{\theta}}_{k-1} \right], \quad (4.35)$$

em que $\boldsymbol{\psi}_k = \boldsymbol{\psi}(k-1)$ é o vetor de regressores que, atualizado na iteração k , contém informação até o instante $k-1$ apenas. $\hat{\boldsymbol{\theta}}$ é o vetor de parâmetros estimados, K^{14} é o ganho de adaptação do vetor estimado (Aguirre, 2004), P é a matriz de covariância, λ_k é o fator de esquecimento no instante k .

Discute-se em seguida algumas possibilidades de determinação do fator de esquecimento, λ . Para o caso estacionário pode-se estimar os parâmetros com $\lambda = 1$, o que equivale exatamente ao RLS clássico em que não há esquecimento. Nesse caso, todos os dados são ponderados igualmente e o algoritmo RLS tem um tamanho de memória *infinita*, o que contribui na redução do efeito do ruído de estimação.

Se em cada iteração k , for usado um mesmo valor do fator de esquecimento $\lambda \in (0,1)$ no caso não-estacionário, a estimação não será estatisticamente ótima (Cho et al., 1991). Uma outra estratégia muito comum é o emprego do fator de esquecimento *exponencial*:

$$\lambda_k = \lambda^{k-i}, \quad 0 < \lambda < 1,$$

em que k é o instante atual e i representa instantes anteriores. Nesse método, assume-se que todos os dados antigos são obsoletos e, portanto, os mesmos são descartados sem nenhuma restrição.

As estratégias supracitadas podem levar a um crescimento exponencial da matriz de covariância, causando *estouro* do estimador¹⁵ (ou estouro da matriz de covariância), ou seja, o estimador se torna instável quando a entrada não é persistentemente excitante. Durante uma excitação deficiente, informações antigas são permanentemente esquecidas enquanto a entrada fornece muito pouca nova informação dinâmica (Fortescue et al., 1981; Cao & Scharz, 1999).

A fim de evitar o problema de estouro da matriz de covariância e restaurar a capa-

¹⁴Muitas vezes referido como o *ganho de Kalman*.

¹⁵Problema conhecido como *estimator wind-up* ou *covariance wind-up*.

cidade de rastreamento da dinâmica do sistema em estudo, pesquisadores tem sugerido diferentes soluções. A idéia básica dessas soluções é a limitação da matriz de covariância introduzindo, por exemplo, um limite superior. Citam-se as seguintes propostas de solução: *i) reset* da matriz de covariância durante excitações não-persistentes (Salgado, Goodwin & Richard, 1988); *ii) esquecimento direcional* ou *esquecimento restrito*, reduzindo a possibilidade de ocorrência de estouro da matriz de covariância quando a informação de entrada não é uniformemente distribuída sobre todos os parâmetros (Ljung & Gunnarsson, 1990; Hagglund, 1995; Cao & Scharz, 1999); *iii) esquecimento seletivo*, cuja idéia básica se assemelha ao esquecimento direcional (Parkum, 1992); *iv) esquecimento múltiplo* ou fatores de esquecimento diferentes relacionados a parâmetros diferentes. O estouro do estimador poderia ocorrer também ao estimar-se múltiplos parâmetros em que todos, ou alguns desses, variem à taxas diferentes (Vahidi, Stefanopoulou & Peng, 2004); *v) fator de esquecimento variante no tempo* (Fortescue et al., 1981; Ljung & Gunnarsson, 1990; Cho et al., 1991; Zhuang, 1998).

O conceito de *fator de esquecimento variável* (VFF) foi introduzido por Fortescue et al. (1981). Conquanto existam outras soluções, tais como as supracitadas, optou-se pelo VFF devido à sua simplicidade. A estratégia de variação do fator de esquecimento, λ , adotada neste trabalho foi proposta por Cho et al. (1991), sobre a qual se discute em seguida.

Dado um pequeno valor do fator de esquecimento, $0 < \lambda \ll 1$, pode-se estimar a tendência global de um sinal não-estacionário ao custo de uma maior variância devido à pequena quantidade disponível de dados. Isso equivale a uma redução do erro devido à não-estacionariedade (*lag effect*) e um aumento do erro de estimação. Por outro lado, dado um valor maior do fator de esquecimento, $0 \ll \lambda < 1$, o tempo de convergência para o parâmetro correto pode ser longo mas eventualmente os parâmetros são estimados com certa precisão quando o sinal experimenta estacionariedade (Söderström & Stoica, 1989; Cho et al., 1991).

O fator de esquecimento é, portanto, escolhido de acordo com o chamado *erro de predição estendido* que é a soma ponderada dos quadrados dos erros de predição *a posteriori*. Em outras palavras, a quantidade de esquecimento corresponderá em cada iteração k , à quantidade de nova informação na iteração anterior $k - 1$, garantindo, portanto, que a estimação seja sempre baseada numa mesma quantidade de informação (Cho et al., 1991; Zhuang, 1998).

A velocidade de adaptação do algoritmo é determinada pelo tamanho da memória

de observações:

$$N = \frac{1}{(1 - \lambda)}, \quad (4.36)$$

o que implica que as informações são desconsideradas paulatinamente com N observações. Num ambiente não-estacionário, tem-se um tamanho mínimo, N_{min} , e máximo, N_{max} , para o tamanho de memória.

A janela de observações é determinada com base no *erro de predição estendido*, conforme definição abaixo:

$$Q_k = \frac{1}{M} \sum_{i=0}^{M-1} \xi_{k-i}^2, \quad (4.37)$$

em que k é o instante de iteração, ξ é o erro de predição. As últimas M iterações são consideradas para mediar proporcionalmente o erro de predição, minimizando os efeitos de erros espúrios devido ao ruído aditivo. Entretanto, M deve ser um número pequeno comparado com o tamanho mínimo de memória, N_{max} , de maneira que a média não mascare a não-estacionariedade do sinal (Cho et al., 1991). Ressalta-se que o erro de predição, que é a diferença entre o sinal e a saída do modelo, pode ter periodicidades desde que o sinal em estudo a tenha. M pode também ser usado para anular a periodicidade do erro.

Ressalta-se em seguida, a idéia básica da equação (4.37). O erro de predição estendido, Q , provê uma informação *a priori* a cerca do estimador. Quando o valor inicial de λ é estabelecido como unitário e o erro é pequeno, pode-se concluir que o estimador é sensível o suficiente para se ajustar às variações dos parâmetros do sistema e portanto reduzir significativamente o erro de estimação. Por conseguinte, é razoável escolher um fator de esquecimento próximo à unidade. No entanto, quando o erro de predição é grande, a sensibilidade do estimador deveria ser aumentada, escolhendo-se um fator de esquecimento menor (Zhuang, 1998; Cho et al., 1991).

Pode-se entender também o erro de predição estendido como um índice de detecção de mudanças na dinâmica do processo. Considerações semelhantes serão feitas na Seção 4.2.2.

Enfim, a estratégia de variação do fator de esquecimento, λ , é dada por:

$$\lambda_k = 1 - \frac{1}{N_k}, \quad (4.38)$$

em que:

$$N_k = \frac{\sigma^2 N_{\max}}{Q_k}, \quad (4.39)$$

sendo k o instante de iteração, Q é o *erro de predição estendido*, σ^2 é a variância do ruído esperado baseado em algum conhecimento real do processo (Cho et al., 1991). O tamanho máximo da memória de observações, N_{\max} , controlará a velocidade de adaptação. Para um processo estacionário, o erro de predição estendido será próximo à variância do ruído, o que resultará em $N_k \approx N_{\max}$. Considerando que esse esquema de variação não garante que λ_k não se torne negativo, um limite inferior, λ_{\min} (relacionado com N_{\min}), será estabelecido.

Exemplo Numérico da Metodologia RLSVFF.

Para ilustrar o algoritmo RLSVFF descrito nesta seção, será usada uma série temporal produzida pelo seguinte processo ARX variante no tempo:

$$y(k) = a(1,k)y(k-1) + b(1,k)u(k-1), \quad (4.40)$$

sendo a entrada $u(k)$ um sinal aleatório de distribuição gaussiana, $\mathcal{N}(0,1)$. Os coeficientes do sistema da equação (4.40) variam no tempo conforme descrição a seguir:

$$\begin{aligned} a(1,k) &= 1; & k &= 1, 2, \dots, 199; \\ a(1,k) &= 1,20 - 0,001k, & k &= 200, \dots, 499; \\ a(1,k) &= -0,3 + 0,002k, & k &= 500, \dots, 649; \\ a(1,k) &= 1; & k &= 650, \dots, 999; \\ a(1,k) &= 0,8; & k &= 1000, \dots, 1349; \\ a(1,k) &= 0,8 + 0,11 \operatorname{sen}(2\pi fk); & k &= 1350, \dots, 1649; \quad f = \frac{1}{300} \text{ Hz}; \\ a(1,k) &= 0,8; & k &= 1650, \dots, 1900; \\ b(1,k) &= 0,4; & k &= 1, 2, \dots, 1900. \end{aligned} \quad (4.41)$$

Definiu-se de maneira empírica as variáveis de projeto necessárias para a execução do algoritmo, tais como: $\sigma^2 = 10^{-12}$ e $N_{\max} = 500$, conforme equação (4.39); $\lambda_{\min} = 0,5$, para limitar os resultados da equação (4.38); $M = 4$, conforme equação (4.37).

Após uma simulação Monte Carlo de 100 realizações, pode-se verificar na Figura 4.7 a média dos resultados da estimação dos parâmetros do sistema da equação (4.40). Percebe-se um eficiente rastreamento dos coeficientes. Tanto no gráfico de variação do fator de esquecimento (Figura 4.7c), quanto no gráfico da evolução do erro de predição

estendido (Figura 4.7b), vemos uma boa indicação dos pontos onde ocorreram mudanças na dinâmica.

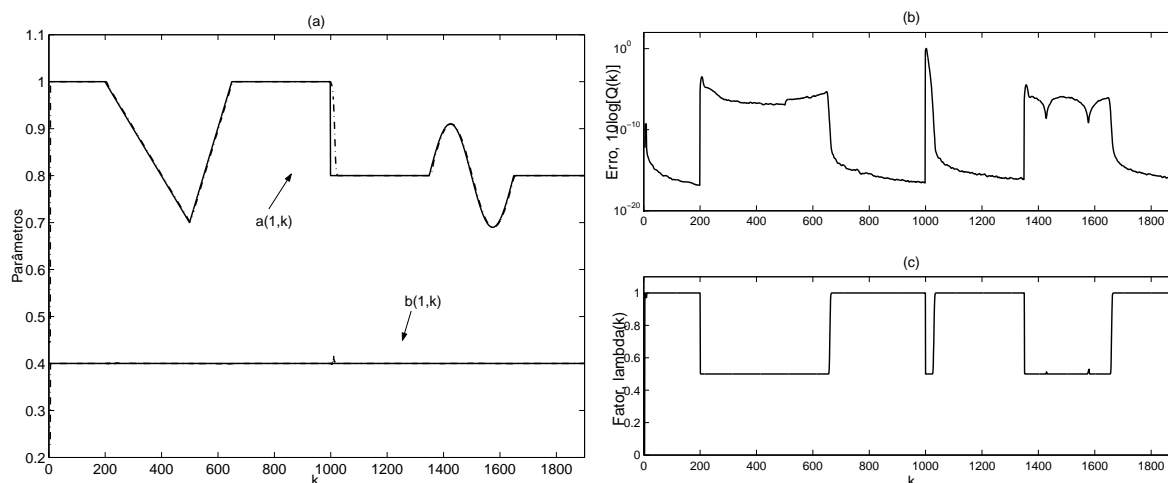


Figura 4.7: Estimação dos parâmetros do sistema representado na equação (4.40), por meio da metodologia RLSVFF. No gráfico (a) temos os parâmetros $a(1,k)$ e $b(1,k)$ originais (—) e estimados (-.-). Em (b) temos o gráfico do logaritmo do erro de previsão estendido, Q_k . Em (c), o gráfico de variação do fator de esquecimento, λ_k .

4.2.2 Estimador em Batelada de Mínimos Quadrados com Janelas Deslizantes de Tamanho Variável.

Foi visto na Seção 3.7 a formulação clássica do estimador de *mínimos quadrados em batelada*. Ao contrário da estimação *recursiva* em que dados são utilizados seqüencialmente na atualização do vetor de parâmetros, na estimação em *batelada* uma quantidade de dados *suficiente*¹⁶ deve estar disponível para que o problema numérico seja resolvido de uma única vez, ou seja, em batelada. Para uma aplicação *on-line*, o processo inteiro de estimação em batelada deverá ser repetido toda vez que um novo dado de observação estiver disponível. Esse formato de implementação será referido doravante como *mínimos quadrados em bloco/batelada* (BLS¹⁷).

Considerando-se que os métodos de estimação recursiva são derivados como aproximações dos métodos em batelada, pode ocorrer que o preço pago por essa aproximação

¹⁶Uma quantidade de dados suficiente no sentido de cobrir as informações da dinâmica do processo em observação.

¹⁷Blockwise Least Squares.

seja a redução da precisão (Söderström & Stoica, 1989). Embora o método BLS produza uma estimação de parâmetros mais precisa que suas variantes recursivas, o mesmo tem em sua carga computacional sua principal desvantagem (Jiang & Zhang, 2004a).

Jiang & Zhang (2004b) consideram que a carga computacional do BLS é da ordem de $\max(n_\theta^3, kn_\theta^2)$, a qual cresce com o número de observações k e o número de parâmetros sendo estimados, n_θ . Essa questão motivou o desenvolvimento de métodos recursivos como o estimador recursivo de mínimos quadrados (RLS), cuja carga computacional é da ordem de apenas n_θ^2 . Entretanto, com os constantes avanços dos microprocessadores¹⁸ dos computadores atuais, a carga computacional do BLS se torna cada vez mais uma questão menor. Não obstante, o método BLS não possui desempenho satisfatório na identificação on-line de sistemas variantes no tempo, cujos parâmetros variem lenta ou abruptamente (Söderström & Stoica, 1989; Aguirre, 2004; Jiang & Zhang, 2004a).

Em vista das vantagens inerentes do BLS, Jiang & Zhang (2004b,a) propuseram uma metodologia para aumentar a habilidade de rastreamento do BLS para sistemas variantes no tempo, enquanto suas características salientes são mantidas. Claramente algum esquema de descarte de informações desatualizadas é necessário. Um maneira efetiva de lidar com essa questão é usar uma janela, de dados, deslizando de tamanho finito. No entanto, se uma janela fixa é usada, pode ser difícil alcançar um bom rastreamento de parâmetros durante o transiente e um alto grau de precisão em condições de estado permanente. A estratégia então é fazer com que o tamanho da janela seja ajustado automaticamente em resposta a uma mudança em potencial dos parâmetros do sistema.

Com esse propósito, foi desenvolvida uma técnica chamada *mínimos quadrados em batelada com janelas deslizantes de tamanho variável* (VLSWBLS¹⁹), cujo objetivo é o rastreamento satisfatório de parâmetros no período transiente e alto nível de precisão da estimação em estado permanente, para sistemas com mudanças abruptas e graduais nos parâmetros. Jiang & Zhang (2004a) ainda sugerem o uso de alguma técnica de ponderação exponencial (ou outra técnica de esquecimento) dentro da janela deslizante para melhorar a capacidade de rastreamento do algoritmo.

Portanto, considerando o modelo ARX da equação (3.41), uma função de custo mo-

¹⁸Em 1974 a Intel Corporation lançou o microprocessador 8080 de 8bits e frequência de *clock* de 2MHz e revolucionou o mercado. Hoje, 31 anos depois, tem-se disponível no mercado microprocessadores de 64bits e *clock* em torno de 3800MHz (Intel, 2005; AMD, 2005).

¹⁹Variable-Length Sliding Window Blockwise Least Squares.

dificada para esses modelos pode ser escrita como:

$$J_{SWBLS} = \sum_{i=k-L+1}^k \lambda^{k-i} \xi(i|i-1, \hat{\boldsymbol{\theta}})^2, \quad (4.42)$$

em que L denota o tamanho da janela deslizante, $\lambda \in (0,1]$ é o fator de esquecimento dentro da janela, ξ é o erro de predição um passo a frente e k denota o instante atual de iteração.

Minimizando a função custo da equação (4.42) em relação ao vetor de parâmetros $\hat{\boldsymbol{\theta}}$, com as L observações mais recentes, obtem-se a seguinte solução para o estimador de mínimos quadrados com janelas deslizantes (Jiang & Zhang, 2004a,b):

$$\hat{\boldsymbol{\theta}}_L(k) = \left[\sum_{i=k-L+1}^k \lambda^{k-i} \boldsymbol{\psi}(i-1) \boldsymbol{\psi}^T(i-1) \right]^{-1} \sum_{i=k-L+1}^k \lambda^{k-i} \boldsymbol{\psi}(i-1) y(i). \quad (4.43)$$

ou, na forma compacta:

$$\hat{\boldsymbol{\theta}}_L(k) = [(\Psi_{k-L+1}^k)^T \Omega_{k-L+1}^k \Psi_{k-L+1}^k]^{-1} [(\Psi_{k-L+1}^k)^T \Omega_{k-L+1}^k \mathbf{Y}], \quad (4.44)$$

em que $\Psi \in \mathbb{R}^{L \times n_\theta}$ é a matriz de regressores definida de maneira similar à equação (3.31), só que para uma janela de L observações. $\Omega \in \mathbb{R}^{L \times L}$ é uma matriz diagonal cujos elementos são função de fatores de esquecimento, tais como, $\lambda^{L-1}, \lambda^{L-2}, \dots, \lambda^0$. Para $\lambda \in (0,1]$, o algoritmo correspondente é chamado *mínimos quadrados em batelada com janelas exponenciais deslizantes* (SEWBLS²⁰). Para $\lambda = 1$, o algoritmo é conhecido como *mínimos quadrados em batelada com janelas retangulares deslizantes* (SRWBLS²¹).

O desempenho do algoritmo acima depende do tamanho da janela L . Para sistemas *invariantes* no tempo, quanto maior a janela de observações maior a precisão da estimação. No entanto, para sistemas em que ocorrem mudanças nos parâmetros, uma janela maior causaria uma resposta mais lenta a essas mudanças. Nesse caso, o tamanho da janela deveria ser ajustada de tal maneira que as observações anteriores à mudança possam ser descartadas efetivamente.

A estratégia de variação do tamanho das janelas deslizantes usada neste trabalho é baseada na proposta apresentada por Jiang & Zhang (2004a,b). Em cada instante de

²⁰Sliding Exponential Window Blockwise Least Squares.

²¹Sliding Rectangular Window Blockwise Least Squares.

iteração k , verifica-se se há mudança nos parâmetros do sistema em estudo e, conforme o resultado, toma-se uma decisão de ajuste no tamanho da janela $L(k+1)$. Observe que agora o tamanho da janela de dados é uma função do instante iteração k .

Define-se a seguir dois mecanismos de detecção de mudanças, os quais são empregados no algoritmo proposto:

Detecção baseada no erro de predição de saída. Considere-se o erro de predição de saída um passo a frente dentro de uma janela de tamanho $L(k)$, conforme solução do estimador SWBLS da equação (4.44):

$$\boldsymbol{\xi}(k) = \mathbf{y}_{k-L(k)+1}^k - \hat{\mathbf{y}}_{k-L(k)+1}^k = \mathbf{y}_{k-L(k)+1}^k - \Psi_{k-L(k)+1}^k \hat{\boldsymbol{\theta}}_L(k), \quad (4.45)$$

em que $\mathbf{y} \in \mathbb{R}^{L(k) \times 1} = [y(k-L(k)+1) \ y(k-L(k)) \ \cdots \ y(k)]^T$ é o vetor de saída (conjunto de observações), $\hat{\mathbf{y}} \in \mathbb{R}^{L(k) \times 1}$ é o vetor de saída estimado, $\Psi \in \mathbb{R}^{L(k) \times n_\theta}$ é a matriz de regressores definida para a janela de observações $L(k)$, $\hat{\boldsymbol{\theta}} \in \mathbb{R}^{n_\theta \times 1}$ é o vetor de parâmetros estimados e $\boldsymbol{\xi} \in \mathbb{R}^{L(k) \times 1}$ é o vetor de erros de predição calculado no instante de iteração k .

A ocorrência de uma mudança de parâmetro será declarada se o seguinte *índice de detecção acumulado*:

$$d_\xi(k) = \frac{1}{M_\xi} \sum_{i=k-M_\xi+1}^k \boldsymbol{\xi}^T(i) \boldsymbol{\xi}(i), \quad (4.46)$$

exceder um limiar pré-estabelecido:

$$d_\xi(k) \begin{cases} > \bar{\delta} & \text{início de mudança,} \\ \leq \underline{\delta} & \text{fim de mudança,} \end{cases} \quad (4.47)$$

em que $\bar{\delta}$ é o limiar para detecção da ocorrência de uma mudança e $\underline{\delta}$ o limiar para detecção do fim de ocorrência de mudança. M_ξ é o número de iterações passadas usado para mediar o índice d_ξ . A escolha dos valores de $\bar{\delta}$ e M_ξ representa um certo grau de compromisso entre a probabilidade de falso alarme e a probabilidade de falha de detecção.

Detecção baseada no erro de estimação de parâmetros. De maneira similar ao esquema anterior, pode-se definir a seguinte média de parâmetros estimados na equa-

çãõ (4.44):

$$\bar{\boldsymbol{\theta}}_{M_\theta}(k) = \frac{1}{M_\theta} \sum_{i=k-M_\theta+1}^k \hat{\boldsymbol{\theta}}_L(i), \quad (4.48)$$

em que $\hat{\boldsymbol{\theta}}_L \in \mathbb{R}^{n_\theta \times 1}$ é o vetor de parâmetros estimados e M_θ é um número que representa o tamanho da janela de iterações usados na mediação.

O seguinte índice pode ser usado na tomada de decisão de mudança ou não de parâmetros:

$$d_\theta(k) = \gamma \times | \|\bar{\boldsymbol{\theta}}_{M_\theta}(k)\| - \|\bar{\boldsymbol{\theta}}_{M_\theta}(k-1)\| | + \\ (1 - \gamma) \times \sum_{i=k-M_\theta+1}^k | \|\bar{\boldsymbol{\theta}}_{M_\theta}(i-1)\| - \|\bar{\boldsymbol{\theta}}_{M_\theta}(i-2)\| |, \quad (4.49)$$

sendo $\gamma \in (0,1)$ um fator de ponderação. Este determina o peso relativo colocado no erro de estimação instantâneo e o acumulado.

O índice $d_\theta(k)$ é observado conforme a regra abaixo, considerando um limiar pré-estabelecido:

$$d_\theta(k) \begin{cases} > \bar{\rho} & \text{início de mudança,} \\ \leq \underline{\rho} & \text{fim de mudança.} \end{cases} \quad (4.50)$$

O algoritmo de detecção de mudanças apresentado nas equações (4.45)-(4.47) pertence a uma classe chamada *algoritmo de soma cumulativa* (CUSUM), que é essencialmente um esquema de média móvel. Basseville & Nikiforov (1993) mostram que d_ξ não é estatisticamente suficiente para mudanças abruptas²² ou graduais nos parâmetros. Teoricamente corre-se o risco de que uma determinada combinação de mudanças de parâmetros possa não ser percebida por meio de alterações em d_ξ . Essa consideração também é pertinente para o critério usado por Cho et al. (1991) na equação (4.37).

Para lidar com a desvantagem supracitada, outros mecanismos de detecção de mudanças mais sofisticados tem sido desenvolvidos (Basseville & Nikiforov, 1993). Jiang & Zhang (2004a) incorporaram o mecanismo de detecção baseado no erro de estimação de parâmetro descrito nas equações (4.48)-(4.50).

²²Definem-se variações abruptas como mudanças em características que ocorrem rapidamente com relação ao período de amostragem das observações, podendo até serem instantâneas (Basseville & Nikiforov, 1993).

Considerando que uma mudança de parâmetro num sistema normalmente se manifesta de maneira mais acentuada no *erro de predição de saída* que no *erro de estimação de parâmetro*, o índice de detecção $d_\xi(k)$ é mais adequado para detecção do início de mudanças e o índice $d_\theta(k)$ é mais sensível para detecção de fim de mudanças (Jiang & Zhang, 2004b).

A essência do algoritmo SWBLS (e derivações) é descrita a seguir. Num determinado instante de iteração k verifica-se, por meio do índice $d_\xi(k)$, se há alguma mudança de parâmetros. Se a mesma não for detectada, uma janela deslizante $L(k) = L_s$ é usada pela equação (4.44). L_s é o tamanho desejável da janela em ambiente estacionário. Caso soe o alarme de mudança de parâmetro, a janela $L(k+1)$ é reduzida drasticamente para um valor mínimo que pode ser definido no seguinte intervalo: $n_\theta \leq L_{\min} \leq L_g$. Considerando o estimador de mínimos quadrados, esse tamanho mínimo da janela não deve ser menor que n_θ (número de termos regressores do modelo). Se essa for uma mudança abrupta, o tamanho da janela expandirá progressivamente nas próximas iterações até atingir o tamanho determinado para ambiente estacionário (L_s). Se for uma mudança gradual, o tamanho da janela se expandirá um pouco até um valor pré-determinado L_g , sendo $L_g \ll L_s$. Nesse ponto, enquanto $d_\theta(k) \leq \underline{\rho}$, o valor de $L(k)$ permanecerá igual a L_g . Quando $d_\theta(k) > \underline{\rho}$, entende-se que a variação gradual terminou (ou que ocorreu mesmo foi uma variação brusca em iterações anteriores) e $L(k)$ continuará a se expandir até o valor igual a L_s .

Resumo do Esquema de Variação da Janela Deslizante:

$L(k) = L_s$, (antes de uma mudança ser detectada);

$L(k) = L_{\min}$, (mudança detectada);

$L(k) = L(k-1) + 1$, (expansão da janela após mudança detectada),

até $L(k) = L_g$, (no caso de mudança de parâmetro gradual),

ou $L(k) = L_s$, (para mudanças abruptas e em estado estacionário).

Exemplo Numérico da Metodologia VLSEWBLS.

Com o propósito de exemplificar a metodologia VLSEWBLS para estimação de parâmetros de sistemas variantes no tempo, usou-se o mesmo processo ARX descrito na equação (4.40) e cujos parâmetros variam também conforme a descrição da equação (4.41). Os resultados representam a média de 100 realizações Monte Carlo.

As variáveis de projeto mais relevantes para a execução do algoritmo, foram definidas empiricamente com os seguintes valores: *i*) limiar de início de mudança, $\bar{\delta} = 10^{-12}$, conforme equação (4.47); *ii*) limiar de fim de mudança, $\underline{\rho} = 0,0005$, conforme equação (4.50).

Vê-se na Figura 4.8, em (a), o gráfico dos parâmetros reais e estimados via RLSFFF. Nos gráficos (b) e (c) da mesma figura, pode-se ver, respectivamente, a variação da janela de mínimos quadrados, $L(k)$, e os índices de erro $d_{\xi}(k)$ (índice de detecção de mudança acumulado) e $d_{\theta}(k)$ (erro de estimação de parâmetros).

Embora sugerido por Jiang & Zhang (2004b), nesse cenário não foi observado se o índice $d_{\theta}(k)$ indicaria maior sensibilidade que $d_{\xi}(k)$, para determinar o fim de mudança nos parâmetros. Verifica-se na Figura 4.8, gráficos c e d, que ambos parecem responder da mesma maneira.

Observou-se também o tempo de processamento para a estimação dos parâmetros do sistema, da equação (4.41), por meio do método VLSEWBLs em comparação com o método RLSVFF. Para uma realização Monte Carlo, o tempo de execução²³ pelo método VLSEWBLs foi de 2,9392s, enquanto pelo método RLSVFF foi 0,1567s. O método recursivo gastou 5% do tempo gasto pelo método em batelada, como era esperado. Ainda assim, o método VLSEWBLs se mostra adequado para estimação on-line de parâmetros.

²³As simulações foram realizadas num microcomputador Intel(R) Pentium(R) 4 CPU 3.00Ghz, com 512MB de memória primária (DRAM) e 1MB de memória cache (SRAM). Programa desenvolvido no software Matlab versão 6.5 e sistema operacional GNU/Linux Mandrake versão 10.2.

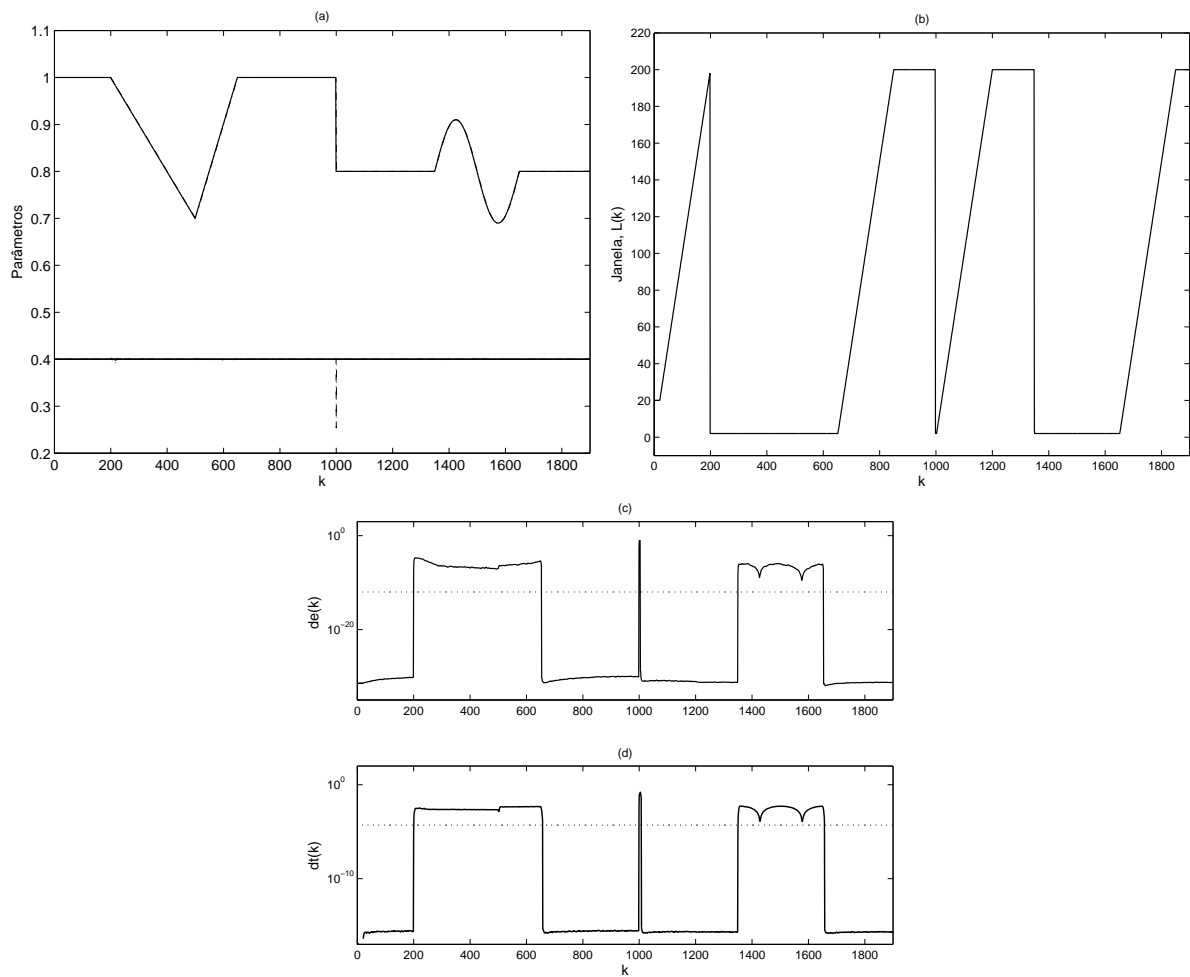


Figura 4.8: Parâmetros do sistema representado na equação (4.40), estimados por meio da metodologia VLSEWBLs. No gráfico (a) temos os parâmetros $a(1,k)$ e $b(1,k)$ originais (—) e estimados (-.-). Em (b) temos o gráfico da variação da janela de mínimos quadrados, $L(k)$. Em (c), no topo, o gráfico de variação do índice de detecção de mudança acumulado, $d_\xi(k)$, em escala logarítmica. (\dots) valor de limiar $\bar{\delta}$. Em (c), embaixo, vemos o gráfico de variação do erro de estimação de parâmetros, $d_\theta(k)$, também em escala logarítmica. (\dots) valor de limiar $\underline{\rho}$. Todos os valores representam a média de 100 realizações Monte Carlo.

Análise da Não-Estacionariedade

Neste capítulo investiga-se a estacionariedade de séries temporais por meio das metodologias abordadas no Capítulo 4. Até que ponto pode-se identificar uma mudança real no padrão dinâmico de um processo? Quão eficiente é o rastreamento de parâmetros variantes no tempo? Os algoritmos conseguiriam distinguir não-estacionariedade de não-linearidade? Uma breve discussão acerca dessas questões será feita.

Todas as metodologias abordadas neste trabalho lidam com modelos polinomiais e assumem que a estrutura desses modelos é constante ao longo do tempo de observação. As variações na dinâmica são então expressas apenas pelas variações nos parâmetros.

Poder-se-ia ainda argumentar que um parâmetro, essencialmente, não é uma variável, e que um modelo com parâmetros constantes e suficientemente compreensível deveria representar adequadamente os regimes dinâmicos. Não obstante, as limitações das técnicas de modelagem, a falta do conhecimento em detalhes a respeito desses regimes dinâmicos e os requisitos do usuário final dificultam a obtenção desse modelo suficientemente compreensível (Norton, 1986). Nessas circunstâncias, a distinção entre parâmetros e variáveis não é clara e, portanto, adotou-se a definição *ad hoc* de que o parâmetro é qualquer coisa que se deseje, por causa de sua interpretação física ou seu lugar na estrutura do modelo.

A fim de mostrar as conseqüências devido à variação temporal da estrutura do modelo ao empregar-se as metodologias consideradas, uma série temporal não-estacionária foi gerada por meio da simulação do seguinte processo:

$$\begin{aligned} y(k) &= 0,8y(k-1) + 0,4u(k-1), & k = 1, \dots, 400; \\ y(k) &= 0,8y(k-1) + 0,4u(k-1) - 0,4y(k-4), & k = 401, \dots, 800; \end{aligned} \quad (5.1)$$

sendo a entrada $u(k)$ um sinal aleatório de distribuição gaussiana, $\mathcal{N}(0; 1)$.

Note que a série temporal gerada pela equação (5.1) é não-estacionária devido a uma mudança na estrutura do processo no tempo de observação $k = 401$, sendo que os parâ-

metros propriamente ditos são constantes. Foram escolhidos alguns modelos polinomiais para explicar o suposto processo gerador dessa série temporal.

O índice $RMSE$, definido a seguir, será usado como um índice de qualidade para comparação dos modelos.

$$RMSE = \frac{\sqrt{\sum_{k=1}^N (y(k) - \hat{y}(k))^2}}{\sqrt{\sum_{k=1}^N (y(k) - \bar{y})^2}}, \quad (5.2)$$

em que $y(k)$ é uma observação do sistema, $\hat{y}(k)$ é a simulação livre da saída do modelo e \bar{y} é o valor médio da série de observações do sistema. Ressalta-se que a simulação livre poderá ser feita ao longo do mesmo tempo de observação em que foram estimados os parâmetros. Isso acontece devido à dificuldade de se fazer previsão real por não conhecermos os parâmetros, que variam no tempo, além do tempo de observação da sistema.

O primeiro modelo, escolhido empiricamente, é composto pelos regressores $\mathbf{y}(k-1)$ e $\mathbf{u}(k-1)$, que são os mesmos do processo da equação (5.1) até o instante de observação $k = 400$. Esse cenário foi proposto considerando-se que a primeira metade dos dados tenha sido usada na identificação do modelo. Pode-se observar na Figura 5.1 o resultado de simulações realizadas. No gráfico (a), vê-se a simulação livre desse primeiro modelo e em (b) a variação temporal de seus parâmetros. Os parâmetros foram estimados por meio da metodologia TV-OPS. Neste cenário, até o instante de observação $k = 400$, a estrutura do modelo corresponde à do processo e obteve-se um índice $RMSE = 0,0092$. Após esse instante o índice $RMSE$ aumentou para 0,8124. Note-se que a variação dos parâmetros não foi suficiente para que esse modelo reproduzisse a variação da dinâmica que, sabe-se, foi causada por mudança na estrutura do sistema original.

O segundo modelo, também escolhido empiricamente, é o mesmo da equação (5.1) após o instante de observação $k = 400$, e portanto é formado pelos regressores $\mathbf{y}(k-1)$, $\mathbf{u}(k-1)$ e $\mathbf{y}(k-4)$. Atente-se para o simples fato de que o regressor $\mathbf{y}(k-4)$ é irrelevante antes do tempo de observação $k = 401$. Observe na Figura 5.1d a variação dos parâmetros estimados para esse modelo via TV-OPS. A simulação livre realizada com os mesmos dados da série temporal, obteve índice $RMSE = 0,0518$ até o instante de observação $k = 400$ e $RMSE = 0,0526$ entre os instantes $k = 401$ e $k = 800$. Esse cenário mostra que se a segunda metade da série temporal tivesse sido usada para determinação do modelo, o mesmo seria adequado para reproduzir os dois regimes dinâmicos do processo da equação (5.1).

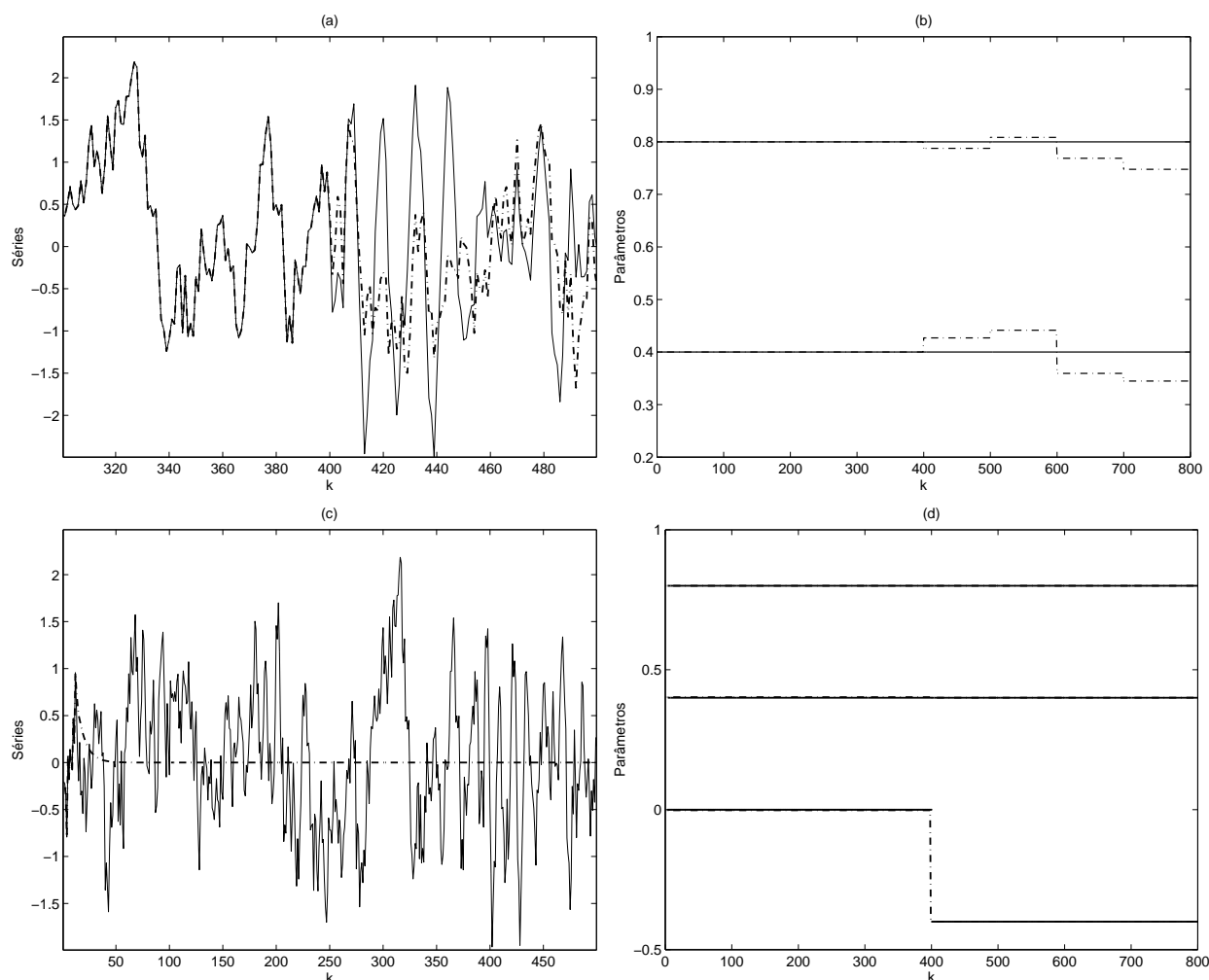


Figura 5.1: Gráficos de simulação livre de modelos ARX cujos parâmetros foram estimados a partir da série temporal simulada do processo da equação (5.1). (a) Simulação livre do modelo ARX cujos regressores são $[\mathbf{y}(k-1) \quad \mathbf{u}(k-1)]$. (b) Parâmetros estimados (-.-) para os termos do modelo proposto em (a), e parâmetros originais (—) dos termos do processo da equação (5.1) até o instante de observação $k=400$. O índice $AICm$ sugeriu o uso de 1 função de Legendre e 6 de Walsh na expansão dos parâmetros. (c) Simulação livre do modelo ARX cujos regressores são $[\mathbf{y}(k-1) \quad \mathbf{y}(k-3) \quad \mathbf{y}(k-4) \quad \mathbf{y}(k-6)]$. O índice $AICm$ sugeriu o uso de 1 função de Legendre e 3 de Walsh na expansão dos parâmetros. (d) Parâmetros estimados (-.-) dos termos do modelo $[\mathbf{y}(k-1) \quad \mathbf{u}(k-1) \quad \mathbf{y}(k-4)]$ e parâmetros originais (—) da equação (5.1).

O terceiro modelo foi escolhido por meio da metodologia TV-OPS usando-se todas as observações. É composto pelos regressores $\mathbf{y}(k-1)$, $\mathbf{y}(k-3)$, $\mathbf{y}(k-4)$ e $\mathbf{y}(k-6)$. Na Figura 5.1c pode-se observar a simulação livre desse modelo. Basta uma inspeção visual

para perceber quão inadequado é esse modelo, a despeito da variação de seus parâmetros.

Vários outros testes foram realizados considerando-se falhas no atraso de regressores, termos regressores inexistentes ou termos em excesso de sistemas com parâmetros variantes no tempo. Ajustes feitos nos parâmetros ao longo do tempo de observação dependem da estrutura correta do modelo do processo. Seriam os parâmetros variantes no tempo limitados pela estrutura escolhida? Parece não ser razoável assumir que, qualquer que seja a estrutura escolhida do modelo, os parâmetros ao variarem no tempo se ajustariam de tal maneira a reproduzirem os padrões dinâmicos do processo mesmo numa estrutura errada ou deficiente.

Em situações práticas, portanto, uma estrutura mais adequada para o modelo de um processo variante no tempo continua a ser extremamente importante. Ressalta-se ainda que os resultados ora apresentados são semelhantes àqueles provenientes dos métodos RLSVFF e VLSEWBLs.

Doravante, em toda discussão envolvendo a questão de não-estacionariedade de séries temporais é assumido que a estrutura do modelo do processo gerador dessa série é constante e que as mudanças no padrão dinâmico do processo podem ser expressas pela variação dos parâmetros do modelo.

5.1 Evidências de Não-Estacionariedade

O objetivo desta seção é investigar a possibilidade de algoritmos, como aqueles abordados no Capítulo 4, diferenciarem não-estacionariedade e estacionariedade em séries temporais. Os algoritmos RLSVFF e VLSEWBLs usam índices de erro como critério de decisão se houve ou não variação na dinâmica. A metodologia TV-OPS usa o critério de Akaike modificado, $AICm$, para determinar a quantidade de funções base para expansão de parâmetros do modelo identificado para a série temporal. Se houver necessidade de funções base para rastrear o parâmetro, poder-se-ia entender que houve variação na dinâmica do processo.

Por meio de alguns exemplos simulados, o critério $AICm$ (ver seção 4.1) será investigado na próximas seções, tendo em vista o propósito supracitado. Em todos os exemplos a serem descritos, foi considerado o seguinte processo ARX:

$$y(k) = a(1,k)y(k-1) + b(1,k)u(k-1), \quad (5.3)$$

sendo a entrada $u(k)$ um sinal PRBS¹ com níveis $+0,5$ e $-0,5$, 10 bits e seqüência de comprimento igual a 3. Os coeficientes $a(1,k)$ e $b(1,k)$ serão descritos oportunamente em cada situação analisada. Em todos os casos foram geradas 600 amostras de observações.

5.1.1 Parâmetros Constantes

Nesta seção descreve-se o cenário em que os parâmetros do processo da equação (5.3) são constantes:

$$\begin{aligned} a(1,k) &= 0,5, \\ b(1,k) &= -0,3. \end{aligned} \quad (5.4)$$

Esse cenário representa uma situação em que não há mudanças na dinâmica do processo. A série temporal gerada a partir do modelo da equação (5.3) com parâmetros constantes possui um padrão dinâmico estacionário. Pode-se ver o resultado do critério de informação de Akaike modificado, $AICm$, na Figura 5.2, sendo seu menor valor referente ao uso de 1 função de Legendre e 1 de Walsh. Como o primeiro grau da função de Legendre e a primeira ordem da função de Walsh são constantes e iguais a um, esse caso indica que não haveria necessidade de expansão dos coeficientes, o que era esperado.

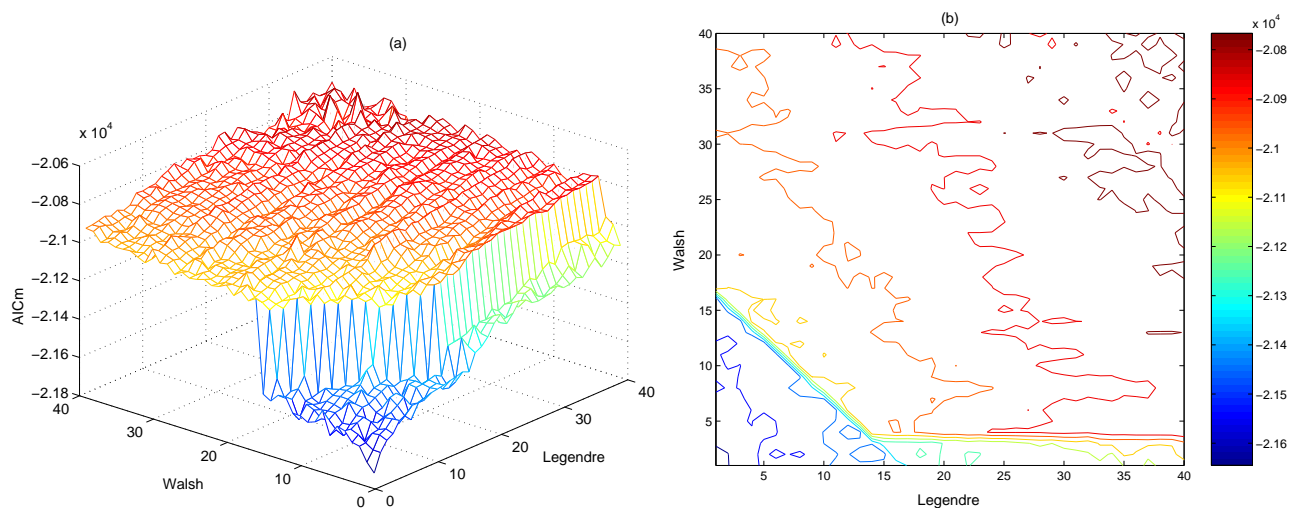


Figura 5.2: Gráficos do critério de informação de Akaike modificado, com relação ao modelo da equação (5.3) com parâmetros constantes. (a) Critério de Informação de Akaike Modificado obtido por meio de uma varredura de até 40 funções de Legendre e de Walsh. (b) Mapa de contorno correspondente. As tonalidades da cor azul indicam os índices mais baixos.

A quantidade mínima de funções base é representada pela ponta do gráfico da Figura

¹Ver descrição detalhada de um sinal PRBS em Aguirre (2004).

5.2a ou a primeira curva de nível (cor aproximadamente azul escuro) da Figura 5.2b. Ressalta-se que a quantidade ótima de funções base sugeridas por $AICm$ são sempre as mesmas, independentemente da realização efetuada, quando os parâmetros a serem estimados são determinísticos.

5.1.2 Parâmetros Aleatórios

Nesta seção descreve-se um cenário em que os parâmetros do processo da equação (5.3) são aleatórios, conforme definição abaixo:

$$\begin{aligned} a(1,k) &= 0,5 + e_a(k), \\ b(1,k) &= -0,3 + e_b(k), \end{aligned} \tag{5.5}$$

em que e_a e e_b são sinais aleatórios de distribuição normal, $N(0; 0,5)$ e $N(0; 1)$ respectivamente.

A Figura 5.3 mostra o resultado da avaliação feita pelo critério de Akaike modificado, a partir de uma realização do processo. O valor de $AICm$ é mínimo para o caso de 6 funções de Legendre e 1 função de Walsh. Essa quantidade de funções base é melhor identificada diretamente nos dados usados na geração dos gráficos da Figura 5.3. Embora os parâmetros sejam aleatórios, o critério $AICm$ apontou um baixo número de funções base. Esse resultado pode ser explicado pelo fato de as rápidas variações dos parâmetros impossibilitarem um rastreamento. Parece que o método entende as variações como ruído e tenta portanto rastrear a média.

Testes realizados com parâmetros $a(1,k) = e_a(k)$ e $b(1,k) = e_b(k)$ (média constante e igual a zero) apresetaram o valor mínimo do critério $AICm$ para uma função de Legendre e uma de Walsh. Pelas mesmas razões sugeridas anteriormente, os parâmetros estimados eram constantes e próximos à média dos parâmetros reais ($\hat{a} \approx \bar{e}_a = -0,0053$ e $\hat{b} \approx \bar{e}_b = -0,007$).

5.1.3 Parâmetros Pseudo-Aleatórios

Descreve-se nesta seção um cenário semelhante ao anterior, porém os parâmetros do processo da equação (5.3) não devem variar tão bruscamente. Definem-se abaixo esses parâmetros:

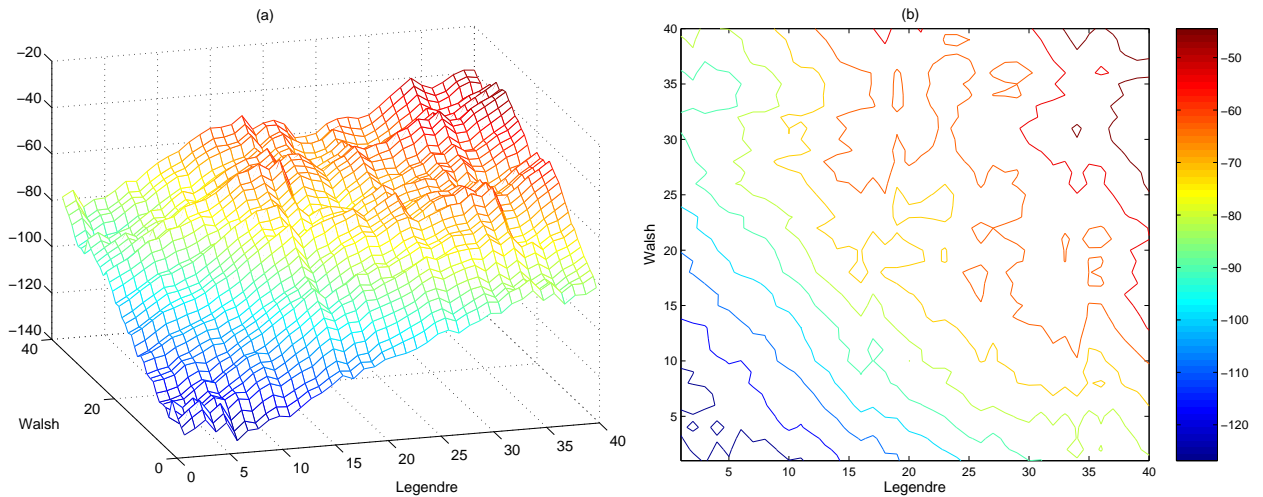


Figura 5.3: Gráficos do critério de informação de Akaike modificado, com relação ao modelo da equação (5.3) com parâmetros aleatórios de distribuição gaussiana, conforme equação (5.5). (a) Critério de Informação de Akaike Modificado obtido por meio de uma varredura de até 40 funções de Legendre e de Walsh. (b) Mapa de contorno correspondente. As tonalidades da cor azul indicam os índices mais baixos.

$$\begin{aligned} a(1,k) &= 0,1 + \text{prbs}(8,2), \\ b(1,k) &= 0,2 + \text{prbs}(10,2), \end{aligned} \quad (5.6)$$

em que $\text{prbs}(i,j)$ é uma função geradora de sinal PRBS, com níveis +1 e 0, i bits e seqüência de comprimento j .

Pode-se perceber na Figura 5.4, que o critério de Akaike modificado, $AICm$, sugere o uso de muitas funções de Legendre e Walsh na tentativa de rastrear os parâmetros aleatórios. Como os parâmetros originais do processo não variam bruscamente, $AICm$ acredita que pode rastreá-los por meio do maior número possível de funções base. Ressalta-se que, embora se aplique as quantidades de funções base sugeridas pelo $AICm$, não foi possível estimar corretamente os parâmetros, como era esperado.

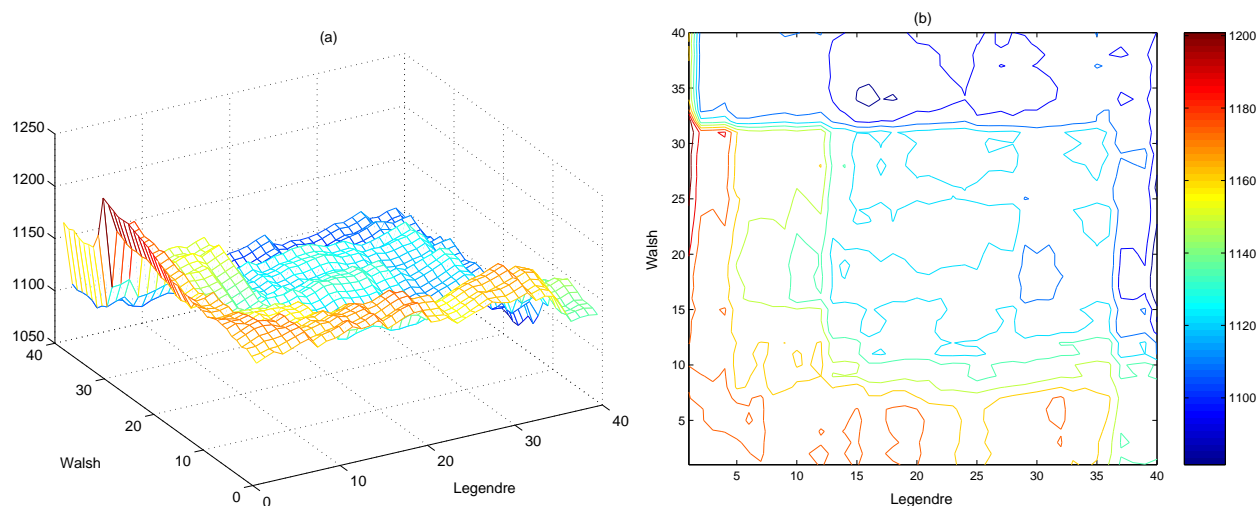


Figura 5.4: Gráficos do critério de informação de Akaike modificado, com relação ao modelo da equação (5.3) com parâmetros aleatórios conforme equação (5.6). (a) Critério de Informação de Akaike Modificado obtido por meio de uma varredura de até 40 funções de Legendre e de Walsh. (b) Mapa de contorno correspondente. As tonalidades da cor azul indicam os índices mais baixos.

5.1.4 Parâmetros com Pequenas Variações Bruscas e Suaves

Esta seção apresenta um cenário em que o parâmetro $a(1,k)$ do modelo da equação (5.3) sofre pequenas variações. Duas situações foram consideradas:

$$\begin{aligned} a(1,k) &= 0,5, & k &= 1, \dots, 300, \\ a(1,k) &= 0,5005, & k &= 301, \dots, 600, \\ b(1,k) &= -0,3, \end{aligned} \quad (5.7)$$

em que $a(1,k)$ sofre uma variação brusca de 0,1% de seu valor, a partir do instante $k = 300$, e

$$\begin{aligned} a(1,k) &= 0,5 + 0,0005 \operatorname{sen}(2\pi f k) \\ b(1,k) &= -0,3, \end{aligned} \quad (5.8)$$

em que o parâmetro $a(1,k)$ sofre uma variação suave conforme meio período de uma senóide com amplitude de 0,1% do valor médio e $f = \frac{1}{1200}$ Hz.

Com os parâmetros da equação (5.7), o critério AIC_m apontou um valor mínimo quando a expansão foi feita com 4 funções de Walsh e 1 de Legendre, conforme pode ser visto na Figura 5.5a. Percebe-se que apesar da variação brusca do parâmetro $a(1,k)$ ser

de apenas 0,1%, o método percebe a variação e ainda sugere somente funções de Walsh por serem mais adequadas nesse caso.

Considerando uma variação suave e pequena para o parâmetro $a(1,k)$, de acordo com a equação (5.8), o critério $AICm$ obteve um valor mínimo para 13 funções de Legendre e 2 de Walsh (ver figura 5.5b). Fica claro a sensibilidade de escolha de funções de Legendre para variações suaves e uma boa sensibilidade para pequenas variações.

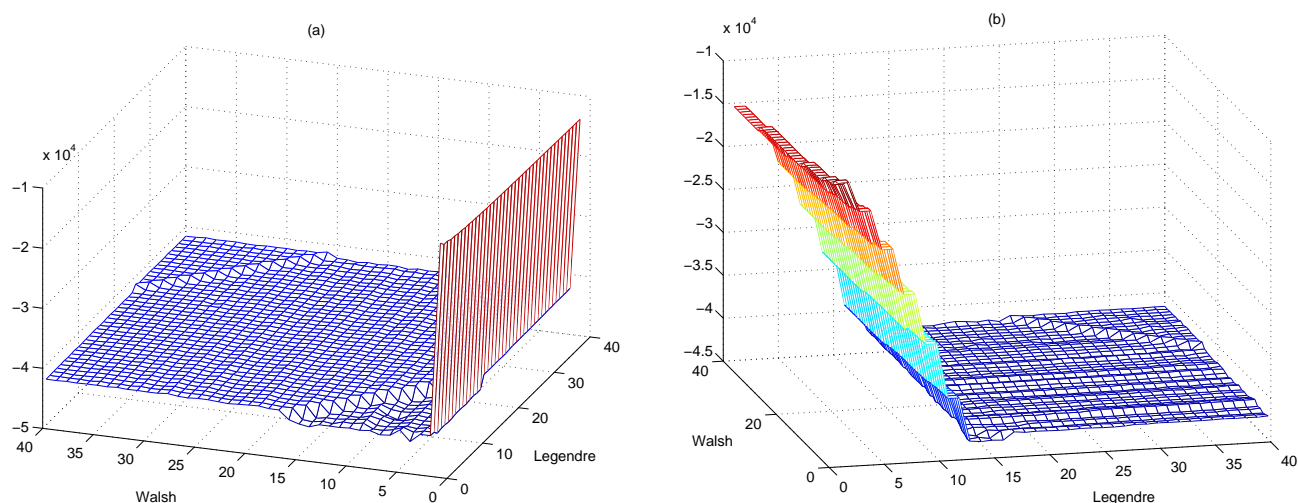


Figura 5.5: Gráficos do critério de informação de Akaike modificado, com relação ao modelo da equação (5.3) com parâmetros com pequena variação. Foi feita uma varredura de até 40 funções de Legendre e de Walsh. (a) Parâmetros variam conforme equação (5.7). (b) Os parâmetros variam conforme equação (5.8).

5.1.5 Discussão

Foi proposto na Seção 5.1 uma investigação sobre a possibilidade de distinção entre não-estacionariedade e estacionariedade em séries temporais, por meio de metodologias como TV-OPS. Essa metodologia foi empregada em alguns exemplos, tendo o critério de Akaike modificado, $AICm$, como um possível meio de discernimento de ocorrência de não-estacionariedades em séries temporais.

Embora o modelo da equação (5.3) tenha uma estrutura simples, vê-se a possibilidade de se considerar como evidência de não-estacionariedade, as múltiplas funções base sugeridas pelo critério de Akaike modificado, conforme os resultados apresentados nas seções 5.1.1, 5.1.2, 5.1.3 e 5.1.4. Ressalta-se novamente que foi assumido que uma mudança no padrão dinâmico seria expressa somente por alterações nos parâmetros e não na estrutura

do modelo.

No caso em que os parâmetros são constantes (Seção 5.1.1) todas as realizações apontaram para a dispensa do uso de funções base. Na situação oposta à anterior, em que os parâmetros variam aleatoriamente (Seção 5.1.2), *AICm* sugeriu o uso de algumas poucas funções base. Com parâmetros aleatórios de média zero, *AICm* sugeriu também a não utilização de funções base. Em ambos os casos com parâmetros aleatórios, o método tentara rastrear a média dos parâmetros aleatórios. O fato é que um parâmetro que varie tão rapidamente e de maneira aleatória, como ocorreu neste exemplo, poderia mesmo ser interpretado como um parâmetro constante perturbado por ruído de média zero. Vê-se ainda na Seção 5.1.3 que quando os parâmetros variam lentamente ainda que aleatoriamente, *AICm* tenta rastreá-los sem sucesso.

A Seção 5.1.4 apresentou casos em que pequenas variações determinísticas foram suficientes para que o índice *AICm* apontasse para o uso de funções base no rastreamento dos parâmetros. Pôde-se novamente observar (rever Seção 4.1.2) que variações bruscas no parâmetro são melhor rastreadas por funções de Walsh e variações suaves por funções de Legendre.

A suma é que em se tratando de parâmetros determinísticos e que os mesmos não variem instantaneamente a cada instante de observação, é possível encontrar um conjunto finito de funções base capaz de rastrear suas trajetórias, sem considerar vários distúrbios intrínsecos a essa metodologia. Estes serão abordados na próxima seção.

5.2 Dificuldades Intrínsecas às Metodologias no Rastreamento de Parâmetros

Ljung (2001) fez a seguinte consideração: *rastrear as propriedades de um sistema é sempre uma questão de avaliação crítica das observações obtidas do processo em questão: as observações contém informações acerca das mudanças no processo ou estão somente dominadas por distúrbios aleatórios?* Esta afirmação está relacionada com o compromisso necessário entre a habilidade de rastreamento e sensibilidade ao ruído. Uma alta taxa de adaptação está relacionada com uma boa capacidade de rastreamento mas também a alta sensibilidade a ruído. Esse compromisso é afetado também pelas variáveis de projeto usadas pelos algoritmos de estimação, como observaram Ljung & Gunnarsson (1990) e Ljung (2001) no caso de algoritmos recursivos.

Apresentam-se nesta seção alguns exemplos em que são mostradas as dificuldades no

rastreamento de parâmetros de sistemas variantes no tempo, devido ao ruído, variáveis de projeto e interferência entre parâmetros. As dificuldades no rastreamento de parâmetros estão relacionadas às dificuldades de identificação dos pontos de mudança na dinâmica do processo.

O sistema a ser usado nos estudos de caso desta seção será o mesmo sistema da equação (4.40). Para facilitar a leitura, rerepresenta-se logo abaixo a equação desse processo.

$$y(k) = a(1,k)y(k-1) + b(1,k)u(k-1), \quad (5.9)$$

em que $u(k)$ é um sinal aleatório de distribuição gaussiana, $\mathcal{N}(0,1)$. Os coeficientes do sistema da equação (5.9) variam no tempo conforme descrição a seguir:

$$\begin{aligned} a(1,k) &= 1; & k &= 1,2,\dots,199; \\ a(1,k) &= 1,20 - 0,001k, & k &= 200,\dots,499; \\ a(1,k) &= -0,3 + 0,002k, & k &= 500,\dots,649; \\ a(1,k) &= 1; & k &= 650,\dots,999; \\ a(1,k) &= 0,8; & k &= 1000,\dots,1349; \\ a(1,k) &= 0,8 + 0,11 \operatorname{sen}(2\pi fk); & k &= 1350,\dots,1649; \quad f = \frac{1}{300} \text{Hz}; \\ a(1,k) &= 0,8; & k &= 1650,\dots,1900; \\ b(1,k) &= 0,4; & k &= 1,2,\dots,1900. \end{aligned} \quad (5.10)$$

5.2.1 Caso I: Ruído na Equação

Nesta seção deseja-se mostrar o efeito nocivo, e esperado, da presença de ruído no processo ao estimar-se parâmetros variantes no tempo. Embora a metodologia usada seja RLSVFF, as mesmas observações se aplicam também às demais metodologias abordadas no capítulo 4, ainda que o método TV-OPS tenha se mostrado um pouco mais robusto.

Para tal cenário, adicionou-se à equação (5.9) um sinal de distribuição aleatória, tal que:

$$y(k) = a(1,k)y(k-1) + b(1,k)u(k-1) + e(k), \quad (5.11)$$

em que e é um sinal aleatório de distribuição gaussiana, $\mathcal{N}(0;0,05)$. A relação sinal/ruído, SNR, foi estimada em 18dB aproximadamente, conforme a equação abaixo:

$$\text{SNR} = 20 \log \left(\frac{\|\mathbf{y}\|}{\|\mathbf{e}\|} \right), \quad (5.12)$$

em que \mathbf{y} o vetor de observações e \mathbf{e} o vetor de ruído.

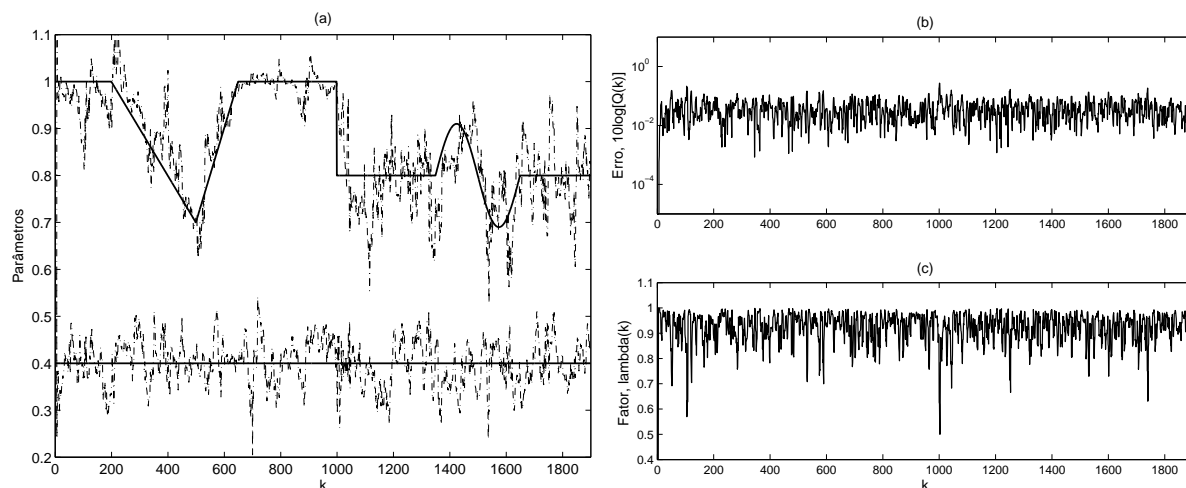


Figura 5.6: Parâmetros estimados para o modelo ARX definido nas equações (5.11) e (5.10), por meio do método RLSVFF. O sistema foi contaminado adicionando-se ruído na equação ($SNR = 18$ dB). (a) Parâmetros estimados (---); Parâmetros originais (—). (b) Erro de predição estendido, $Q(k)$. (c) Fator de esquecimento variável, $\lambda(k)$.

Pode-se observar na Figura 5.6a os parâmetros estimados via RLSVFF em comparação aos parâmetros reais, os quais foram definidos na equação (5.10). A capacidade de rastreamento dos parâmetros foi bastante reduzida. Além desse fato, vê-se também nas Figuras 5.6b e 5.6c, respectivamente, o erro de predição estendido, $Q(k)$, e o fator de esquecimento variável, $\lambda(k)$, que, como consequência, ambos os mecanismos de detecção de pontos de mudanças no padrão dinâmico se tornam quase estéreis. Embora $\lambda(k)$ sugira uma mudança mais brusca no instante de observação $k = 1000$, poder-se-ia somente afirmar que, baseado nos gráficos de $Q(k)$ e $\lambda(k)$, os parâmetros do sistema em observação estariam variando regularmente ou que foi perdida a informação dos instantes de mudanças mais significativos.

Outros testes mostraram que, de fato, variações bruscas são mais prováveis de serem detectadas. À medida que que a taxa SNR aumenta, pontos de mudanças mais bruscas são os primeiros a se diferenciarem.

5.2.2 Caso II: Escolha das Variáveis de Projeto

Pode-se perceber no Capítulo 4 que as metodologias de estimação de parâmetros variantes no tempo dependem da definição de valores para algumas variáveis, doravante chamadas *variáveis de projeto*. A determinação desses valores pode se dar de maneira subjetiva quando se tem pouco ou nenhum conhecimento prévio das características do sistema que está em observação. Na prática essa é a realidade quando se lida com modelagem caixa cinza ou caixa preta.

A escolha das variáveis de projeto afeta consideravelmente os resultados da análise de uma série temporal sobre a qual se deseja conhecer como os coeficientes do modelo escolhido variam no tempo e/ou em que instantes de observação o padrão dinâmico tenha porventura variado.

Nesta seção é apresentado um exemplo em que se observa variações nos resultados devido a diferentes escolhas de uma variável de projeto. Considerou-se a metodologia RLSVFF e uma de suas variáveis de projeto. Na equação (4.39), a variável de projeto σ^2 representa a variância esperada do ruído e ajuda a definir o tamanho da janela ou memória de observação, afetando principalmente a velocidade de adaptação. Usou-se o modelo da equação (5.11) para simular diferentes cenários.

O cenário em que não há ruído na equação foi mostrado anteriormente na Figura 4.7, Seção 4.2.1. Nesse caso, escolheu-se o valor de $\sigma^2 = 10^{-12}$, sendo este próximo de zero conforme sugerido por Cho et al. (1991). Apesar desse cenário não estar próximo de situações reais, é interessante observar o bom rastreamento dos parâmetros estimados bem como a identificação dos pontos de mudança no padrão dinâmico por meio do erro de predição estendido, $Q(k)$, na Figura 4.7b e o fator de esquecimento variável, λ_k , na Figura 4.7c.

Um outro cenário foi considerado ao escolher-se o valor de $\sigma^2 = 5 \times 10^{-4}$ que é o mesmo valor da variância do ruído sugerido para a equação (5.11). Neste caso a taxa SNR foi estimada em 42dB. Vê-se na Figura 5.7 os resultados obtidos. O parâmetro $a(k)$, gráfico (a), foi estimado com atrasos na tentativa de adaptá-lo às mudanças. O fator de esquecimento variável, gráfico (c), traz alguma indicação dos pontos de mudança principalmente no instante $k = 1000$, ao passo que o erro de predição estendido, gráfico (b), deixou dúvidas na diferenciação entre uma perturbação, por exemplo, proveniente de procedimentos de medição, e uma real mudança no padrão dinâmico.

Um terceiro cenário é apresentado na Figura 5.8, em que à variável de projeto σ^2

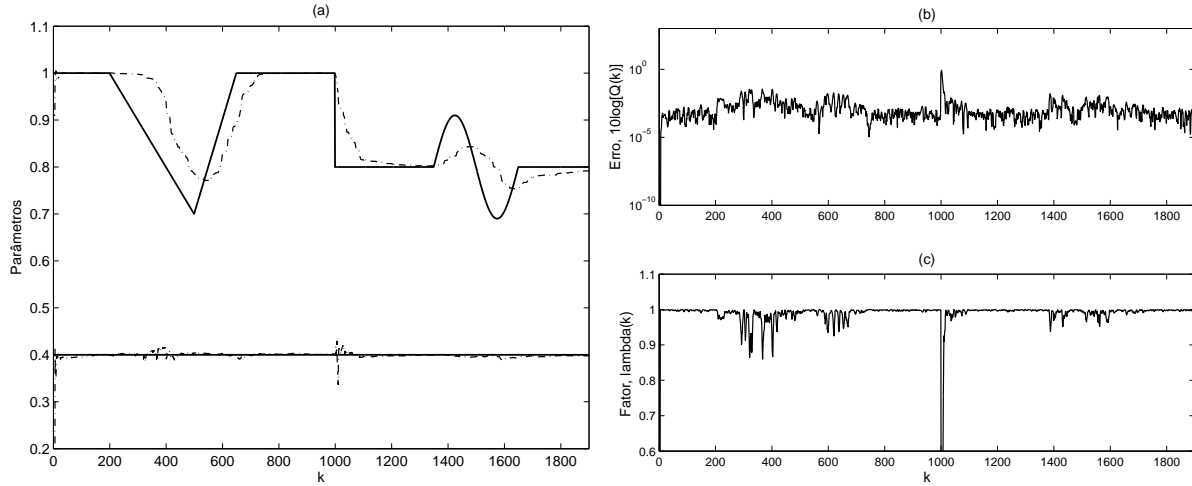


Figura 5.7: Parâmetros estimados para o modelo ARX definido na equação (5.11) por meio do método RLSVFF com taxa SNR estimada em 42dB. Considerou-se a variável de projeto $\sigma^2 = 5 \times 10^{-4}$ (ver equação (4.39)), cujo valor é igual à variância do ruído. (a) Parâmetros estimados (-.-); parâmetros originais (—). (b) Erro de predição estendido, $Q(k)$, e (c) Fator de esquecimento variável, $\lambda(k)$.

foi atribuído um valor menor que a variância do ruído da equação (5.11), ou seja, $\sigma^2 = 1 \times 10^{-7}$. Na Figura 5.8a vê-se que houve ganho na capacidade de rastreamento do parâmetro variante no tempo ao preço de uma maior sensibilidade ao ruído e deficiente capacidade de detecção de pontos de mudança (ver gráficos (b) e (c)).

No último cenário foi considerado $\sigma^2 = 1 \times 10^{-2}$, o qual é maior que a variância do ruído da equação (5.11). Na Figura 5.9a vê-se que, pelo fato do fator de esquecimento ter se mantido próximo da unidade, perdeu-se a capacidade de adaptação às mudanças no padrão dinâmico. Nos gráficos (b) e (c) da Figura 5.9 percebe-se uma melhor possibilidade de detecção de mudanças, ainda que maior ênfase seja dada à detecção da mudança brusca no instante $k = 1000$.

5.2.3 Caso III: Interferência Mútua entre Parâmetros

Conforme comentado por Söderström & Stoica (1989) e Vahidi et al. (2004), ocorre muitas vezes perturbações em um parâmetro estimado, devido a outro parâmetro que varie a uma taxa diferenciada do primeiro. Essa observação foi feita com relação às metodologias recursivas de mínimos quadrados. Coelho & Coelho (2004) também afirmam, referindo-se especificamente à metodologia RLSVFF, que com um fator de esquecimento constante ou

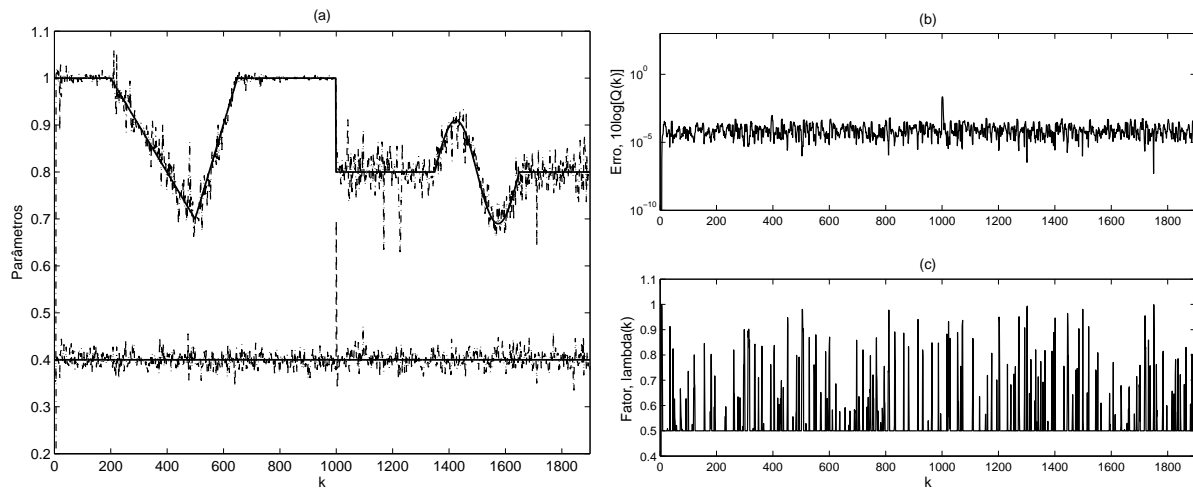


Figura 5.8: Parâmetros estimados para o modelo ARX definido na equação (5.11) por meio do método RLSVFF com taxa SNR estimada em 41dB. Considerou-se a variável de projeto $\sigma^2 = 1 \times 10^{-7}$ (ver equação (4.39)), cujo valor é bem menor que a variância do ruído. (a) Parâmetros estimados (-.-); parâmetros originais (—). (b) Erro de predição estendido, $Q(k)$, e (c) Fator de esquecimento variável, $\lambda(k)$.

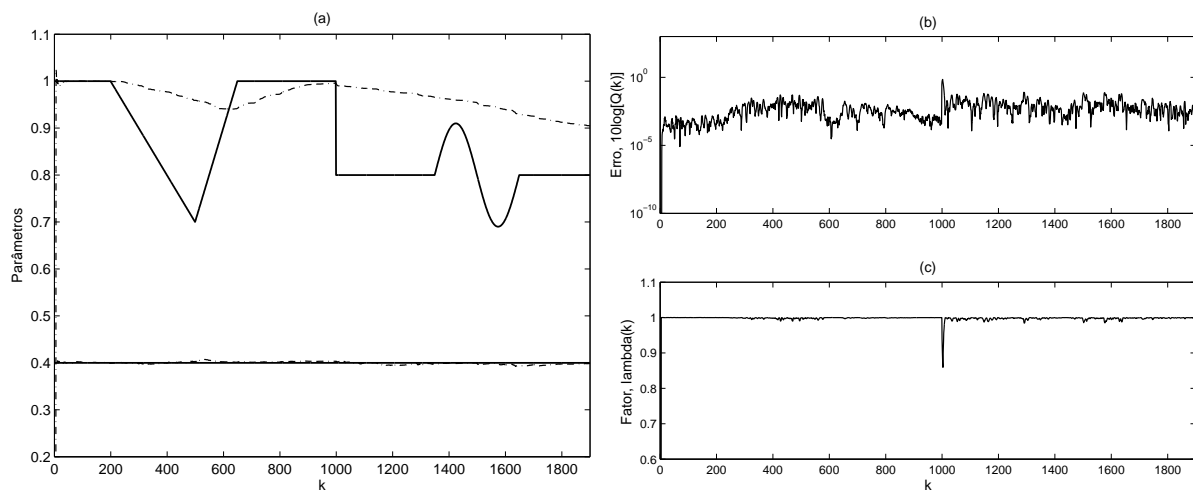


Figura 5.9: Parâmetros estimados para o modelo ARX definido na equação (5.11) por meio do método RLSVFF com taxa SNR estimada em 36dB. Considerou-se a variável de projeto $\sigma^2 = 1 \times 10^{-2}$ (ver equação (4.39)), cujo valor é maior que a variância do ruído. (a) Parâmetros estimados (-.-); parâmetros originais (—). (b) Erro de predição estendido, $Q(k)$, e (c) Fator de esquecimento variável, $\lambda(k)$.

variável, aumentam-se de um mesmo valor todos os elementos da matriz de covariância P (ver equação (4.35)). Por conseqüência, os parâmetros que não são variantes no tempo oscilam.

O modelo da equação (5.9) foi usado para demonstrar o problema de interferência entre parâmetros, entretanto seus parâmetros são definidos conforme equação abaixo:

$$\begin{aligned} a(1,k) &= 1; & k &= 1, 2, \dots, 300; \\ a(1,k) &= 0,8; & k &= 301, \dots, 600; \\ b(1,k) &= 0,5; & k &= 1, 2, \dots, 600. \end{aligned} \tag{5.13}$$

A Figura 5.10 mostra os parâmetros estimados para o modelo das equações (5.9) e (5.13), conforme a metodologia RLSVFF. Verificou-se que o parâmetro estimado, $\hat{b}(1,k)$, originalmente constante, sofre perturbações devido à variação brusca do parâmetro $a(1,k)$. Observou-se também esse tipo de interferência quando foram empregadas as metodologias VLSEWBLs e TV-OPS.

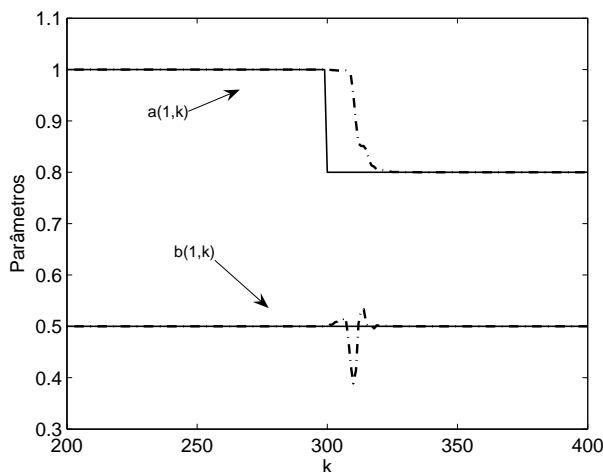


Figura 5.10: Exemplo de interferência entre os parâmetros estimados para o modelo das equações (5.9) e (5.13), usando o método RLSVFF. (—) parâmetro original; (-.-)parâmetro estimado.

5.2.4 Discussão

Nesta seção foram ilustrados vários problemas intrínsecos às metodologias de estimação de parâmetros variantes no tempo. Foi visto, por exemplo, como o ruído e a escolha empírica ou subjetiva de variáveis de projeto podem interferir negativamente no processo de rastreamento dos parâmetros e na determinação de pontos de mudanças no padrão dinâmico do sistema em estudo. Observou-se também que parâmetros que variam a taxas diferenciadas podem causar interferência mútua.

5.3 Não-Estacionariedade × Não-Linearidade

Deseja-se investigar nesta seção, a possibilidade de se diferenciar não-estacionariedade de não-linearidade por meio das metodologias abordadas no Capítulo 4. Pode-se fazer o raciocínio inverso, questionando a possibilidade dessas metodologias distingüirem não-linearidades de não-estacionariedades.

O cenário de testes apresentado em seguida é semelhante ao cenário considerado na introdução deste capítulo mas dentro de uma perspectiva diferente, ou seja, confrontando-se não-linearidade com não-estacionariedade. Um modelo NARX com parâmetros fixos será usado para gerar uma série temporal e em seguida um modelo ARX com parâmetros variáveis é considerado para identificação de seus parâmetros a partir da série temporal mencionada.

Vê-se em seguida o modelo polinomial NARX usado para gerar uma série temporal:

$$y_1(k) = 0,5y_1(k-1)y_1(k-1) - 0,3u(k-1), \quad (5.14)$$

em que a entrada $u(k)$ será definida oportunamente nos casos a serem analisados. Vê-se que o padrão dinâmico não-linear desse processo é invariante no tempo ao definir-se seus parâmetros com valores constantes. Foi gerada uma série temporal, \mathbf{y}_1 , com 600 observações.

O modelo ARX abaixo foi escolhido admitindo-se que o mesmo seja capaz de reproduzir o mesmo padrão dinâmico do processo expresso na equação (5.14):

$$y_2(k) = a(1,k)y_2(k-1) + b(1,k)u(k-1), \quad (5.15)$$

em que os parâmetros $a(1,k)$ e $b(1,k)$ serão estimados por meio da série temporal \mathbf{y}_1

(equação (5.14)), ou seja, a série temporal \mathbf{y}_2 será feita igual à série \mathbf{y}_1 . Sabe-se que os valores originais desses coeficientes quando compara-se essa equação com a equação (5.14) são: $a(1,k) = 0,5y_1(k-1)$ e $b(1,k) = -0,3$.

Segue-se nas próximas seções, a análise de dois casos montados a partir do cenário descrito acima.

5.3.1 Caso I: Sinal de Entrada Pseudo-Aleatório

O sinal de entrada, $u(k)$, dos sistemas das equações (5.14) e (5.15), foi definido como:

$$u(k) = -0,5 + \text{prbs}(600,10,2), \quad (5.16)$$

em que o sinal pseudo-aleatório possui níveis $+0,5$ e $-0,5$, 10 bits e seqüência de comprimento igual a 2.

Aplicando-se a metodologia TV-OPS, o critério $AICm$ sugeriu o uso de 15 funções de Legendre e 1 de Walsh. A Figura 5.11 mostra os parâmetros do modelo linear ARX definido na equação (5.15), $a(1,k)$ e $b(1,k)$.

Pode-se observar que o parâmetro $a(1,k)$ estimado para o modelo da equação (5.15) tende a acompanhar a média dos valores originais, comportamento que já foi observado na Seção 5.1.2. O parâmetro $b(1,k)$ foi estimado satisfatoriamente. Embora os valores rastreados possam ser considerados praticamente constantes em uma eventual aplicação, a quantidade de funções base escolhida indicaria uma possível não-estacionariedade.

Testes realizados com a metodologia RLSVFF mostraram resultados muito semelhantes no que diz respeito aos parâmetros estimados $a(1,k)$ e $b(1,k)$, e os respectivos gráficos não serão apresentados. Ainda com relação à metodologia RLSVFF, a Figura 5.12 mostra os gráficos do erro de predição estendido, $Q(k)$, e do fator de esquecimento variável, $\lambda(k)$. Pode-se verificar que a variável $Q(k)$ mostra uma variação uniforme e não permite se concluir corretamente a respeito da não-estacionariedade. Por outro lado, o fator de esquecimento varia constantemente em torno do valor médio igual a 0,85, o que pode indicar uma variação uniforme dos parâmetros, o que de fato é verdadeiro para o parâmetro $a(1,k)$.

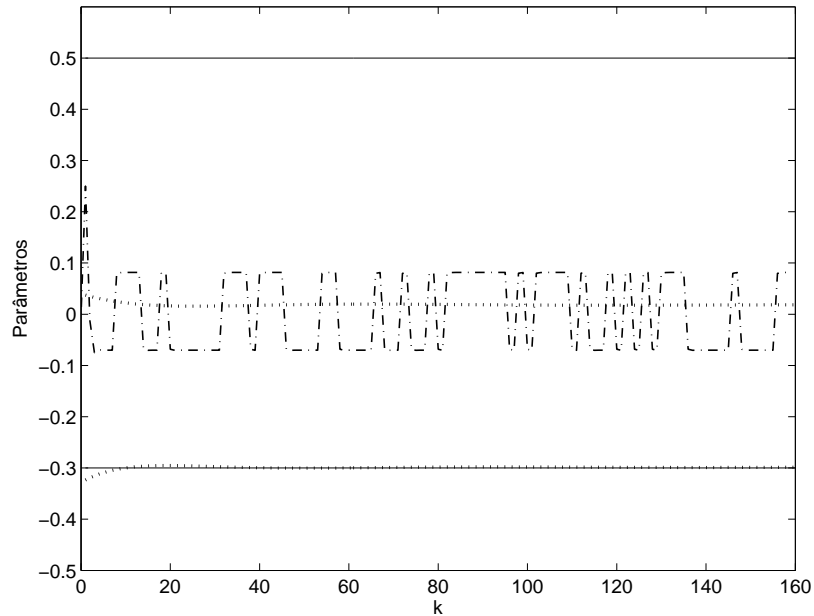


Figura 5.11: Gráfico que mostra os parâmetros originais e estimados, $a(1,k)$ e $b(1,k)$, do modelo ARX definido na equação (5.15), por meio do método TV-OPS. Empregou-se 15 funções de Legendre e 1 de Walsh na expansão dos parâmetros. (—) parâmetros do sistema não-linear NARX da equação (5.14); (\cdots) parâmetros $a(1,k)$ e $b(1,k)$ estimados; (-.-) parâmetro $a(1,k)$ original.

5.3.2 Caso II: Sinal de Entrada Suave e Determinístico

Neste caso foi definido um sinal de entrada mais suave para os sistemas das equações (5.14) e (5.15), tal como:

$$u(k) = 0,4 \text{sen}(2\pi f_1 k) + 0,5 \text{sen}(2\pi f_2 k), \quad (5.17)$$

em que $f_1 = \frac{1}{300}$ Hz e $f_2 = \frac{1}{200}$ Hz.

Para o método TV-OPS, o critério $AICm$ sugeriu 23 funções de Legendre e 16 de Walsh para expansão dos coeficientes da equação (5.15), cuja entrada $u(k)$ é definida na equação (5.17). A Figura 5.13 mostra os resultados do rastreamento dos parâmetros.

Vê-se que, tanto pelas múltiplas funções base sugeridas pelo critério $AICm$ quanto pela variação dos parâmetros estimados, se poderia sugerir alguma variação do regime dinâmico do sistema em estudo.

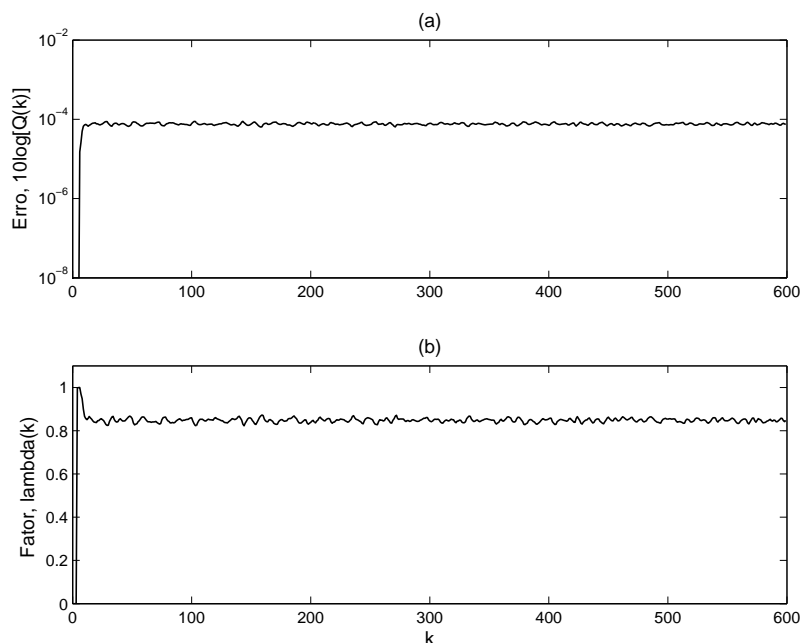


Figura 5.12: Gráficos que mostram as variações do erro de predição estendido, $Q(k)$, e do fator de esquecimento variável, $\lambda(k)$, na estimação dos parâmetros do modelo ARX definido na equação (5.15), por meio do método RLSVFF. Representam a média de 100 realizações Monte Carlo. (a) Erro de predição estendido, $Q(k)$. (b) Fator de esquecimento variável, $\lambda(k)$.

5.3.3 Discussão

Nesta seção questionou-se a possibilidade das metodologias de modelagem de sistemas variantes no tempo, apresentadas no Capítulo 4, distinguírem não-linearidades de não-estacionariedades.

Os resultados dos exemplos apresentados sugeriram que pode-se concluir sobre a não-estacionariedade de um determinado processo em detrimento de não-linearidades não observadas. Nos cenários estudados, as múltiplas funções base sugeridas pelo critério $AICm$ do método TV-OPS e a variação dos próprios parâmetros estimados reforçam essa hipótese.

O problema parece se apresentar como uma questão de erro na estrutura, juntando-se a esta as observações já feitas na introdução deste capítulo. Nesse sentido, se a estrutura do modelo não representar bem as não-linearidades presentes no sistema, *é possível se confundir não-linearidade com não-estacionariedade*.

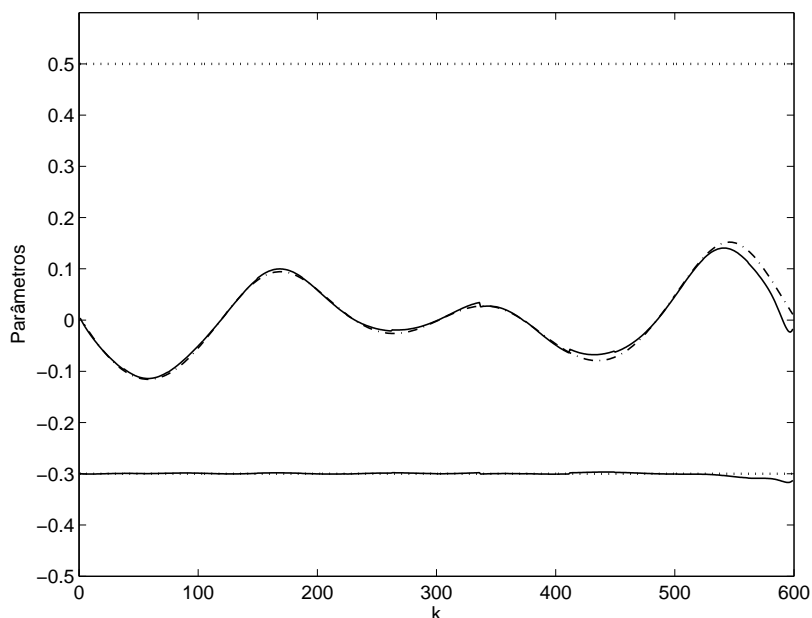


Figura 5.13: Gráfico que mostra os parâmetros originais e estimados, $a(1,k)$ e $b(1,k)$, do modelo ARX definido na equação (5.15) com entrada não aleatória (equação (5.17)), por meio do método TV-OPS. Foram escolhidas, via AICm, 23 funções de Legendre e 16 de Walsh para expansão dos parâmetros. (⋯) parâmetros do modelo NARX da equação (5.14); (—) parâmetros $a(1,k)$ e $b(1,k)$ estimados ; (-.-) parâmetro $a(1,k)$ original.

5.4 Conclusões

Abordou-se neste capítulo o tema da não-estacionariedade de séries temporais à luz das metodologias de análise de sistemas variantes no tempo apresentadas no Capítulo 4. Ressaltou-se inicialmente a importância de os modelos polinomiais possuírem uma estrutura constante ao longo dos instantes de observação. Sob essa consideração, qualquer mudança no regime dinâmico de um processo seria então expressa por variações nos parâmetros do modelo. A não observância desse ponto poderia anular de antemão os resultados da análise feita pelas metodologias em questão. Vê-se na Seção 5.3, por exemplo, que uma falha do modelo polinomial em expressar as não-linearidades do processo pode levar à uma confusão entre não-estacionariedade e não-lineariedade.

Observou-se também que ainda que o modelo usado seja adequado ao processo, a não-estacionariedade expressa nas variações dos parâmetros e também a determinação do instante de ocorrência dessas mudanças podem ser facilmente mal interpretadas devido à

presença de ruído, o ajuste incorreto das variáveis de projeto das metodologias de análise e interferência mútua entre os parâmetros do modelo.

Embora em condições bem adequadas as metodologias estudadas consigam observar a não-estacionariedade em séries temporais, a tarefa de modelagem de um sistema real variante no tempo torna-se bastante complexa. Essa conclusão pode ser sustentada com relação a todas metodologias abordadas no Capítulo 4 deste trabalho.

Ainda que em meio às dificuldades já discutidas, o índices de erro de predição estendido, $Q(k)$, e o fator de esquecimento variável, $\lambda(k)$, usados para determinação de instantes de mudanças no padrão dinâmico se mostraram um pouco mais robustos quando ocorrem mudanças bruscas. No Capítulo 6 deste trabalho pretende-se abordar sistemas reais sob esse prisma.

Detecção de Mudança na Dinâmica do Processo

Considerando-se as dificuldades abordadas no Capítulo 5, deseja-se investigar neste capítulo somente a questão de detecção de pontos de mudança na dinâmica de séries temporais não-estacionárias. A partir de exemplos simulados em comparação com séries temporais reais, serão abordadas não-estacionariedades típicas como mudança brusca na média e tendência.

As séries não-estacionárias simuladas foram obtidas em dois passos. O primeiro passo, que é comum em ambos os casos, foi gerar uma série temporal estacionária a partir de um processo ARX cuja dinâmica seja bem conhecida. O segundo passo diz respeito às modificações necessárias nessa série temporal a fim de reproduzir os tipos comuns de não-estacionariedades supracitas. Esse último passo será detalhado oportunamente. A partir da série temporal não-estacionária busca-se então um modelo que seja adequado para reproduzir a dinâmica do suposto processo que a tenha gerado. Usou-se a metodologia TV-OPS como auxílio na identificação de modelos.

Vê-se abaixo o processo ARX usado no primeiro passo citado:

$$y(k) = 0,5y(k-1) - 0,3u(k-1), \quad (6.1)$$

em que a entrada $u(k)$ é um sinal PRBS com níveis $+0,5$ e $-0,5$, 10 bits e seqüência de comprimento igual a 2.

A validação dos modelos será feita por inspeção visual do gráfico das séries reais e preditas, bem como por meio do índice $RMSE$ definido na equação (5.2).

Tabela 6.1: Modelos ARX usados na modelagem do processo resultante das equações (6.1) e (6.2).

Modelo 1	Modelo 2
$y(k-1)$	$u(k-1)$
$y(k-2)$	$y(k-1)$
$u(k-2)$	1
$u(k-1)$	
1	
$u(k-3)$	
$y(k-4)$	
$u(k-5)$	
$u(k-3)$	
$y(k-5)$	

6.1 Mudança Brusca da Média

Investiga-se nessa seção a modelagem paramétrica de uma série temporal cuja não-estacionariedade é devido à mudança brusca da média. Obteve-se inicialmente uma série temporal da maneira que se segue:

$$\begin{aligned} y_m(k) &= y(k) & k = 1, \dots, 350, \\ y_m(k) &= y(k) + 0,5, & k = 351, \dots, 600, \end{aligned} \quad (6.2)$$

em que $y(k)$ é uma observação gerada pelo processo da equação (6.1) e $y_m(k)$ é a observação modificada com uma mudança de média a partir de $k = 351$.

A partir da série temporal produzida pelas equações (6.1) e (6.2), buscou-se então um modelo, por meio do método TV-OPS, que fosse adequado para reproduzir essa série. A quantidade de 600 observações foi usada na definição da estrutura do modelo. Inicialmente sugeriu-se modelos polinomiais do tipo NARX de grau de não-linearidade $\ell = 2$. Alguns desses modelos foram testados mas se mostraram instáveis. Na segunda tentativa optou-se por modelos ARX.

O modelo 1 da Tabela 6.1 foi escolhido com a ajuda do método TV-OPS. Fazendo-se uma varredura de 50 funções de Legendre e de Walsh, o critério *AICm* sugeriu o uso 40 funções de Legendre e 34 de Walsh. Os parâmetros estimados via TV-OPS para esse modelo atingem valores de tal magnitude que o modelo fica completamente instável. Sugere-se que isso tenha ocorrido pela impossibilidade de combinação das funções base e/ou até pelos erros numéricos gerados pelas manipulações da matriz de regressores $W \in \mathbb{R}^{596 \times 494}$,

conforme pode ser verificado nas equações (4.21) e (4.23).

Os parâmetros estimados para o modelo 1 por meio das metodologias RLSVFF e VLSEWBLs não provocam instabilidade. Pode-se ver na Figura 6.1 os resultados da estimação dos parâmetros via RLSVFF. Os resultados obtidos com o uso do VLSEWBLs não serão mostrados devido à similaridade com o método RLSVFF.

O gráfico (a) da Figura 6.1 mostra os resultados da predição livre, com índice $RMSE = 0,3039$. Embora vários parâmetros, incluindo alguns mostrados no gráfico (c), sofram mudanças em torno do tempo de observação $k = 351$, houve um atraso razoável na convergência após a mudança brusca da média. No gráfico (c) se observa que tanto o índice de predição estendido quanto o fator de esquecimento variável ressaltam o ponto de mudança, como era de se esperar.

De uma maneira empírica, optou-se por fazer a mesma análise com o modelo 2 da Tabela 6.1, baseando-se no conhecimento prévio de que o comportamento dinâmico do processo da equação (6.2) é o mesmo ao longo do tempo de observação exceto pelo ganho estático obtido a partir do instante $k = 351$.

A Figura 6.2 mostra os resultados da predição feita pelo modelo 2 da Tabela 6.1, com os parâmetros estimados via RLSVFF ($RMSE = 0,0899$). Percebe-se na Figura 6.2b, que os parâmetros estimados para os termos $u(k-1)$ e $y(k-1)$ são iguais aos originais da equação (6.1), exceto pela interferência causada pela variação súbita do parâmetro do termo unitário. A responsabilidade pela mudança na média do processo ficou a cargo do termo unitário do modelo 2.

6.1.1 Caso1: Circuito de Chua

Na seção anterior discorreu-se a respeito da possibilidade de detecção de mudança de média de série temporal obtida por meio de simulação. Pretende-se na presente seção, fazer uma análise similar de uma série temporal caótica gerada a partir de um circuito de Chua real.

O circuito de Chua é um circuito eletrônico capaz de exibir oscilações periódicas ou caóticas (Kennedy, 1992). Chua é composto por dois capacitores C_1 e C_2 , um resistor R , um indutor L e um elemento não-linear conhecido como *diodo de Chua*. O diodo de Chua possui resistência negativa e é o elemento que confere energia ao sistema, permitindo assim, manter o circuito oscilando autonomamente (Junior, 2003; Tôres, 2001).

As equações que descrevem o circuito de Chua são (Chua, 1994):

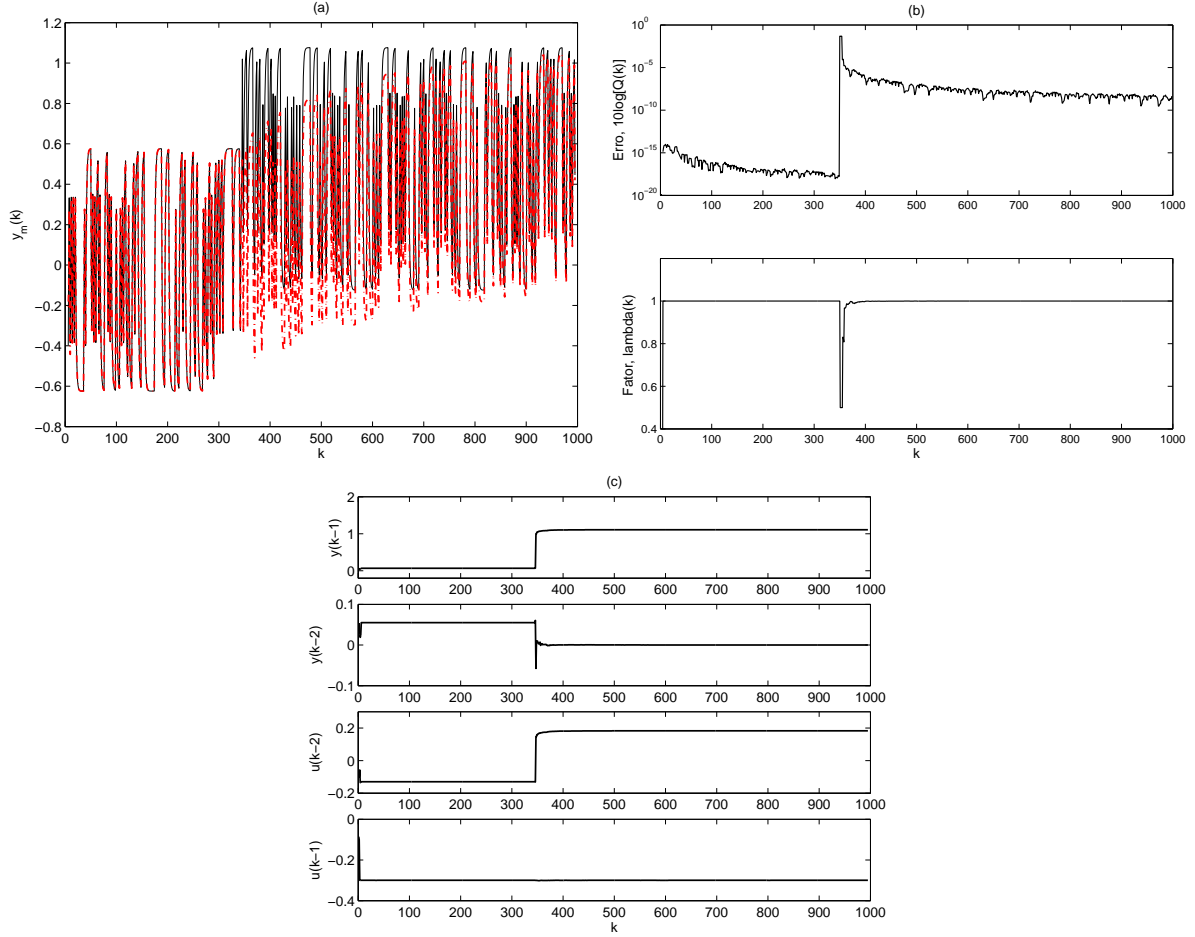


Figura 6.1: Modelagem do processo definido pelas equações (6.1) e (6.2). Análise realizada para o modelo 1 da Tabela 6.1 via RLSVFF. (a) Predição livre (-.-); (—) Série original. (b) No topo, erro de predição estendido, Q_k ; Embaixo, fator de esquecimento variável, λ_k . (c) De cima para baixo, respectivamente, os parâmetros estimados para os termos $y(k-1)$, $y(k-2)$, $u(k-2)$ e $u(k-1)$ do modelo.

$$\left\{ \begin{array}{l} C_1 \frac{dv_{C_1}}{dt} = \frac{v_{C_2} - v_{C_1}}{R} - i_d(v_{C_1}) \\ C_2 \frac{dv_{C_2}}{dt} = \frac{v_{C_1} - v_{C_2}}{R} w - i_L \\ L \frac{di_L}{dt} = -v_{C_2}, \end{array} \right. \quad (6.3)$$

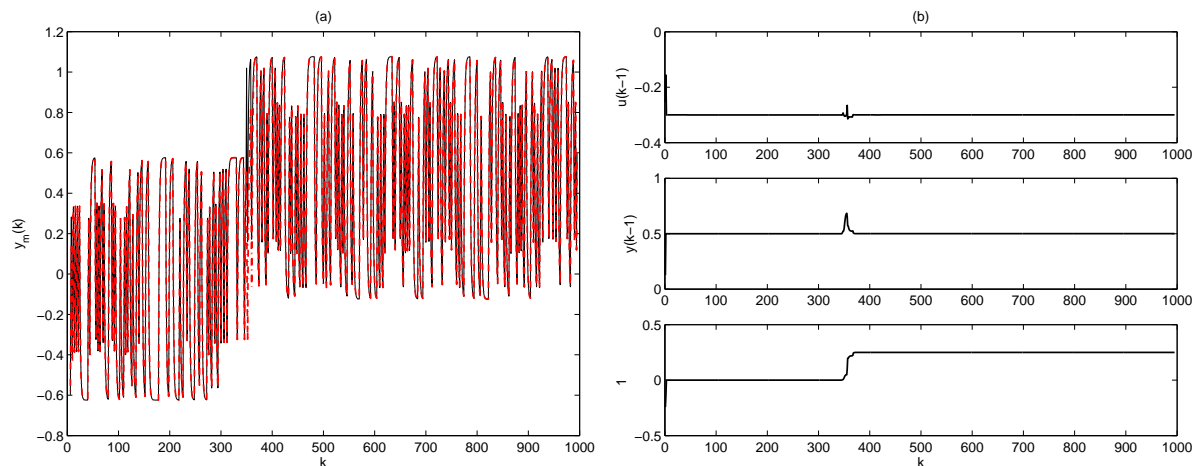


Figura 6.2: Modelagem do processo definido pelas equações (6.1) e (6.2). O modelo 2 da Tabela 6.1 foi usado e seus parâmetros estimados via RLSVFF, com a variável de projeto $\sigma^2 = 10^{-6}$. (a) Predição livre (-.-); Série original (—). (b) De cima para baixo, respectivamente, os parâmetros variantes no tempo dos termos $u(k-1)$, $y(k-1)$ e termo unitário.

em que v_{C_1} é a tensão sobre o capacitor C_1 , v_{C_2} a tensão sobre o capacitor C_2 e i_L a corrente que passa através do indutor.

Utilizou-se dados obtidos por Junior (2003) por meio do Protótipo para Controle do Circuito de Chua (PCChua), descrito e implementado por Tôrres (2001). A tensão no capacitor C_1 (v_{C_1}) é a variável observada e amostrada a cada 0,01s.

A Figura 6.3 mostra um período de observação em que se vê uma mudança de média. Para identificação de um modelo usou-se as observações a partir do instante $k = 186$ desses dados. Por meio do método TV-OPS escolheu-se o modelo 1 da Tabela 6.2. Na simulação um passo a frente de todo o período obteve-se índice $rmse = 0,0274$ após estimar-se os parâmetros variantes no tempo via RLSVFF.

Na Figura 6.4 pode-se ver o resultado da análise realizada com o objetivo de identificar o instante de mudança de dinâmica do sinal v_{C_1} observado no período supracitado. Vê-se nos gráficos (a) e (b) que os coeficientes dos regressores 1 e $y(k-22)$ respectivamente parecem responder à mudança de média que ocorre em torno do instante $k = 185$. Como se pode esperar, esses coeficientes assumem valores constantes após convergência, que são diferentes antes e depois da mudança de média. Nos gráficos (c) e (d) vê-se que o erro de predição estendido, $Q(k)$, e o fator de esquecimento variável, $\lambda(k)$, identificam o momento de mudança brusca da média da série temporal estudada. Atentando-se para a média do

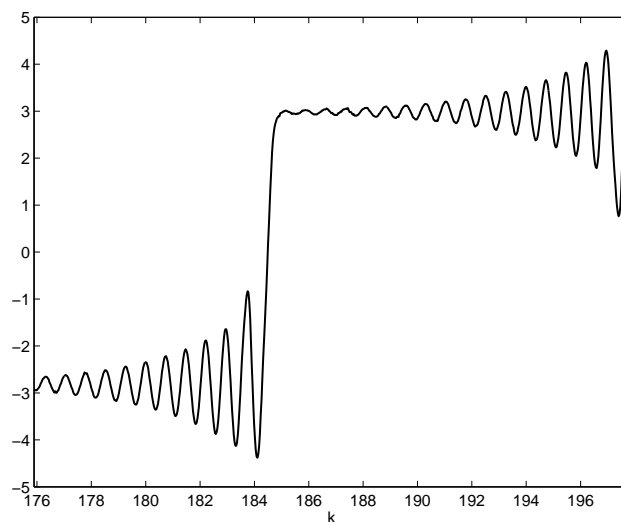


Figura 6.3: Série temporal obtida por meio da observação do sinal de tensão v_{C_1} , ver equação (6.3), do circuito de Chua implementado por Tôrres (2001).

Tabela 6.2: Modelos polinomiais AR/NAR usados na modelagem da série temporal obtida do circuito de Chua. Modelos obtidos por meio da metodologia TV-OPS.

Modelo 1	Modelo 2
1	$y(k-2)y(k-3)$
$y(k-22)$	$y(k-2)y(k-2)$
$y(k-30)$	$y(k-2)y(k-4)$
$y(k-18)$	$y(k-3)y(k-3)$
$y(k-11)$	$y(k-2)y(k-7)$
$y(k-10)$	$y(k-1)y(k-3)$

índice $Q(k)$, este também sugere que, para o período observado, a dinâmica antes da mudança da média é diferente da dinâmica depois de mudança.

6.1.2 Caso2: Circuito de Chua

A placa do circuito de Chua implementada por Tôrres (2001), foi construída de forma a permitir variações no resistor R (ver equação (6.3)). Este componente foi utilizado por Junior (2003) para mudar a dinâmica do sistema durante o tempo de observação de v_{C_1} , sendo que as demais variáveis foram mantidas constantes nesse período. Investiga-se a possibilidade de se determinar o instante de mudança da dinâmica por meio da análise de

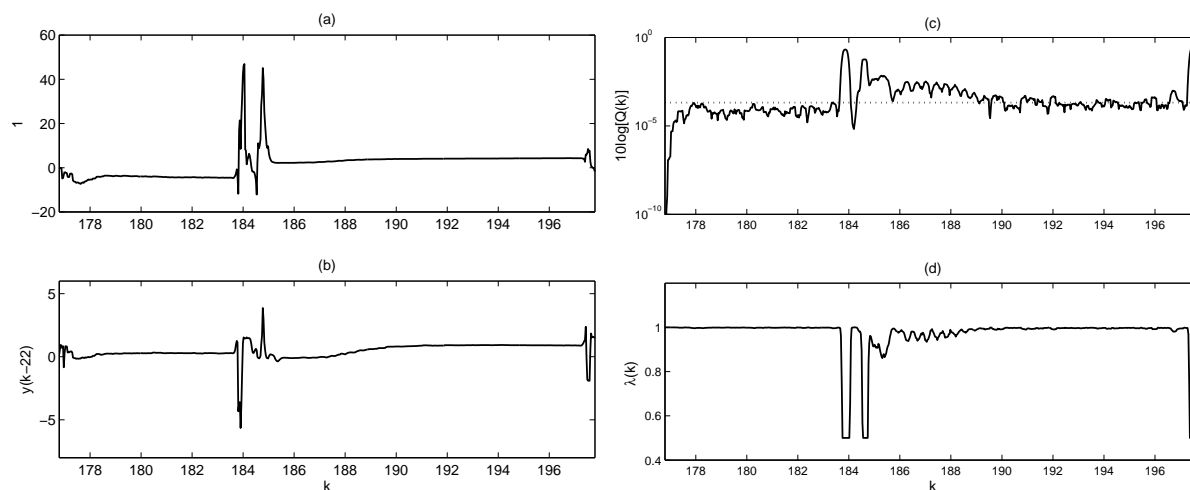


Figura 6.4: Análise via método RLSVFF da série de Chua com mudança de média, conforme modelo 1 da Tabela 6.2. (a) Parâmetro do regressor 1. (b) parâmetro do regressor $y(k-22)$. (c) Erro de predição estendido, $Q(k)$. (d) Fator de esquecimento variável, $\lambda(k)$.

série temporal.

Junior (2003) realizou o seguinte experimento: dados foram gerados durante 120 s com $R = 1,549 \Omega$, e o sistema operando em regime caótico dupla volta. Neste instante, ainda com o circuito de Chua ligado, variou-se a resistência para um determinado valor, até então desconhecido, porém mantendo o regime dupla volta. Após passados mais 120 s, desligou-se o circuito e mediu-se o valor da resistência. O valor obtido foi $R = 1,615 \Omega$. O período de amostragem foi de 0,01 s.

Vê-se na Figura 6.5 a série temporal obtida pela observação de v_{C_1} , durante o experimento citado. É interessante citar que os dados usados na seção anterior, Figura 6.3, correspondem a um período de pouco mais de 20 segundos dos mesmos dados obtidos neste experimento.

A série temporal $v_{C_1}(k)$ foi reamostrada com fator de decimação igual a 3. Aproximadamente as amostras obtidas nos primeiros 31 segundos de observação foram usadas para identificação de um modelo paramétrico NAR. Por meio da metodologia TV-OPS, o modelo 2 da Tabela 6.2 foi escolhido. Os parâmetros variantes no tempo foram estimados via RLSVFF e um índice $rmse = 0,0537$ foi obtido para predição 1 passo a frente usando-se todas as amostras.

Atente-se para a Figura 6.6 com o resultado dessa análise. A seção A indicada nos

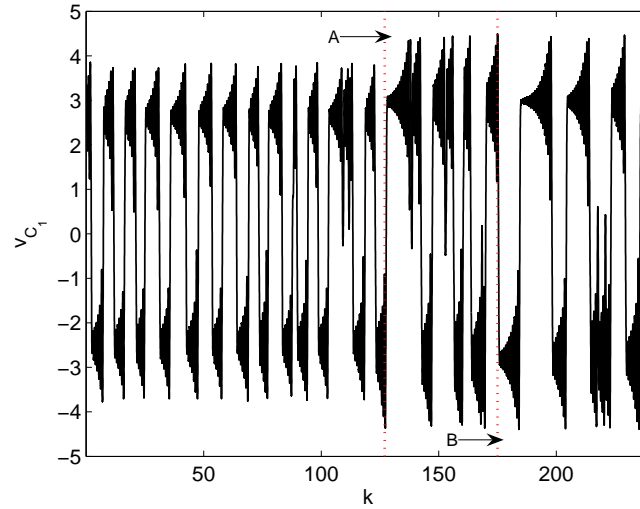


Figura 6.5: Série temporal obtida por meio da observação do sinal de tensão v_{C_1} , ver equação (6.3), do circuito de Chua implementado por Tôres (2001). O instante de observação k é incrementado a cada 0,01s.

gráficos desta figura, como também na Figura 6.5, apontam o instante de 120 s quando foi alterado o valor da resistência R do circuito de Chua e, por conseqüência, seu padrão dinâmico. Os gráficos (a) e (b) da Figura 6.6 mostram, respectivamente, os parâmetros estimados para os regressores $y(k-2)y(k-3)$ e $y(k-2)y(k-2)$. Estes não trazem indícios claros de alguma alteração após os primeiros 120 s.

Nos gráficos (c) e (d) da Figura 6.6, o erro de predição estendido, $Q(k)$, e o fator de esquecimento variável, $\lambda(k)$, parecem sugerir que algo aconteceu em torno do tempo de 120 s (secção A). Mas é importante ressaltar o fato que no instante apontado pela secção B os mesmos sugerem uma nova alteração no regime dinâmico. Isso parece coincidir com o que se observa a partir do mesmo instante na Figura 6.5. Junior (2003) diz que houve uma perda de sincronismo após o tempo de 120 s mas que o mesmo durou apenas alguns segundos e não o espaço de tempo entre as secções A e B. Sugere-se que nesse trecho tenha ocorrido um problema de convergência para o valor final da resistência R , considerando o fato de que sua mudança no circuito PCChua (Tôres, 2001) é feita manual e gradualmente, embora Junior (2003) declara tê-la feito rapidamente.

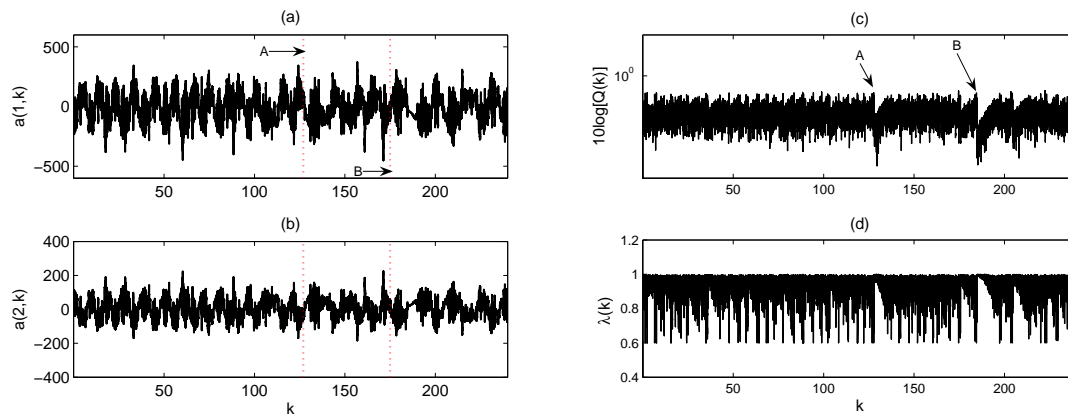


Figura 6.6: Análise da série de Chua conforme modelo 2 da tabela 6.2 via RLSVFF. (a) Parâmetro estimado para o regressor $y(k-2)y(k-3)$. (b) Parâmetro estimado para o regressor $y(k-2)y(k-2)$. (c) Erro de predição estendido, $Q(k)$. (d) Fator de esquecimento variável, $\lambda(k)$.

6.2 Tendência

Nesta seção, deseja-se inicialmente produzir, por meio de simulação, uma série temporal com tendência. A série temporal obtida do modelo da equação (6.1) é modificada com a adição de uma tendência determinística (ver Seção 2.1.1.2), conforme a seguinte equação:

$$y_t(k) = y(k) + (0,0033 \times k - 0,0033), \quad k = 1, \dots, 600, \quad (6.4)$$

em que $y(k)$ é uma observação gerada pelo processo da equação (6.1) e $y_t(k)$ é a nova observação modificada.

Dentre um conjunto de modelos NARX e ARX, o método TV-OPS indicou claramente a estrutura do modelo 1 da Tabela 6.3.

Atente-se à Figura 6.7, a qual mostra os resultados da análise via TV-OPS. No gráfico (b) vê-se os parâmetros estimados para o modelo. Ressalta-se o fato de que os parâmetros dos termos regressores $y(k-1)$ e $u(k-1)$ são iguais aos parâmetros do modelo ARX original da equação (6.1). O regressor 1 é responsável por modelar a tendência.

Vale observar que para efetuar a predição livre (Figura 6.7a) além do período de observação usado na escolha do modelo, ou seja, $k > 600$, aproximou-se os parâmetros

Tabela 6.3: Modelos ARX/AR usados na modelagem de séries com tendência. O modelo 1 foi usado para modelar o processo formado pelas equações (6.1) e (6.4). O modelo 2 foi usado para modelar a série de produção mensal de energia elétrica da Austrália. O modelo 3 modela a mesma série após remoção da tendência. Metodologia TV-OPS.

Modelo 1	Modelo 2	Modelo 3
$y(k-1)$	1	1
1	$y(k-1)$	$y(k-16)$
$u(k-1)$	$y(k-3)$	$y(k-1)$
	$y(k-4)$	$y(k-17)$
	$y(k-16)$	$y(k-4)$
	$y(k-25)$	$y(k-18)$
	$y(k-24)$	$y(k-27)$
	$y(k-28)$	$y(k-29)$
	$y(k-12)$	$y(k-13)$
	$y(k-23)$	$y(k-11)$

variantes no tempo por um polinômio de primeiro grau. Obteve-se um índice $rmse = 6,1578 \times 10^{-15}$ para a predição feita nessas condições.

Os resultados obtidos pela metodologia RLSVFF foi similar ao alcançado pelo método TV-OPS, entretanto obteve-se um índice $rmse = 0,4382$ ($\sigma^2 = 10^{-6}$).

6.2.1 Caso: Produção Mensal de Energia Elétrica

Como exemplo de uma série temporal real com tendência, analisou-se a produção mensal de energia elétrica da Austrália entre janeiro de 1956 e agosto de 1995 (Hyndman, 2004). Esses dados foram normalizados e apresentados na Figura 6.8.

Para a identificação de um modelo que descrevesse o comportamento dessa série, foram usadas todas as amostras pelo método TV-OPS. O modelo 2 da Tabela 6.3 foi escolhido e seus parâmetros estimados pelo método RLSVFF. Um índice $rmse = 0,0991$ foi obtido ao se realizar predição livre com os mesmos dados usados para identificação do modelo e estimação dos parâmetros. Com o objetivo de se fazer uma comparação, o modelo 3 foi usado para modelar a mesma série porém com a tendência de média removida. A tendência foi modelada por meio de um polinômio de grau 5.

A Figura 6.9 mostra os resultados da análise realizada usando-se o modelo 2 da Tabela 6.3. Vê-se que os parâmetros dos termos regressores mais relevantes, regressor 1 no gráfico (a) e regressor $y(k-1)$ em (b), não assumem com clareza a tendência da série como ocorreu

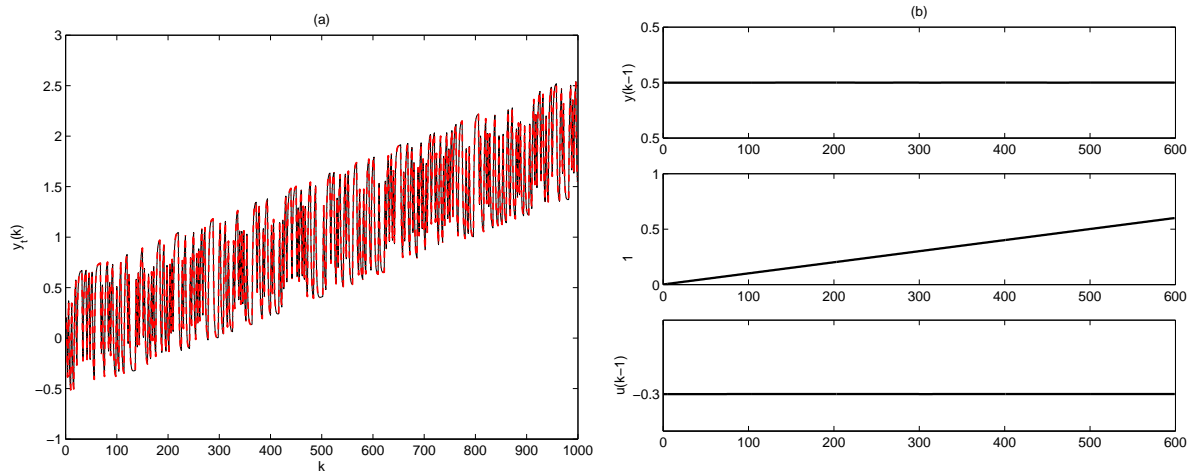


Figura 6.7: Modelagem do processo ARX não-estacionário definido pelas equações (6.1) e (6.4). Por meio método TV-OPS foi escolhido o modelo 2 da Tabela 6.1. Para a expansão dos coeficientes desse modelo foram usadas 2 funções de Legendre e 6 de Walsh. (a) Predição livre (-.-), que é real após $k = 600$; Série original (—). (b) De cima para baixo, respectivamente, os coeficientes dos termos $y(k - 1)$, termo unitário e $u(k - 1)$.

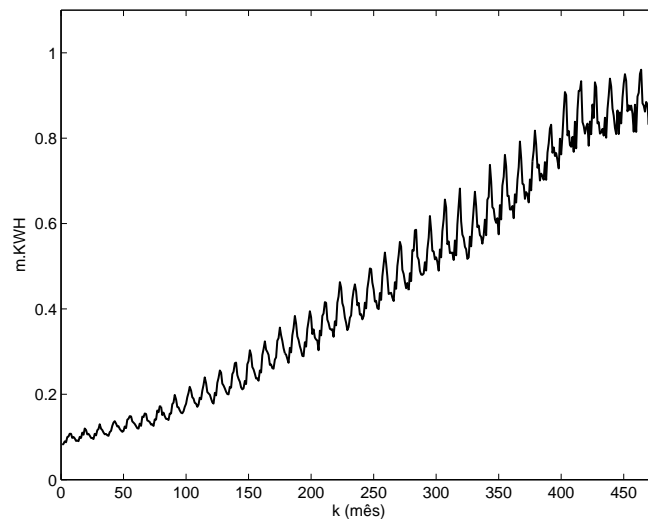


Figura 6.8: Série de produção mensal de energia elétrica na Austrália, observada entre janeiro de 1956 e agosto de 1995. Valores de produção normalizados.

no exemplo simulado. De fato nenhum dos outros parâmetros o fazem. A Figura 6.10 mostra a análise realizada com o modelo 3 da Tabela 6.3, considerando-se a série sem tendência.

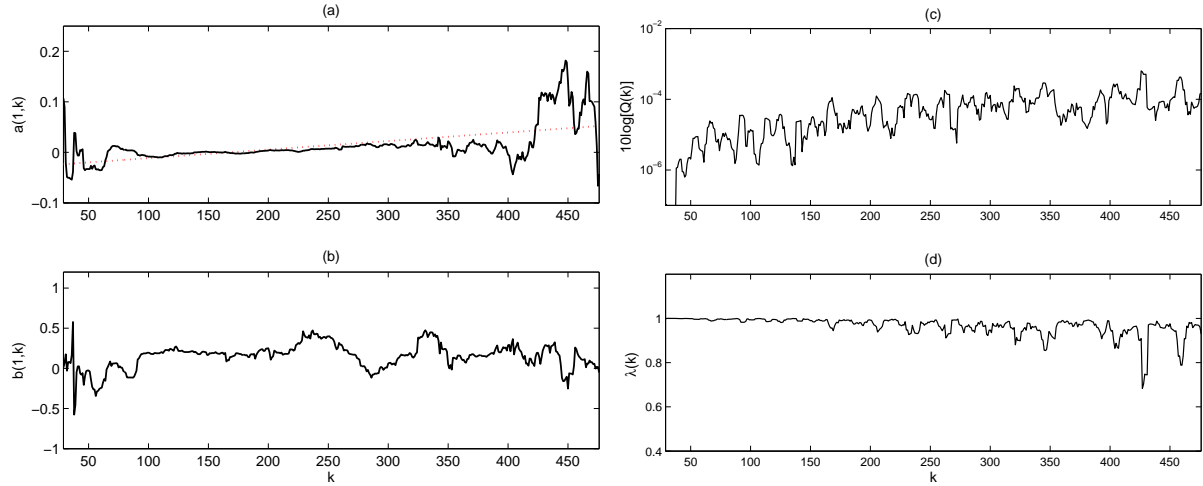


Figura 6.9: Análise da série de produção de energia da Austrália, conforme modelo 2 da Tabela 6.2 via RLSVFF. (a) (—) Parâmetro estimado para o regressor 1; ajuste polinomial de uma reta (...). (b) Parâmetro estimado para o regressor $y(k-1)$. (c) Erro de predição estendido, $Q(k)$. (d) Fator de esquecimento variável, $\lambda(k)$.

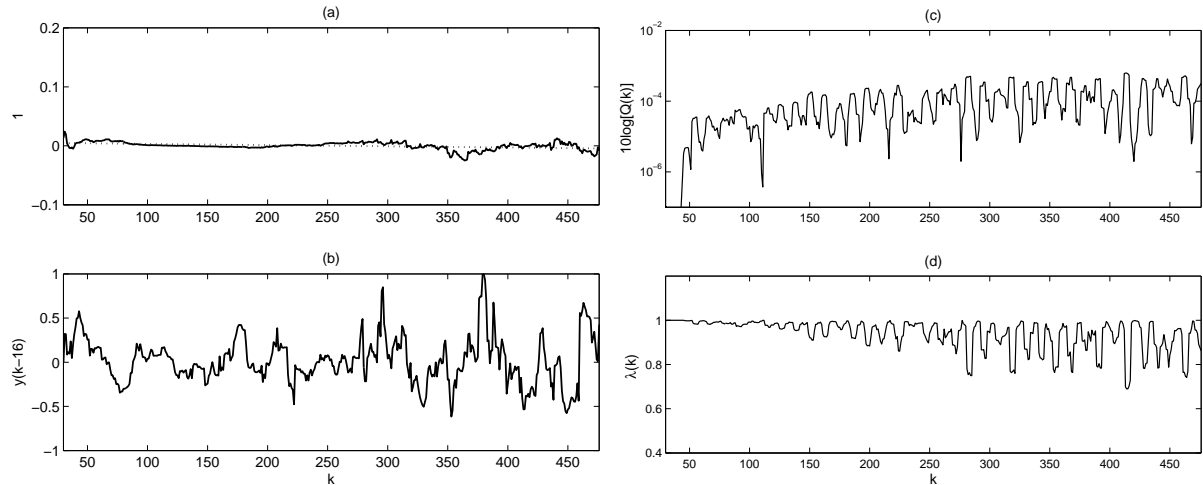


Figura 6.10: Análise da série de produção de energia da Austrália sem tendência, conforme modelo 3 da Tabela 6.2 via RLSVFF. (a) (—) Parâmetro estimado para o regressor 1; ajuste polinomial de uma reta (...). (b) Parâmetro estimado para o regressor $y(k-16)$. (c) Erro de predição estendido, $Q(k)$. (d) Fator de esquecimento variável, $\lambda(k)$.

Comparando-se o parâmetro do regressor unitário da Figura 6.9a com o parâmetro do mesmo regressor na Figura 6.10a, cenário em que a tendência da série é removida,

observa-se uma leve ascendência no valor médio do coeficiente do primeiro cenário. Por outro lado, comparando-se as Figuras 6.9c/d com as Figuras 6.10c/d, não se observam diferenças relevantes no comportamento do índice de erro estendido, $Q(k)$, e do fator de esquecimento variável, $\lambda(k)$. Em ambos os cenários o índice $Q(k)$ aumenta à medida que o tempo de observação aumenta e $\lambda(k)$ responde gradualmente em declínio à medida que o erro aumenta

6.3 Conclusões

Neste capítulo investigou-se a possibilidade de detecção do instante em que ocorrem mudanças no regime dinâmico de algumas séries temporais. Foram estudados cenários em que ocorreram mudança brusca na média da série temporal e mudança brusca num variável interna do sistema em observação. Séries temporais com tendência também foram investigadas.

Aplicando-se a metodologia RLSVFF, foi possível detectar com facilidade a simples mudança de média de um pequeno trecho da série de Chua. Quando efetuou-se uma mudança brusca numa variável do sistema PCChua, a detecção do momento em ocorre essa alteração não pôde ser facilmente determinado. Os índices $Q(k)$ e $\lambda(k)$ sugerem mudanças não somente em torno do tempo de 120 s mas também em um momento posterior. Problemas de convergência do novo valor de R podem ter contribuído com a incerteza.

Analisando-se a série mensal de produção de energia elétrica da Austrália, não é possível afirmar com clareza a cerca do comportamento com tendência da série temporal. Não-obstante, observou-se variações graduais nos índices $Q(k)$ e $\lambda(k)$, tanto no cenário original quanto no cenário em que a tendência foi removida. Esse padrão mostra uma variação crescente do padrão dinâmico e pode ser entendido como tendência. Ressalta-se que a tendência removida da série de produção de energia no cenário comparativo, corresponde aproximadamente aos valores médios dos vários períodos sazonais que compõem essa série. Entretanto, há uma outra tendência não removida que corresponde ao aumento da amplitude dos picos de cada período sazonal ao longo do tempo de observação (Figura 6.8). Esse fato explica a aparente contradição entre os cenários original e o cenário em que a tendência de média é removida, no que diz respeito aos índices $Q(k)$ e $\lambda(k)$.

Baseado nos resultados obtidos, pode-se dizer que se a estrutura do modelo do processo em estudo tende à correta, é possível verificar a detecção de mudanças na dinâmica em poucos parâmetros, como nos casos em que houve variação da média da série ou até mesmo no caso com tendência. Muitos parâmetros variando poderiam indicar uma es-

estrutura incorreta, um sinal com perfil mais aleatório ou ainda conseqüências da mútua interferência entre os parâmetros como foi exposto na Seção 5.2.3.

Discussão e Conclusões Finais

Parece ser, de fato, unânime entre pesquisadores que os sistemas reais em geral, sofrem, em alguma medida, variação. Essa variação poderia estar relacionada simplesmente à alteração de um parâmetro ou até mesmo a mudanças no padrão dinâmico do sistema. Neste último caso, esse comportamento é referido como *não-estacionariedade* quando refere-se a processos estocásticos. Para alguns casos de estudo a não-estacionariedade é vista como um complicador desnecessário, mas para outros é tida como fonte de informação relevante. Apesar da relevância que possa ser atribuída à não-estacionariedade, o fato é que se trata de um fenômeno que precisa ser considerado na identificação e modelagem de sistemas, ainda que, oportunamente, conclua-se que possa ser desconsiderado.

Não obstante, há uma certa subjetividade no tratamento empírico da não-estacionariedade, embora se tenha uma definição matemática muito bem definida. Uma série temporal pode ser considerada não-estacionária por um período de observação mas num instante futuro a mesma poderia mostrar-se ciclo-estacionária ou até mesmo possuir um comportamento *não-linear* antes não considerado. É certo, portanto, que há um ganho para interpretação adequada do comportamento dinâmico de um sistema quando o tempo de observação da variável de interesse é satisfatório para cobrir as variações no regime dinâmico, incluindo modificações bruscas e lentas ou até mesmo um comportamento não-linear do sistema. Ressalta-se que, por razões semelhantes às supracitadas, o período de amostragem da variável em observação também é importante para a posterior caracterização adequada do sistema.

Ainda que os tempos *essenciais* do sistema em observação sejam cobertos, os procedimentos de medição podem contribuir positiva ou negativamente dependendo da taxa de ruído advindo desses procedimentos. Esse ruído seria adicionado ao ruído inerente à estrutura interna do sistema, levando a um desafio ainda maior na determinação da não-estacionariedade.

Uma outra reflexão que pode-se fazer ainda antes de discutir-se os resultados deste

trabalho, é a questão da representatividade da variável observada. Uma determinada variável, em hipótese, pode refletir o comportamento de várias outras que estão ocultas ao observador. Desconsiderando-se os sistemas em que essas variáveis são bem definidas, a questão da não-estacionariedade e outras características poderiam ser caracterizadas de maneira inadequada simplesmente pelo desconhecimento ou incapacidade de obter-se informações que revelem de fato o comportamento do sistema.

Ljung (2001) fez uma interessante consideração: *rastrear as propriedades de um sistema é sempre uma questão de avaliação crítica das observações obtidas do processo em questão: as observações contém informações acerca das mudanças no processo ou estão somente dominadas por distúrbios aleatórios?*

Atento a essas observações preliminares, optou-se pela investigação da característica de não-estacionariedade por meio da modelagem paramétrica de sistemas dinâmicos discretos variantes no tempo, com uso de modelos polinomiais regressivos, lineares e/ou não-lineares. A idéia central dessa estratégia é expressar as variações temporais no padrão dinâmico de um determinado sistema por meio de variações temporais nos coeficientes do modelo. Para tanto, foi necessário assumir que a estrutura do modelo não muda ao longo do tempo de observação, conforme foi visto no Capítulo 5. Nos sistemas simulados a estrutura dos modelos foi mantida constante. Nos casos reais, o modelo identificado foi considerado invariante no tempo para que as análises fossem feitas.

Se admitir-se variação na estrutura do processo, a série temporal observada desse processo deve ser dividida e um modelo específico identificado para cada parte. A partir de cada modelo considerado foram feitas as análises propostas neste trabalho. Embora admita-se o fato de que a não-estacionariedade possa ser também decorrente da variação temporal da estrutura do modelo, esse caso não foi considerado neste trabalho.

7.1 Discussão

Foram empregadas três metodologias na análise de séries temporais observadas de sistemas com características não-estacionárias em sua dinâmica. O método *TV-OPS* envolve uma metodologia de seleção de estrutura (*OPS*) e promove o rastreamento dos coeficientes do modelo polinomial a partir da expansão dos mesmos por meio das funções base de Walsh e Legendre. Os métodos *RLSVFF* e *VLSWBL* promovem a busca por variações no padrão dinâmico rastreando os coeficiente por meio da técnica de mínimos quadrados regressivo e em batelada respectivamente.

Dentre as três metodologias, TV-OPS inclui a técnica OPS¹, que auxilia na seleção da estrutura de modelos polinomiais. Em toda seleção de estrutura feita neste trabalho, exceto onde indicado, usou-se o método OPS. Não houve um exercício comparativo entre os métodos OPS e OS/ERR, mas, baseando-se em resultados de seleção de estrutura de sistemas simulados, considerou-se satisfatório o desempenho de OPS. Uma vantagem no uso desse método é que ele está inserido no contexto de modelagem de sistemas não-estacionários. Uma outra vantagem em potencial é a possibilidade do uso de outras funções base além de Legendre e Walsh, o que poderia agregar valor na definição da estrutura. Uma desvantagem é a subjetividade envolvida na definição de um valor *limiar* usado na comparação com o resíduo de ajuste entre regressores. Essa comparação determina se os regressores são ou não linearmente independentes. Neste trabalho, o valor do *limiar* usado foi de 10^{-6} . Esse valor foi determinado empiricamente após testes em sistemas simulados.

Baseado nas investigações do capítulo 5 e 6, pôde-se refinar a declaração de Ljung (2001) mencionando-se vários fatores que estão envolvidos na análise de um sistema não-estacionário, especialmente com o uso das metodologias empregadas neste trabalho:

- houve mudanças no processo ou as variáveis de projeto usadas pelas metodologias de análise não possuem valores adequados?
- a variação de um coeficiente é real ou estimulada pela variação em outro coeficiente?
- houve mudanças no processo ou o modelo escolhido representa incorretamente a dinâmica de todo processo?
- foi feita uma distinção adequada entre não-estacionariedade e não-linearidade?

Pôde-se verificar ao longo deste trabalho que todas as metodologias estudadas são capazes de rastrear parâmetros que variam em consequência de mudanças na dinâmica. A metodologia RLSVFF usa o índice de erro $Q(k)$ e/ou o fator de esquecimento variável $\lambda(k)$ para sugerir mudanças na dinâmica. O método VLSEWBLS também usa índices de erro e/ou também a variação do tamanho das janelas de dados para indicar não-estacionariedade. Considerando-se que os resultados foram semelhantes para as três metodologias, na Seção 5.1 optou-se por expor somente o método TV-OPS quanto à sua capacidade de apontar não-estacionariedades.

O critério de Akaike modificado, $AICm$, sugere a quantidade de funções base (Legendre e/ou Walsh) necessária para rastrear os parâmetros do modelo. Observou-se que tanto

¹O método OPS é apresentado na Seção 3.6.2.

para parâmetros constantes quanto para parâmetros que variam rápido e aleatoriamente, AIC_m aponta para a quantidade mínima de funções, ou seja, sugere que há estacionariedade. No último caso, embora haja variações bruscas e aleatórias, o algoritmo só consegue rastrear a média desses coeficientes. Outros cenários mostraram que esse índice pode indicar a característica de não-estacionariedade pela tentativa de montar um conjunto de funções base para descrever a variação temporal do parâmetro. Apesar de os testes terem sido realizados em cenários bem definidos, o procedimento matemático mostra-se eficiente. Atente-se para o fato de que os gráficos “ $AIC_m \times No. Funções Base$ ” apresentam formas geométricas que poderiam ser relacionadas a um *grau* de não-estacionariedade para o sistema. Não foi feito um estudo comparativo desses gráficos mas somente uma inspeção visual. Sugere-se essa investigação em trabalhos futuros.

Discute-se a seguir dificuldades na interpretação do resultado da análise de séries temporais, causadas pelo uso inadequado dos métodos abordados.

Todas as metodologias abordadas possuem variáveis, chamadas *variáveis de projeto*, cujos valores são empiricamente determinados. Ficou claro que o valor escolhido para essas variáveis de projeto é muito importante para a interpretação adequada dos resultados da análise. O conhecimento prévio do sistema em estudo pode ajudar na definição das variáveis de projeto. Como é o caso do conhecimento do taxa de ruído que, como visto na seção 5.2.1, também pode comprometer a análise de resultados. Na metodologia RLSVFF por exemplo, o valor da variável σ^2 deve ser determinado a partir do valor da relação sinal-ruído (SNR) e pode minimizar o efeito do ruído.

Para análise de sistemas reais, determinados valores de variáveis, considerados *default* a partir da simulação de outros sistemas, devem ser inicialmente empregados. Em seguida novas análises devem ser feitas a partir de variações nessas mesmas variáveis. Os resultados podem ser considerados à luz de algum conhecimento prévio a cerca do processo em estudo. Sugere-se que esse mesmo conhecimento possa conduzir um ajuste nos valores das variáveis de projeto, de maneira que valores *default* possam ser definidos para determinados tipos de processos. Esse procedimento visa contribuir para a interpretação adequada do comportamento dinâmico do processo.

Outro fato também interessante é a variação de um determinado parâmetro causar uma perturbação ou variação errada em outro parâmetro. Pudemos certificar que variações bruscas ou suaves em um parâmetro causa perturbações em outro parâmetro constante. As metodologias RLSVFF e VLSWBLs usam o mesmo princípio geral de redução da janela de dados recentes à medida que o erro de predição aumenta. Quando isso acontece, os parâmetros constantes são mais afetados ao serem estimados com uma

quantidade de dados menor. Com relação aos parâmetros variantes no tempo, esse efeito de redução da janela de dados passa a ser referido como maior ou menor sensibilidade de rastreamento. Quanto maior o poder de adaptação às mudanças, maior a sensibilidade aos efeitos indesejados. É interessante, portanto, empregar modelos com estrutura mínima. Termos desnecessários ou com uma desprezível contribuição no rastreamento da dinâmica podem levar à uma detecção espúria de não-estacionariedade.

Os fatores de dificuldades comentados acima podem ocorrer independentemente da estrutura do modelo usada. Em outras palavras, pode-se afirmar que ainda que a estrutura do modelo seja a correta, podem ocorrer interferência mútua entre coeficientes e as variáveis de projeto ainda precisam ser bem definidas. Não obstante, num cenário como esse as metodologias se mostraram capazes de rastrear parâmetros variantes no tempo e detectar instantes de mudança de dinâmica.

Um outro ponto investigado foi a questão da distinção entre não-linearidade e não-estacionariedade. Concluiu-se que é possível diferenciar-se um de outro desde que a estrutura do modelo representativo do processo seja correto ou que pelos menos seja capaz de *absorver* suas não-linearidades. Caso contrário não se pode afirmar nada. Pode-se definir que a *não-linearidade* é uma questão ligada à estrutura do modelo e a *não-estacionariedade* expressa nas variações dos parâmetros desse modelo. Semelhantemente, Söderström & Stoica (1989) mostra que ciclos (sazonalidades) podem ser modelados por modelos de segunda ordem. De fato, em experimentos com sazonalidades não incluídos nesse trabalho, pôde-se ver que quando a sazonalidade foi corretamente *absorvida* pela estrutura do modelo, os parâmetros do modelo se mostraram constantes, sugerindo a estacionariedade do processo em estudo.

A questão da seleção correta da estrutura do modelo escolhido para o processo torna-se crucial para o sucesso no uso dessas metodologias. Além do problema de sobreparametrização em modelos com estruturas incorretamente especificadas, esses modelos estão sujeitos a regimes dinâmicos espúrios. Concluiu-se que a variação dos parâmetros de um modelo incorreto em geral não compensa o erro da estrutura. Ainda que para processos reais seja difícil determinar uma estrutura “correta” para o modelo, acredita-se que certas características da dinâmica do processo possam ser analisadas por meio do “melhor” modelo obtido.

Atente-se, por exemplo, à questão da detecção do instante de mudança da dinâmica abordado no Capítulo 6. Diante de todos os desafios à modelagem paramétrica de sistemas não-estacionários, a detecção do instante de mudança do padrão dinâmico se mostrou mais robusto, principalmente para mudanças bruscas. Certamente o problema da sobre-

parametrização (Mendes & Billings, 1998) dificulta a tomada de decisão dos instantes de mudança devido principalmente à interferência mútua entre parâmetros. Não obstante, viu-se no casos de séries temporais simuladas com variações na média e séries com tendência, que, numa estrutura menor e mais próxima da correta (linear ou não-linear), um parâmetro ou um número mínimo deles poderiam se encarregar dessas variações.

Foi visto na Seção 6.1.1 que numa série real obtida de um circuito de Chua foi possível identificar satisfatoriamente o instante em que houve uma mudança brusca na média dessa série. Investigou-se também a possibilidade de detectar uma mudança interna do circuito de Chua na Seção 6.1.2. Percebe-se que há indícios de alguma mudança no instante em que de fato ela ocorreu, mas a mesma ferramenta aponta para uma outra suposta mudança, como pôde ser visto na Figura 6.6, seção *B*. Baseado no conhecimento prévio a cerca de um período necessário para convergência após a alteração no valor do resistor R (Seção *A*) (Junior, 2003) e também uma inspeção visual da Figura 6.5, RLSVFF parece indicar tanto o momento em que alteração da variável R perturbou o sistema e o momento em que R se estabiliza.

No caso da análise da série de produção de energia elétrica da Austrália, Seção 6.2.1, não foi possível identificar com clareza a tendência na *média*, por meio do coeficiente do regressor unitário conforme esperado. Entretanto, o índice de erro, $Q(k)$, e o fator de esquecimento, $\lambda(k)$, sugerem uma mudança gradual na dinâmica mesmo depois de retirada da tendência da média. Acredita-se que o aumento da amplitude do período sazonal contribuiu para esse fato. Acredita-se que se a sazonalidade fosse retirada ou mesmo absorvida adequadamente pela estrutura do modelo, a tendência da média poderia ser melhor rastreada.

Pode-se anuir que, baseados nos resultados fundamentais dos Capítulos 5 e 6, modelos mais adequados podem aumentar as possibilidades de uma análise com menos conhecimento prévio, com o objetivo de detecção de instantes de mudança no regime dinâmico do processo em estudo.

7.2 Conclusões Finais

Neste trabalho investigou-se a modelagem paramétrica de processos não-estacionários. Desejou-se compreender até que ponto as três metodologias abordadas no Capítulo 4 podem caracterizar um regime dinâmico variante no tempo.

No que diz respeito à questão matemática e considerando processos simulados bem de-

finidos, as metodologias mostraram-se bastante eficientes no rastreamento de parâmetros de sistemas variantes à luz das seguintes considerações básicas: *i*) a estrutura do modelo seja constante ao longo do tempo de observação; *ii*) o modelo corresponda exatamente ao processo gerador da série temporal;

As considerações supracitadas são fundamentais também na análise de séries temporais reais. Como não é possível garantir que um modelo represente *exatamente* um processo dinâmico real, deve-se buscar o melhor modelo com o menor número possível de regressores. Muitos parâmetros variando poderiam indicar uma estrutura incorreta, um sinal com perfil mais aleatório ou ainda conseqüências da mútua interferência entre os parâmetros como foi exposto na Seção 5.2.3.

Dentro do contexto deste trabalho, definiu-se que não-linearidade está relacionada à estrutura do modelo enquanto a não-estacionariedade é expressa nos coeficientes variantes no tempo dos termos do modelo. Não-linearidades não traduzidas na estrutura do modelo podem levar à uma detecção espúria de não-estacionariedade.

As variáveis de projeto das metodologias devem ter seus valores cuidadosamente definidos, alguns por meio de conhecimento do processo, como a taxa de ruído (SNR), e outros pelo método de tentativa e erro partindo de valores de casos já estudados.

Baseando-se nos resultados obtidos, a despeito de todas as dificuldades inerentes aos métodos utilizados, pode-se dizer que se a estrutura do modelo do processo em estudo *tende* à correta, é possível detectar mudanças na dinâmica de processos reais por meio da variação de parâmetros do modelo e/ou por meio de índices como o fator de esquecimento variável da metodologia RLFVFF. Acredita-se que a detecção de mudanças no regime dinâmico de processos não-estacionários seja a maior contribuição da modelagem paramétrica desses processos.

BIBLIOGRAFIA

- Abramowitz, M. & Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. U.S. Department of Commerce.
- Aguirre, A. & Aguirre, L. A. (2000). Time series analysis of monthly beef cattle prices with non-linear autoregressive models. *Journal of applied economics*, 32, 265–275.
- Aguirre, L. A. (1994). Some remarks on structure selection for nonlinear model. *International Journal of Bifurcation and Chaos*, 4(6), 1707–1714.
- Aguirre, L. A. (2004). *Introdução à identificação de Sistemas: Técnicas Lineares e Não-Lineares aplicadas a Sistemas Reais*. Editora UFMG. 2a edição.
- Aguirre, L. A. & Billings, S. A. (1995). Improved structure selection for nonlinear models based on term clustering. *International Journal of Control*, 62(3), 569–587.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automatic Control*, 19(6), 716–723.
- AMD, C. (2005). <http://www.amd.com>. Acesso em 06 de julho de 2005.
- Bandt, C. & Pompe, B. (2002). Permutation entropy: A natural complexity measure for time series. *Physical Review Letters*, 88(17).
- Basseville, M. & Nikiforov, I. V. (1993). *Detection of Abrupt Changes - Theory and Application*. Prentice-Hall: Englewood Cliffs, NJ. Version available for download at: <http://www.irisa.fr/sisthem/kniga/>.
- Belge, M. & Miller, E. L. (2000). A sliding window RLS-like adaptive algorithm for filtering alpha-stable noise. *IEEE Signal Processing Letters*, 7(4), 86–89.
- Bendat, J. S. & Piersol, A. G. (1986). *Random Data: Analysis and Measurement Procedures*. Wiley, New York.

- Bianchi, A. M., Mainardi, L., Petrucci, E., Signorini, M. G., Mainardi, M., & Cerutti, S. (1993). Time-variant power spectrum analysis for the detection of transient episodes in HRV signal. *IEEE Trans. on Biomedical Engineering*, 40(2), 136–144.
- Billings, S. A. & Aguirre, L. A. (1995). Effects of the sampling time and identification of nonlinear models. *Int. J. of Control*, 5(6), 1541–1556.
- Billings, S. A., Chen, S., & Korenberg, M. J. (1989). Identification of MIMO nonlinear systems using a forward-regression orthogonal estimator. *Int. J. of Control*, 49(6), 2157–2189.
- Billings, S. A. & Coca, D. (1999). Discrete wavelet models for identification and qualitative analysis of chaotic systems. *Int. J. Bifurcation and Chaos*, 9(7), 1263–1284.
- Billings, S. A. & Fakhouri, S. Y. (1978). Identification of nonlinear systems using wiener model. *Eletronics Letters*, 13(17), 502–504.
- Billings, S. A., Korenberg, M. J., & Chen, S. (1988). Identification of nonlinear output-affine systems using an orthogonal least-squares algorithm. *Int. Journal of Systems Science*, 19, 1559–1568.
- Boldrini, J. L., Costa, S. I. R., Figueiredo, V. L., & Wetzler, H. G. (1980). *Álgebra Linear* (3 ed.). Harper & Row do Brasil.
- Bouzeghoub, M. C., Ellacott, S. W., Easdown, A., & Brown, M. (2000). On the identification of non-stationary linear processes. *Int. J. of Systems Science*, 31(3), 273–286.
- Box, G. E. P. & Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day (Revised edition, 1976).
- Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control* (3rd ed.). Prentice-Hall International, Inc.
- Braga, A. P., Carvalho, A. C., & Ludemir, T. B. (2000). *Redes Neurais Artificiais: teoria e aplicações*. Livraria Técnica Científica.
- Braun, C., Kowallik, P., Freking, A., Haderer, D., Kniffki, K., & Meesmann, M. (1998). Demonstration of nonlinear components in heart rate variability of healthy persons. *Am. J. Physiology*, 275(44), H1577–H1584.
- Brockwell, P. J. & Davis, R. A. (1996). *Introduction to Time Series and Forecasting*. Springer-Verlag New York.

- Brooks, C. & Hinich, M. J. (1998). Episodic nonstationarity in exchange rates. *Applied Economics Letters*, 5, 719–722.
- Campello, R. J. G. B., Fávier, G., & Amaral, W. C. (2004). Optimal expansions of discrete-time Volterra models using Laguerre functions. *Automatica*, 40(5), 815–822.
- Cao, L. & Scharztz, H. M. (1999). A novel recursive algorithm for directional forgetting. *Proc. of the American Control Conference*, 1334–1338. San Diego, California.
- Cao, Y., wen Tung, W., Gao, J. B., Protopopescu, V. A., & Hively, L. M. (2004). Recurrence time statistics for chaotic systems and their applications. *Physical Review E*, 70(16).
- Casdagli, M. C. (1997). Recurrence plots revisited. *Physica D*, 108, 12–44.
- Cassini, C. (1999). Estimação recursiva de não-linearidades estáticas usando modelos narmax polinomiais. Master's thesis, Universidade Federal de Minas Gerais, PPGEE.
- Castro, L. M. S. S. (1996). *Wavelets e Séries de Walsh em Elementos Finitos*. PhD thesis, Universidade Técnica de Lisboa, Portugal.
- Cesarelli, M., Bifulco, P., & Bracale, M. (1997). Evaluating time-varying heart-rate variability power spectral density. *IEEE Engineering in Medicine and Biology*, 16(6), 76–79.
- Chen, C.-T. (1999). *Linear System: Theory and Design* (third ed.). Oxford University Press.
- Chen, S. & Billings, S. A. (1989). Representation of nonlinear systems the NARMAX model. *Int. J. of Control*, 49(3), 1013–1032.
- Chen, S., Cowan, C. F. N., & Grant, P. M. (1991). Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Trans. on Neural Networks*, 2(2), 302–309.
- Chien, T. M. (1975). On representations of Walsh functions. *IEEE Trans. Electromag. Compat.*, 7(3), 170–177.
- Cho, Y. S., Kim, S. B., & Powers, E. J. (1991). Time-varying spectral estimation using AR models with variable forgetting factors. *IEEE Transaction on Signal Processing*, 39(6), 1422–1426.

- Chua, L. O. (1994). Chua circuit: An overview 10 years later. *Journal of Circ., Sys and Computers*, 4(2), 117–159.
- Coelho, A. A. R. & Coelho, L. S. (2004). *Identificação de Sistemas Dinâmicos Lineares*. Editora da UFRJ, Florianópolis, SC.
- Coelho, M. C. S. (2002). Modelos de hammerstein e de wiener: conexões com modelos narx e sua aplicação em identificação de sistemas não-lineares. Master's thesis, Universidade Federal de Minas Gerais, (PPGEE).
- Coifman, R. & Wickerhauser, M. (1992). Entropy-based algorithms for best basis selection. *IEEE Trans. Infor. Theory*, 38, 713–718.
- Corrêa, M. V. (2001). *Identificação Caixa-Cinza de Sistemas Não-Lineares Utilizando Representações NARMAX Polinomiais e Racionais*. PhD thesis, Universidade Federal de Minas Gerais. (PPGEE).
- Davies, A. C. (1972). On the definition and generation of Walsh functions. *IEEE Trans. Comput.*, 21, 187–189.
- Desrochers, A. A. (1981). On an improved model reduction technique for nonlinear systems. *Automatica*, 17, 407–409.
- Eckmann, J. P., Kamphorst, S. O., & Ruelle, D. (1987). Recurrence plots of dynamical systems. *Europhys. Lett.*, 973(4).
- Eykhoff, P. (1974). *System Identification: parameter and state estimation*. Chichester, England: John Wiley & Sons.
- Fortescue, T. R., Kershenbaum, L. S., & Ydstie, B. E. (1981). Implementation of self-tuning regulators with variable forgetting factors. *Automatica*, 17, 831–835.
- Freitas, U. S. (2001). Uso de técnicas de detecção de estrutura na identificação de modelos dinâmicos contínuos não-lineares polinomiais. Master's thesis, Universidade Federal de Minas Gerais, (PPGEE).
- Gamero, L. G., Risk, M., Sobh, J. F., Ramirez, A. J., & Saul, J. P. (1996). Heart rate variability analysis using wavelet transform. *IEEE Conf. Computers in Cardiology*, 177–180.
- Gao, J. B. (1999). Recurrence time statistics for chaotic systems and their applications. *Physical Review Letters*, 83(16), 3178–3181.

- Gao, J. B. (2001). Detecting nonstationarity and state transitions in time series. *Physical Review E*, 63(6).
- Garcia, C. (1997). *Modelagem e simulação de processos industriais e de sistemas eletromecânicos*. EDUSP. 1a. edição, São Paulo.
- Gomes, M. E. D. (2001). *Técnicas de Sistemas Dinâmicos Não Lineares na Análise da Modulação Autônoma da Variabilidade da Frequência Cardíaca*. PhD thesis, Universidade Federal de Minas Gerais. (PPGEE).
- Gomes, M. E. D., Souza, A. V. P., Guimarães, H. N., & Aguirre, L. A. (2000). Investigation of determinism in heart rate variability. *Chaos*, 10(2), 398–410.
- Grassberger, P. & Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica D*, 9, 189–208.
- Grenier, Y. (1983). Time-dependent ARMA modeling of nonstationary signals. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 31(4), 899–911.
- Hagglund, T. (1995). Recursive estimation of slowly time-varying parameters. *The 7th IFAC Symposium on Identification and Parameter Estimation*, 1137–1142.
- Harmuth, H. F. (1972). *Transmission of Information by orthogonal functions*. Springer-Verlag, Berlin.
- Henderson, K. W. (1970). Comment on 'computation on fast Walsh-Fourier transform'. *IEEE Trans. Comput.*, 19, 850–851.
- Hwang, C. & Shih, Y. P. (1982). Parameter identification via Laguerre polynomials. *International Journal of Systems Science*, 13(2), 209–211.
- Hyndman, R. J. (2004). Time series data library. <http://www-personal.buseco.monash.edu.au/~hyndman/TSDL/>.
- Intel, C. (2005). <http://www.intel.com>. Acesso em 04 de julho de 2005.
- Islsker, H. & Kurths, J. (1993). A test for nonstationarity: Finding parts in a time series apt for correlation dimension estimates. *Int. J. Bifurcation Chaos Appl. Sci. Eng.*, 3(6), 1573–1579.
- Jácome, C. R. F. (1996). Uso de conhecimento prévio na identificação de modelos polinomiais narmax. Master's thesis, Universidade Federal de Minas Gerais. (PPGEE).

- Jiang, J. & Zhang, Y. (2004a). A novel variable-length sliding window blockwise least-squares algorithm for on-line estimation of time-varying parameters. *Int. J. Adapt. Control Signal Process.*, 18, 505–521.
- Jiang, J. & Zhang, Y. (2004b). A revisit to block and recursive least squares for parameter estimation. *Computers and Electrical Engineering*, 30, 403–416.
- Judd, K. & Small, M. (2000). Towards long-term prediction. *Physica D*, 136, 31–44.
- Junior, H. B. (2003). Métodos de estimação recursiva baseados no filtro de kalman aplicados a sistemas não-lineares. Master's thesis, Universidade Federal de Minas Gerais, (PPGEE).
- Kandil, M. S., El-Debeiky, S. M., & Hasanien, N. E. (2001). Overview and comparison of long-term forecasting techniques for a fast developing utility: part-I. *Electric Power System Research*, 58, 11–17.
- Kennedy, M. P. (1992). Robust op amp realization of chua's circuit. *Frequenz*, 46(3-4), 66–80.
- Kennel, M. B. (1997). Statistical test for dynamical nonstationarity in observed time-series data. *Physical Review E*, 56(1), 316–321.
- Korenberg, M. J. (1985). Orthogonal identification of nonlinear difference equation models. *Proceedings of the 28th Midwest Symposium on Circuits and Systems*, 90–95.
- Korenberg, M. J. (1989a). A fast orthogonal search method for biological time-series analysis and system identification. *IEEE Inter. Conf. on Systems, Man and Cybernetics*, 2, 459–465.
- Korenberg, M. J. (1989b). A robust orthogonal algorithm for system identification and time-series analysis. *Biological Cybern.* 1, 1(69), 267–276.
- Korenberg, M. J., Billings, S. A., Liu, Y. P., & McIlroy, P. J. (1988). Orthogonal parameter estimation algorithm for nonlinear stochastic systems. *Int. J. of Control*, 48, 193–210.
- Korenberg, M. J. & Paarmann, L. D. (1991). Orthogonal approaches to time-series analysis and system identification. *IEEE Signal Processing Magazine*, 8(3), 29–43.
- Lee, Y. W. (1933). Synthesis of electrical networks by means of the Fourier transform of Laguerre functions. *Journal of Maths and Phys.*, 11, 83–113.

- Leon-Garcia, A. (1994). *Probability and Random Processes for Electrical Engineering* (2nd ed.). Addison-Wesley.
- Leontaritis, I. J. & Billings, S. A. (1985a). Input-output parametric models for nonlinear systems. Part I: Deterministic nonlinear systems. *Int. J. of Control*, *41*, 303–328.
- Leontaritis, I. J. & Billings, S. A. (1985b). Input-output parametric models for nonlinear systems. Part II: Stochastic nonlinear systems. *Int. J. of Control*, *41*, 329–344.
- Liao, G.-C. & Tsao, T.-P. (2004). Application of fuzzy neural networks and artificial intelligence for load forecasting. *Electric Power System System Research*, *70*, 237–244.
- Ljung, L. (1987). *System Identification, theory for the user*. New Jersey: Prentice Hall.
- Ljung, L. (2001). Recursive identification algorithms. Technical report, Department of Electrical Engineering, Linköping, Sweden. Report no.: LiTH-ISY-R-2366.
- Ljung, L. & Gunnarsson, S. (1990). Adaptation and tracking in system identification - a survey. *Automatica*, *26*(1), 7–21.
- Lu, S., Ju, K. H., & Chon, K. H. (2001). A new algorithm for linear and nonlinear ARMA model parameter estimation using affine geometry. *IEEE Trans. on Biomedical Engineering*, *48*(10), 1116–1124.
- Mandelbrot, B. B. (1982). *The Fractal Geometry of Nature*. Freeman, São Francisco.
- Manuca, R. & Savit, R. (1996). Stationarity and nonstationarity in time series analysis. *Physica D*, *99*, 134–161.
- Martin, N. (1986). An AR spectral analysis of nonstationary signals. *Signal Processing*, *10*, 61–74.
- Mendes, E. M. A. M. & Billings, S. A. (1998). On overparametrization of nonlinear discrete systems. *Int. J. Bifurcation and Chaos*, *8*(3), 535–556.
- Mingoti, S. A. & Neves, O. F. (1999). A metodologia de geoestatística como alternativa na análise de séries temporais. *Revista Escola de Minas, Ouro Preto*, *52*(3), 182–187.
- Mohan, B. & Datta, K. (1988). Lumped and distributed parameter system identification via shifted Legendre polynomials. *Trans. of the ASME, J. of Dynamic System, Measurement and Control*, *110*(4), 436–440.

- Mohanty, S. D. (2000). Robust test for detecting nonstationarity in data from gravitational wave detectors. *Physical Review D*, 61(122002), 1–12.
- Morettin, P. A. & Toloi, C. M. C. (2004). *Análise de Séries Temporais*. Editora Edgard Blücher Ltda., 1a. edição.
- Norton, J. P. (1986). *An Introduction to Identification*. Academic Press INC., London - UK.
- Oliveira, D. J. (2002). Síntese e análise do sinal de variabilidade da frequência cardíaca baseada em modelos identificados. Master's thesis, Universidade Federal de Minas Gerais. (PPGEE).
- Paarmann, L. D. & Korenberg, M. J. (1992). Estimation of the parameters of an ARMA signal model based on an orthogonal search. *IEEE Trans. on Automatic Control*, 37(3), 347–352.
- Padmakumari, K., Mohandas, K. P., & Thiruvengadam, S. (1999). Long term distribution demand forecasting using neuro fuzzy computations. *Electrical Power and Energy System*, 21, 315–322.
- Paley, R. E. A. (1932). A remarkable series of orthogonal functions. *Proc. London Math Soc.*, 34, 241–279.
- Parkum, J. E. (1992). Recursive forgetting algorithms. *International Journal of Control*, 55(1), 109–128.
- Penny, W. D. & Roberts, S. J. (1999). Dynamic models for nonstationary signal segmentation. *Computers and Biomedical Research*, 32(6), 483–502.
- Pierce, D. A. (1980). A survey of recent developments in seasonal adjustment. *The American Statistician*, 34, 125–134.
- Plackett, R. L. (1950). Some theorems in least squares. *Biometrika*, 37, 149–157.
- Pola, S., Macerata, A., Emdin, M., & Marchesi, C. (1996). Estimation of the power spectral density in nonstationary cardiovascular time series: Assessing the role of the time-frequency representation (TRF). *IEEE Transactions. on Biomedical Engineering*, 43(1), 46–59.
- Potamianos, A. & Maragos, P. (2001). Time-frequency distributions for automatic speech recognition. *IEEE Transactions on Speech and Audio Processing*, 9, 196–200.

- Ralston, J. C., Zoubir, A. M., & Boashash, B. (1996). Identification of a class of time-varying nonlinear systems using basis sequences. *Proc. of IEEE Time-Frequency and Time-scale analysis. Int. Symp. on.*, 161–164.
- Rieke, C., Andrzejak, R. G., Mormann, F., & Lehnertz, K. (2004). Improved statistical test for nonstationarity using recurrence time statistics. *Physical Review E*, 69(4).
- Rieke, C., Mormann, F., Andrzejak, R. G., Kreuz, T., David, P., Elger, C. E., & Lehnertz, K. (2003). Discerning nonstationarity from nonlinearity in seizure-free and preseizure EEG recordings from epilepsy patients. *IEEE Trans. on Biomedical Engineering*, 50(5), 634–639.
- Rieke, C., Sternickel, K., Andrzejak, R. G., David, P., Elger, C. E., & Lehnertz, K. (2002). Measuring nonstationarity by analysing the loss of recurrence in dynamical systems. *Physical Review Letters*, 88(24).
- Salgado, M., Goodwin, G. C., & Richard, H. M. (1988). Modified least square algorithm incorporating exponential resetting and forgetting. *International Journal of Control*, 47(2), 477–491.
- Schreiber, T. (1997). Detecting and analyzing nonstationarity in a time series using nonlinear cross predictions. *Physical Review Letters*, 78(5), 843–846.
- Schreiber, T. (1999). Interdisciplinary application of nonlinear time series methods. *Physics Reports*, 308(1), 1–64.
- Söderström, F. & Stoica, P. (1989). *System Identification*. London-UK: Prentice Hall International.
- Sorenson, H. W. (1970). Least-square estimation: from gauss to kalman. *IEEE Spectrum*, 7, 63–68.
- Stark, J. (1993). Recursive prediction of chaotic time series. *J. Nonlinear Sci.*, 3, 197–223.
- Steenis, H. G. V., Martens, W. L. J., & Tulen, J. H. M. (2003). Quantification of the dynamic behavior over time of narrow-band components present in heart rate variability by means of the instantaneous amplitude and frequency. *Computers in Biology and Medicine*.
- Stolen, D. P. & Hardness, J. (1994). Walsh functions for offline identification of nonlinear plants and their control. *International Journal of Control*, 60(6), 1121–1136.

- Tanguy, N., Morvan, R., Vilbé, P., , & Calvez, L.-C. (2002). Pertinent choice of parameters for discrete Kautz approximation. *IEEE Trans. on Automatic Control*, 47(5), 783–787.
- Tong, H. (1990). *Nonlinear Time Series - A Dynamical System Approach*. Oxford Science Publication.
- Tôrres, L. A. B. (2001). *Controle e Sincronismo de Osciladores Caóticos*. PhD thesis, Universidade Federal de Minas Gerais. (PPGEE).
- Tsatsanis, M. K. & Giannakis, G. B. (1993). Time-varying system identification and model validation using wavelets. *IEEE Trans. on Signal Processing*, 41(12), 3512–3523.
- Vahidi, A., Stefanopoulou, A., & Peng, H. (2004). Recursive least squares with forgetting for online estimation of vehicle mass and road grade: Theory and experiments. To appear in *International Journal of Control*.
- Van Den Hof, P. M. J., Heuberger, P. S. C., & Bokor, J. (1995). System identification with generalized orthonormal basis functions. *Automatica*, 31(12), 1821–1834.
- Virili, F. & Freisleben, B. (2000). Nonstationarity and data preprocessing for neural network predictions of an economic time series. *IEEE, Int. J. Joint Conf. on NN*, 5, 129–134.
- Volterra, V. (1930). *Theory of functions*. Blackie and Sons.
- Walsh, J. L. (1923). A closed set of normal orthogonal functions. *Amer. J. Math.*, 45, 5–24.
- Wei, H.-L. & Billings, S. A. (2002). Identification of time-varying systems using multiresolution wavelet models. *Int. J. of Systems Science*, 33(15), 1217–1228.
- Wiener, N. (1949). *Extrapolation, inetrpolation and smoothing of stationary time series*. MIT Press, Cambridge, MA.
- Witt, A., Kurths, J., & Pikovsky, A. (1998). Testing stationarity in time series. *Physical Review E*, 58(2), 1800–1810.
- Wold, H. (1954). *A study in the analysis of stationary time series*. 2nd edition Estocolmo: Almqvist & Wiksel.
- Young, P. (1984). *Recursive Estimation and Time-Series Analysis*. Springer Verlag. Berlin.

- Yuen, C. K. (1972). Remarks on the ordering of Walsh functions. *IEEE Trans. Comput.*, *21*, 1452.
- Yule, G. U. (1927). On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers. *Philos. Trans. R. Soc., A*(226), 267–298.
- Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting*, *14*, 35–62.
- Zhang, G. P. & Qi, M. (2003). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, *160*, 501–514.
- Zhuang, W. (1998). RLS algorithm with variable forgetting factor for decision feedback equalizer over time-variant fading channels. *Wireless Personal Communications*, *8*(1), 15–29.
- Zou, R. & Chon, K. H. (2004). Robust algorithm for estimation of time-varying transfer functions. *IEEE Trans. on Biomedical Engineering*, *51*(2), 219–228.
- Zou, R., Wan, H., & Chon, K. H. (2003). A robust time-varying identification algorithm using basis function. *Ann. Biomed. Eng.*, *31*, 840–853.

