

Alexandre Barbosa de Lima

Contribuições à Modelagem de Teletráfego Fractal

Texto apresentado à Escola Politécnica
da Universidade de São Paulo para
obtenção do Título de Doutor em Enge-
nharia .

Alexandre Barbosa de Lima

Contribuições à Modelagem de Teletráfego Fractal

Texto apresentado à Escola Politécnica
da Universidade de São Paulo para
obtenção do Título de Doutor em Enge-
nharia .

Área de concentração:
Sistemas Eletrônicos

Orientador:
Prof. Dr. José Roberto de Al-
meida Amazonas

Para Shirlei e nossas filhas Ana Luísa e Maria Eduarda.

Agradecimentos

A Deus, por ter me sustentado ao longo da caminhada.

A minha esposa Shirlei, que sempre me apoiou incondicionalmente em todos os meus projetos profissionais e acadêmicos.

Ao meu orientador e amigo Prof. Dr. Amazonas.

Ao Prof. Dr. Pedro A. Morettin, do IME-USP, pelas discussões (altamente frutíferas) sobre modelagem de séries temporais com memória longa e pela sua disponibilidade para me receber em seu escritório, não obstante a sua agenda lotada.

A todos os amigos que me apoiaram e ajudaram na realização deste trabalho. Em especial a todos os colegas do Laboratório de Comunicações e Sinais (LCS) da Escola Politécnica da USP.

Resumo

Estudos empíricos [1],[2] demonstraram que o tráfego das redes *Internet Protocol* (IP) possui propriedades *fractais* tais como *impulsividade*, *auto-similaridade* e *dependência de longa duração* em diversas escalas de agregação temporal, na faixa de milissegundos a minutos. Essas características têm motivado o desenvolvimento de novos modelos fractais de teletráfego e de novos algoritmos de controle de tráfego em redes convergentes. Este trabalho propõe um novo modelo de tráfego no espaço de estados baseado numa aproximação finito-dimensional do processo *AutoRegressive Fractionally Integrated Moving Average* (ARFIMA). A modelagem por meio de processos auto-regressivos (AR) também é investigada.

A análise estatística de séries simuladas e de séries reais de tráfego mostra que a aplicação de modelos AR de ordem alta em esquemas de previsão de teletráfego é fortemente prejudicada pelo problema da identificação da ordem do modelo. Também demonstra-se que a modelagem da memória longa pode ser obtida às custas do posicionamento de um ou mais pólos nas proximidades do círculo de raio unitário. Portanto, a implementação do modelo AR ajustado pode ser instável devido a efeitos de quantização dos coeficientes do filtro digital.

O modelo de memória longa proposto oferece as seguintes vantagens: a) possibilidade de implementação prática, pois não requer memória infinita, b) modelagem (explícita) da região das baixas frequências do espectro e c) viabilização da utilização do filtro de Kalman.

O estudo de caso apresentado demonstra que é possível aplicar o modelo de memória longa proposto em trechos estacionários de sinais de teletráfego fractal. Os resultados obtidos mostram que a dinâmica do parâmetro de Hurst de sinais de teletráfego pode ser bastante lenta na prática. Sendo assim, o novo modelo proposto é adequado para esquemas de previsão de tráfego, tais como Controle de Admissão de Conexões (CAC) e alocação dinâmica de banda, dado que o parâmetro de Hurst pode ser estimado em tempo real por meio da aplicação da transformada *wavelet* discreta (*Discrete Wavelet Transform* (DWT)).

Abstract

Empirical studies [1],[2] demonstrated that heterogeneous IP traffic has fractal properties such as impulsiveness, self-similarity, and long-range dependence over several time scales, from milliseconds to minutes. These features have motivated the development of new traffic models and traffic control algorithms. This work presents a new state-space model for teletraffic which is based on a finite-dimensional representation of the ARFIMA random process. The modeling via *AutoRegressive* (AR) processes is also investigated.

The statistical analysis of simulated time series and real traffic traces show that the application of high-order AR models in schemes of teletraffic prediction can be highly impaired by the model identification problem. It is also demonstrated that the modeling of the long memory can be obtained at the cost of positioning one or more poles near the unit circle. Therefore, the implementation of the adjusted AR model can be unstable due to the quantization of the digital filter coefficients.

The proposed long memory model has the following advantages: a) possibility of practical implementation, inasmuch it does not require infinite memory, b) explicit modeling of the low frequency region of the power spectrum, and c) forecasts can be performed via the Kalman predictor.

The presented case study suggests one can apply the proposed model in periods where stationarity can be safely assumed. The results indicate that the dynamics of the Hurst parameter can be very slow in practice. Hence, the new proposed model is suitable for teletraffic prediction schemes, such as CAC and dynamic bandwidth allocation, given that the Hurst parameter can be estimated on-line via DWT.

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	1
1.1	Motivação	1
1.2	Objetivos e Contribuições	3
1.3	Estrutura da Tese	4
I	Fundamentação Teórica	5
2	Conceitos Básicos em Séries Temporais	6
2.1	Notação de Operadores	6
2.2	Processos Estocásticos	7
2.3	Modelagem de Séries Temporais	12
2.4	Modelos Auto-Regressivos	16
2.4.1	Função de Autocorrelação	16
2.4.2	Identificação	17
2.4.3	Estimação de Modelos AR	20
2.5	Modelagem de Séries Não-Estacionárias	20
2.5.1	Séries Estacionárias com Relação a Tendências	20
2.5.2	Modelo ARIMA	21
2.5.3	Passeio Aleatório	22
3	A Natureza Fractal do Teletráfego	24

3.1	Fractais	24
3.2	O Expoente de Hurst	30
3.3	LRD e Auto-Similaridade	32
3.3.1	Dependência de Longa Duração	32
3.3.2	Auto-Similaridade	35
3.4	Impulsividade	37
3.4.1	Distribuições Estáveis	37
3.4.2	Medidas de Dependência para Processos Impulsivos	43
3.5	Por que o tráfego das redes de dados é fractal?	45
4	Modelagem de Teletráfego com Memória Longa	46
4.1	Introdução	46
4.2	Modelagem Não-Paramétrica com FGN e MWM	47
4.2.1	Processo FGN	47
4.2.2	Transformada <i>Wavelet</i>	49
4.2.3	Modelo MWM	69
4.3	Modelagem Paramétrica	71
4.3.1	Modelo ARFIMA	71
4.3.2	Previsão de Modelos ARFIMA	74
4.4	Testes Estatísticos de Memória Longa	77
4.4.1	Estatística R/S	77
4.4.2	Teste GPH	79
4.5	Alguns Métodos para Estimação de H e d	79
4.5.1	Abordagens Heurísticas	79
4.5.2	Método de Whittle	81
4.5.3	Estimador Aproximado de MV de Haslett e Raftery	81
4.5.4	Estimador <i>Wavelet</i> de Abry e Veitch	82
4.6	Biespectro e Testes de Linearidade	84

4.7	Teste de Estacionariedade KPSS	89
-----	--	----

II Resultados 91

5 Modelagem de Teletráfego no Espaço de Estados 92

5.1	Introdução	92
5.2	Modelagem no Espaço de Estados	94
5.2.1	Modelo TARFIMA	94
5.2.2	Previsão Multipasso com o Filtro de Kalman	96
5.3	O Poder de Previsão do Modelo TARFIMA	97
5.4	Análise Exploratória de Séries Simuladas e da Série do Rio Nilo .	100
5.4.1	Série ARFIMA(0; 0, 4; 0)	101
5.4.2	Série MWM com $H = 0,9$	106
5.4.3	Série ARFIMA(0; 0, 3; 0)	111
5.4.4	Série MWM com $H = 0,8$	116
5.4.5	Série do Rio Nilo	120
5.5	Estudo Empírico da Previsão com o Modelo TARFIMA	125
5.5.1	Séries ARFIMA(0, d , 0)	125
5.5.2	Séries MWM	135
5.5.3	Série do Rio Nilo entre os anos de 1007 e 1206	140
5.6	Conclusão	141

6 Modelagem de Tráfego de Redes 146

6.1	Modelagem do <i>Trace</i> UNC02	147
6.1.1	Análise Exploratória	147
6.1.2	Análise Local da Memória Longa do <i>Trace</i> UNC02	160
6.1.3	Previsão Empírica com o Modelo TARFIMA	168
6.2	Conclusão	171

7 Conclusões e Trabalhos Futuros	175
7.1 Sumário	175
7.2 Trabalhos Futuros	176
Referências	178
Apêndice A – Códigos MATLAB	189
Apêndice B – Código em S para S-PLUS	209

Lista de Figuras

2.1	Ciclo iterativo Box-Jenkins.	13
3.3	Foto que ilustra a auto-similaridade de uma couve-flor [3]. Note que as partes menores se parecem com a couve-flor original. . . .	27
3.4	Ilustração das cinco primeiras iterações para obtenção do conjunto de Cantor. Repare-se que o formato de uma determinada parte do objeto, quando magnificada, se assemelha ao formato do todo (auto-similaridade).	28
3.5	Visualização do tráfego <i>Fast Ethernet</i> (100 Mbps) coletado num servidor WWW/Email/FTP da Universidade de Drexel [4] em quatro diferentes níveis de agregação: 10 ms, 100 ms, 1 s e 10 s (de cima para baixo). Este <i>trace</i> tem um tamanho total de 3 h.	28
3.6	Comparação de tráfego Ethernet real e sintetizado [1]. À esquerda tem-se o tráfego real, ao centro tráfego simulado utilizando-se o modelo de Poisson e à direita tráfego gerado por meio de um modelo auto-similar.	29
3.7	Tráfego multifractal simulado por meio do modelo MWM. Note-se que a impulsividade da série varia no tempo (heterocedasticidade).	30
3.8	FAC de uma série LRD com $H = 0,9$ e $N = 4096$ amostras. A linha contínua mostra a função de autocorrelação do melhor modelo auto-regressivo (AR) que pôde ser ajustado pela função <code>ar</code> do <i>software S+FinMetrics</i> [5] à série segundo o critério <i>Akaike Information Criterion</i> (AIC) [6]. No caso, obteve-se um modelo AR(15). Note que o decaimento da autocorrelação do modelo AR(15) ajustado decai mais rapidamente (exponencialmente) para zero do que o da série LRD.	31
3.9	DEPs para modelos AR(4) e FD(0,4) de mesma potência.	32

3.10	Funções densidade de probabilidade de Pareto I para vários valores de $k = \alpha$ com $x_m = 1$. Observe que a função decai lentamente para zero para valores grandes de x . A função tende para o impulso $\delta(x - x_m)$ quando $k \rightarrow \infty$	38
3.11	Funções distribuição de probabilidade de Pareto I para vários valores de $k = \alpha$ com $x_m = 1$	39
3.12	Simulação de uma realização com distribuição $S_{1,2}(1, 1, 0)$ (256 amostras).	40
3.13	Funções densidade de probabilidade estáveis simétricas centradas em zero [7]. Os parâmetros β e c correspondem aos parâmetros η e σ , respectivamente, de (3.23).	40
3.14	Funções densidade de probabilidade estáveis assimétricas [7].	41
3.15	Funções distribuição de probabilidade simétricas [7].	41
3.16	Funções distribuição de probabilidade assimétricas [7].	42
4.1	Realizações de processos FBM para vários valores do parâmetro de Hurst. As simulações foram feitas com o código MATLAB de geração e identificação de processos FBM desenvolvido por Coeurjolly [8],[9].	48
4.2	Quatro exemplos de funções <i>wavelet</i> . Esta figura foi gerada pela função <code>toon0111.m</code> do <i>toolbox WaveLab850</i> para MATLAB [10].	51
4.3	<i>Wavelet</i> de Meyer. Esta figura foi gerada pela função <code>WaveLab850 toon0114.m</code>	51
4.5	A imagem da parte inferior da figura é a CWT $W_\psi(s, \tau)$ do sinal da parte superior, calculada com uma <i>wavelet</i> que é a primeira derivada da <i>Probability Density Function</i> (PDF) Gaussiana. Os parâmetros $\tau = t$ e s variam ao longo dos eixos horizontal e vertical, respectivamente. Grandes valores (positivos) de coeficientes <i>wavelet</i> são indicados pela cor preta. Observe que as singularidades isoladas na parte esquerda da figura produzem grandes coeficientes em seus respectivos cones de influência, que convergem para as localizações das singularidades. Esta figura foi criada com a função <code>WaveLab WTBrowser.m</code>	55
4.6	Taxonomia da transformada <i>wavelet</i>	57

4.7	Amostragem crítica do plano tempo-escala por meio da discretização dos parâmetros da CWT ($s = 2^j$ e $\tau = 2^j k$). A CWT é definida em todos os pontos do plano (s, τ) e corresponde a uma representação redundante da informação presente no sinal. Note que o número de coeficientes dobra quando se vai de uma escala $s_1 = 2^{j+1}$ para uma escala mais rápida (ou mais refinada) $s_2 = 2^j$.	58
4.8	Uma ilustração da DWT de 3 níveis do sinal de tempo discreto $x(k) = \sin(3k) + \sin(0,3k) + \sin(0,03k)$. O gráfico concatena as seqüências dos coeficientes de escala $\{u_3(k)\}_{128}$ e dos coeficientes <i>wavelet</i> $\{w_3(k)\}_{128}$, $\{w_2(k)\}_{256}$ e $\{w_1(k)\}_{512}$ da esquerda para a direita, ou seja, os primeiros 128 pontos correspondem à seqüência $\{u_3(k)\}_{128}$; seguem-se os 128 pontos da seqüência $\{w_3(k)\}_{128}$, os 256 pontos da seqüência $\{w_2(k)\}_{256}$ e os 512 pontos da seqüência $\{w_1(k)\}_{512}$.	61
4.9	DEP do sinal $x(k) = \sin(3k) + \sin(0,3k) + \sin(0,03k)$.	61
4.10	Síntese do sinal $x(k) = \sin(3k) + \sin(0,3k) + \sin(0,03k)$ em termos da soma $\mathcal{S}_3(t) + \mathcal{D}_1(t) + \mathcal{D}_2(t) + \mathcal{D}_3(t)$ (aproximação na escala $j = 3$ e detalhes nas escalas 1, 2 e 3). Na Fig., $a_3 = \mathcal{S}_3$ e $d_j = \mathcal{D}_j$, $j = 1, 2, 3$.	62
4.11	Resposta em frequência de filtros QMF (gráfico da parte superior) <i>vs</i> resposta em frequência de filtros do tipo <i>brickwall</i> (fisicamente não-realizáveis).	64
4.12	Os gráficos da parte inferior mostram as <i>wavelets</i> de Daubechies com $N = 2, 3, 4$ momentos nulos (<i>vanishing moments</i>), da esquerda para a direita, respectivamente. As funções de escala correspondentes estão na parte superior. Esta figura foi criada com a função WaveLab WTBrowser.m.	66
4.15	Diagrama de blocos que mostra que a DWT funciona de modo similar a um esquema de codificação de sub-bandas. O espectro $U_0(f)$ do sinal $u_0(n)$ da Fig. 4.14 é subdividido em três bandas de frequência (que cobrem duas oitavas): $0 \leq f < 1/8$, $1/8 \leq f < 1/4$ e $1/4 \leq f \leq 1$.	69
4.16	DEPs para modelos AR(4) e FD(0,4) de mesma potência.	73
5.3	Série do rio Nilo entre os anos de 1007 e 1206.	104
5.5	Espectro <i>wavelet</i> do sinal ARFIMA(0; 0, 4; 0).	106

5.8	Diagrama de pólos e zeros do modelo AR(15).	109
5.10	Espectro <i>wavelet</i> do sinal MWM($H = 0, 9$).	111
5.13	Diagrama de pólos e zeros do modelo AR(28).	114
5.15	Espectro <i>wavelet</i> do sinal ARFIMA(0; 0, 3; 0).	116
5.18	Diagrama de pólos e zeros do modelo AR(16).	119
5.20	Espectro <i>wavelet</i> do sinal MWM($H = 0, 8$).	122
5.23	Diagrama de pólos e zeros do modelo AR(29).	125
5.25	Espectro <i>wavelet</i> da série do rio Nilo entre os anos de 1007 e 1206.	127
5.28	Diagrama de pólos e zeros do modelo AR(8).	130
6.1	Tráfego coletado na rede local do LCS: série de contagem de <i>bytes</i> por <i>bin</i> de 100 milissegundos.	149
6.4	Ajuste de uma distribuição estável com parâmetros $\{\alpha = 1, 92, \sigma = 495, \eta = 0, 86, \mu = 2, 15 \times 10^4\}$, à série <code>UNC02Pktsbin1s</code> (gráfico do canto superior esquerdo), QQ-plot da distribuição estável <i>vs</i> os dados empíricos (canto superior direito), ZZ-plot da distribuição estável <i>vs</i> os dados empíricos (canto inferior esquerdo) e PP-plot da distribuição estável ajustada <i>vs</i> os dados empíricos (canto inferior direito). Os gráficos ZZ-plot e PP-plot sugerem que o ajuste da distribuição $S_{1,92}(495; 0, 86; 2, 15 \times 10^4)$ aos dados empíricos é bastante satisfatório.	154
6.7	Ajuste de uma PDF estável com parâmetros $\{\alpha = 1, 79, \sigma = 4, 09 \times 10^5, \eta = 0, 99, \mu = 6, 35 \times 10^6\}$, à série <code>UNC02bin1s</code> (gráfico do canto superior esquerdo), QQ-plot da distribuição estável estimada <i>vs</i> os dados empíricos (canto superior direito), ZZ-plot da distribuição estável ajustada <i>vs</i> os dados empíricos (canto inferior esquerdo) e PP-plot da distribuição estável estimada <i>vs</i> os dados empíricos (canto inferior direito). Os gráficos ZZ-plot e PP-plot sugerem que os dados da série são bem modelados pela distribuição $S_{1,79}(4, 09 \times 10^5; 0, 99; 6, 35 \times 10^6)$	157

Lista de Tabelas

4.1	Valores assintóticos do parâmetro de forma p dos multiplicadores β em função de α (or H)[11].	71
5.1	Valores de P_h^2 para modelos AR(1) e TARFIMA(L), $L = \{1, 10\}$. Os valores das correlações de $lag-1$ do AR(1) foram escolhidos de tal forma que $\rho_1 = 1/9$ corresponde à correlação de $lag-1$ do ARFIMA(0; 0, 1; 0) e $\rho_1 = 2/3$ corresponde à correlação de $lag-1$ do ARFIMA(0; 0, 4; 0).	98
5.2	Valores de P_h^2 para modelos TARFIMA(L), $L = \{100, 150\}$ e ARFIMA(0, d , 0) (coluna em que $L = \infty$).	99
5.3	Estimativas do parâmetro d para a realização ARFIMA(0; 0, 4; 0).	101
5.4	Série ARFIMA(0; 0, 4; 0): testes de estacionariedade, memória longa, normalidade e linearidade, e número de raízes unitárias (coluna \hat{m}).	104
5.5	Estimativas do parâmetro H para a realização MWM($H = 0, 9$).	106
5.6	Série MWM($H = 0, 9$): resultados dos testes estatísticos e número estimado de raízes unitárias.	109
5.7	Estimativas do parâmetro d para a realização ARFIMA(0; 0, 3; 0).	111
5.8	Série ARFIMA(0; 0, 3; 0): resultados dos testes estatísticos e número estimado de raízes unitárias.	114
5.9	Estimativas do parâmetro H para a realização MWM($H = 0, 8$).	116
5.10	Série MWM($H = 0, 8$): resultados dos testes estatísticos e número estimado de raízes unitárias.	119
5.11	Estimativas do parâmetro d para a série do rio Nilo.	120
5.12	Série do rio Nilo: resultados dos testes estatísticos e número estimado de raízes unitárias.	122

5.13	Série ARFIMA(0; 0, 4; 0): $PCEMSE_{x,y}(t = 3996, K)$ dos modelos estimados AR(15), TARFIMA(10), TARFIMA(50) e TARFIMA(100) (colunas em que $L = 10, 50$, e 100, respectivamente) relativos ao modelo AR(2). Os modelos TARFIMA(L) utilizaram $\hat{d} = 0,3684$.	130
5.14	Série ARFIMA(0; 0, 3; 0): $PCEMSE_{x,y}(t = 3996, K)$ dos modelos estimados AR(16), TARFIMA(10), TARFIMA(50) e TARFIMA(100) relativos ao modelo AR(2). Os modelos TARFIMA(L) utilizaram $\hat{d} = 0,2895$.	135
5.15	Série MWM($H = 0,9$): $PCEMSE_{x,y}(t = 3996, K)$ dos modelos estimados AR(28), TARFIMA(10), TARFIMA(50) e TARFIMA(100) relativos ao AR(2). Os modelos TARFIMA(L) utilizaram $\hat{d} = 0,3572$.	140
5.16	Série MWM($H = 0,8$): $PCEMSE_{x,y}(t = 3996, K)$ dos modelos estimados AR(29), TARFIMA(10), TARFIMA(50) e TARFIMA(100) relativos ao AR(2). Os modelos TARFIMA(L) utilizaram $\hat{d} = 0,3164$.	140
5.17	Série do rio Nilo: $PCEMSE_{x,y}(t = 1106, K)$ dos modelos estimados AR(8), TARFIMA(10), TARFIMA(50) e TARFIMA(100) relativos ao AR(2). Os modelos TARFIMA(L) utilizaram $\hat{d} = 0,4125$.	141
6.1	Série UNC02bin1s : resultados de testes estatísticos (estacionariedade, memória longa, normalidade e linearidade) e quantificação do número de raízes unitárias (coluna \hat{m}). Os testes de Hinich, que utilizam a <i>Fast Fourier Transform</i> (FFT) (<i>Fast Fourier Transform</i>), foram aplicados ao primeiro bloco de 4096 amostras de UNC02bin1s (que contém 7340 amostras, um número que não é uma potência de 2).	156
6.2	Estimativas do parâmetro d para a série UNC02bin1s .	158
6.3	Série UNC02bin1s ($\hat{d} = 0,3717$): $PCEMSE_{x,y}(t = 2700, K)$ dos modelos estimados AR(18), TARFIMA(10), TARFIMA(50) e TARFIMA(100) relativos ao modelo AR(2).	171

LISTA DE ABREVIATURAS

ACF *AutoCorrelation Function*

AIC *Akaike Information Criterion*

ANOVA *Analysis of Variance*

AR *AutoRegressive*

ARIMA *AutoRegressive Integrated Moving Average*

ARFIMA *AutoRegressive Fractionally Integrated Moving Average*

ARMA *AutoRegressive Moving Average*

ATM *Asynchronous Transfer Mode*

BIC *Bayesian Information Criteria*

bps bits por segundo

CAC Controle de Admissão de Conexões

CDF *Cumulative Distribution Function*

CEMSE *Cumulative Empirical Mean Squared Error*

CWT *Continuous Wavelet Transform*

DEP Densidade Espectral de Potência

DFBM *Discrete-time Fractional Brownian Motion*

DiffServ *Differentiated Services*

DWT *Discrete Wavelet Transform*

EM *Expectation-Maximization*

EMQ Estimador de Mínimos Quadrados

EMSE *Empirical Mean Squared Error*

EMV Estimador de Máxima Verossimilhança

EQM Erro Quadrático Médio

EQMM Erro Quadrático Médio Mínimo

FAC Função de Autocorrelação

FACP Função de Autocorrelação Parcial

FACPA Função de Autocorrelação Parcial Amostral

FARIMA *Fractional Autoregressive Integrated Moving Average*

FBM *Fractional Brownian Motion*

FD *Fractionally Differenced*

FFT *Fast Fourier Transform*

FIR *Finite Impulse Response*

FGN *Fractional Gaussian Noise*

FTP *File Transfer Protocol*

ICWT *Inverse Continuous Wavelet Transform*

IDWT *Inverse Discrete Wavelet Transform*

IETF *Internet Engineering Task Force*

IntServ *Integrated Services*

IID Independente e Identicamente Distribuído

IP *Internet Protocol*

IIR *Infinite Impulse Response*

ISP *Internet Service Provider*

ITU-T *International Telecommunications Union - Telecommunications
Standardization Sector*

HTTP *Hyper Text Transfer Protocol*

LAN *Local Area Network*

LMS *Least Mean Squares*

LRD *Long-Range Dependence*

MA *Moving Average*

MBAC *Measurement-Based Admission Control*

MMSE *Minimum Mean Squared Error*

MODWT *Maximal Overlap Discrete Wavelet Transform*

MPEG *Moving Picture Experts Group*

MPLS *Multiprotocol Label Switching*

MQ *Mínimos Quadrados*

MRA *Multiresolution Analysis*

MV *Máxima Verossimilhança*

MTU *Maximum Transfer Unit*

MWM *Multifractal Wavelet Model*

OC *Optical Carrier*

PC *Personal Computer*

PCEMSE *Proportionate Change in the Cumulative Empirical Mean Squared Error*

PDF *Probability Density Function*

PDT *Packet Delay Transfer*

PDV *Packet Delay Variance*

PLR *Packet Loss Rate*

QMF *Quadrature Mirror Filters*

QoS *Quality of Service*

QQ-plot *Quantile-Quantile plot*

RB *Ruído Branco*

RBI *Ruído Branco Independente*

RDSI-FL *Rede Digital de Serviços Integrados de Faixa Larga*

RLS *Recursive Least Squares*

R/S *Range Over Standard Deviation*

SACF *Sample AutoCorrelation Function*

SEMIFAR *Semiparametric Fractional Autoregressive*

SONET *Synchronous Optical network*

SRD *Short-Range Dependence*

STM *Synchronous Transport Module*

TDF *Transformada Discreta de Fourier*

TCP *Transport Control Protocol*

TFTD *Transformada de Fourier de Tempo Discreto*

THRU *Throughput*

UPC *Usage Parameter Control*

VBR *Variable Bit Rate*

WAN *Wide Area Network*

WFT *Windowed Fourier Transform*

WOSA *Welch's Overlapped Segment Averaging*

1 Introdução

Este Capítulo apresenta a motivação, os objetivos e as principais contribuições deste trabalho.

1.1 Motivação

Atualmente, as operadoras de telecomunicações têm como meta a integração dos serviços de voz, vídeo, imagem e dados sobre uma única rede convergente chaveada por pacotes. Esta rede, que também é conhecida como rede de serviços integrados, multisserviço, ou de próxima geração, deve possuir características de desempenho aceitáveis para cada tipo de serviço e as seguintes vantagens: redução de custos operacionais, flexibilidade para suportar os serviços existentes e futuros serviços ainda não previstos, alocação dinâmica de banda, transporte integrado de todos os tipos de informação e utilização eficiente dos recursos da rede através da multiplexação estatística. A crescente demanda por aplicações multimídia também é outro fator motivador da implementação dessa rede.

Em 1988, a *International Telecommunications Union - Telecommunications Standardization Sector* (ITU-T) padronizou o Modo de Transferência Assíncrono (*Asynchronous Transfer Mode* (ATM)) como tecnologia de transporte a ser adotada na Rede Digital de Serviços Integrados de Faixa Larga (RDSI-FL). Entretanto, a utilização da tecnologia ATM foi bastante prejudicada pela sua complexidade, dificuldades de padronização e de integração ao IP e, principalmente, porque poucas aplicações suportam o ATM de forma nativa. O ATM foi “derrotado” pela simplicidade e enorme sucesso da pilha de protocolos *Transport Control Protocol* (TCP)/IP. O IP tornou-se o padrão de fato e é a “cola” que une toda a Internet.

O núcleo das redes convergentes consiste numa única infraestrutura IP, com suporte a Qualidade de Serviço - *Quality of Service* (QoS)¹ - redes privativas virtu-

¹Os parâmetros de QoS são os seguintes: *Packet Delay Transfer* (PDT), *Packet Delay Variance* (PDV), *Throughput* (THRU) e *Packet Loss Rate* (PLR) [12].

ais e protocolos IP versões 4 e 6 (IPv4/IPv6). A transformação da Internet numa rede com suporte a QoS é um dos grandes desafios a serem vencidos e é por isso que o *Internet Engineering Task Force* (IETF) tem proposto várias tecnologias e padrões para a implementação de QoS na Internet, tais como: *Multiprotocol Label Switching* (MPLS) [13], [14], [15], roteamento baseado em restrições (*constrained-based routing*) [13], engenharia de tráfego, Serviços Integrados (*Integrated Services* (IntServ)) [16] e Serviços Diferenciados (*Differentiated Services* (DiffServ)) [17], [18], [19], [20].

A implementação de mecanismos que regulem e monitorem o teletráfego é essencial para a operação eficiente das redes convergentes. Sem controle de tráfego, a demanda irrestrita pelos recursos compartilhados (*buffers*, banda e processadores) pode degradar seriamente o desempenho da rede. O controle de tráfego é necessário para proteger a QoS percebida pelos usuários e para assegurar a eficiente utilização dos recursos da rede. O seguinte conjunto de funções de controle de tráfego deve ser implementado nas redes convergentes [21]: a) CAC, b) policiamento do usuário - *Usage Parameter Control* (UPC), c) controle de prioridade e d) controle de congestionamento. As funções a), c) e d) envolvem a funcionalidade de previsão *on-line* de tráfego.

Medições [1], [2] demonstraram que o tráfego agregado (ou heterogêneo) das redes de dados² possui propriedades estatísticas fundamentalmente diferentes do tráfego existente nas redes telefônicas tradicionais, que usam a tecnologia de comutação de circuitos. O matemático e engenheiro dinamarquês A. K. Erlang mostrou em 1909 [22] que o número de chamadas telefônicas geradas durante um determinado intervalo de tempo pode ser modelado pelo processo de Poisson [23]³. Em contraste, *traces* (séries) de tráfego de dados possuem propriedades fractais tais como dependência de longa duração⁴ - *Long-Range Dependence* (LRD) ou memória longa - e impulsividade (grande variabilidade) em diversas escalas de agregação temporal, que não são capturadas pelo processo de Poisson. O tráfego em algumas redes locais - *Local Area Network* (LAN) - e de grande área - *Wide Area Network* (WAN) é extremamente impulsivo porque tem uma distribuição marginal com cauda pesada (não-Gaussiana)⁵ [25], [26], [27], [28], [29], [30].

²O tráfego de dados é a série temporal de contagem de *bytes* por unidade de tempo (*bytes/bin*) em um enlace da rede.

³O modelo de Poisson é uma cadeia de Markov [23, p. 502].

⁴No contexto desta pesquisa, os termos memória longa e auto-similaridade são sinônimos. A rigor, auto-similaridade e dependência de longa duração são conceitos distintos. Entretanto, conforme será mostrado mais adiante, processos estocásticos auto-similares de segunda ordem são LRD e vice-versa. É por esta razão que a literatura não faz distinção entre os termos auto-similaridade e LRD.

⁵Um comportamento parecido foi observado, no início da década de 1960, por Mandelbrot em

Sabe-se que a probabilidade de transbordamento numa fila (*buffer overflow probability*), com um único servidor que trabalha a uma taxa de serviço constante, submetida a tráfego Markoviano é uma função exponencial do tamanho do *buffer* [31]. Portanto, um aumento do *buffer* diminui sensivelmente a PLR (taxa de perda de pacotes) quando o tráfego é Markoviano. Por outro lado, vários estudos [1], [32], [33], mostraram que a memória longa e a impulsividade do tráfego de dados degradam o desempenho da rede (a taxa de perda de pacotes aumenta), porque a *buffer overflow probability* em sistemas submetidos a tráfego fractal é uma função potência do tamanho do *buffer*, o que implica um decaimento hiperbólico, muito mais lento do que o do caso Markoviano⁶ [30], [34]. Portanto, tráfego auto-similar provoca *buffer overflows* muito mais freqüentemente do que tráfego Markoviano. Este fenômeno foi constatado nas redes ATM do início da década de 1990, em que os comutadores foram implementados com pequenos *buffers* (de 10 a 100 células de 53 *bytes*) [35].

Modelos de teletráfego fractal constituem o núcleo dos mecanismos de controle de tráfego em redes convergentes, tais como algoritmos de CAC baseados em medições [36], e dos geradores de tráfego agregado, que são usados para o dimensionamento de redes de alta velocidade e para a homologação de novos algoritmos de controle de tráfego via simulações e experimentos em redes de teste (*testbed networks*).

1.2 Objetivos e Contribuições

O objetivo principal deste trabalho é propor um modelo de teletráfego que seja adequado para esquemas de previsão *on-line*. As seguintes questões foram examinadas no decorrer desta pesquisa:

- Como fazer a previsão de sinais com dependência de longa duração?
- Como aplicar a teoria de previsão de séries temporais ao tráfego real das redes de telecomunicações?
- Teletráfego é um sinal linear ou não-linear? É Gaussiano ou não-Gaussiano?
- Como deve ser feita a estimação dos parâmetros do modelo?
- Como identificar o modelo adequado?

séries financeiras: elas também são não-Gaussianas, seguindo distribuições do tipo Lévy-*stable* com variância infinita [24].

⁶A taxa de decaimento é dominada pelo parâmetro H ou pelo índice de cauda (*tail index*) da distribuição marginal.

- Se a previsão puder ser feita com a teoria de estimação linear, então quais seriam as técnicas disponíveis?

Segue-se uma lista com as contribuições originais desta tese:

- Desenvolvimento de um novo modelo de teletráfego fractal no espaço de estados.
- Proposta de uma metodologia de análise estatística de teletráfego LRD.
- Demonstração, por meio de simulações, que a modelagem de sinais com memória longa por meio de processos AR de ordem alta pode ser inviável na prática, dado que pode ser obtida às custas do posicionamento de um ou mais pólos do modelo nas proximidades do círculo de raio unitário.

1.3 Estrutura da Tese

O Capítulo 2 apresenta alguns conceitos fundamentais da teoria de séries temporais e introduz alguns modelos clássicos, tais como o processo AR. O Capítulo 3 introduz a noção de fractal e formula os conceitos de LRD e auto-similaridade de segunda ordem. Também são abordados os conceitos de variável aleatória com distribuição de cauda pesada, distribuições estáveis e impulsividade. O Capítulo 4 apresenta alguns modelos de memória longa importantes da literatura. Os Capítulos 2, 3 e 4 compõem a parte de fundamentação teórica deste trabalho. Os resultados obtidos são apresentados pelos Capítulos 5 e 6. O Capítulo 5 propõe um novo modelo de tráfego. O Capítulo 6 apresenta um estudo de caso baseado na análise de um *trace* de tráfego Internet. Finalmente, o Capítulo 7 apresenta as conclusões e novas idéias para trabalhos futuros.

Parte I

Fundamentação Teórica

2 Conceitos Básicos em Séries Temporais

Este Capítulo apresenta alguns conceitos fundamentais da teoria de séries temporais e introduz os modelos AR, *Moving Average* (MA), *AutoRegressive Moving Average* (ARMA) e *AutoRegressive Integrated Moving Average* (ARIMA) [37]. É dada uma ênfase especial ao modelo AR, que é essencial para o entendimento do novo modelo de tráfego proposto por este trabalho (vide Capítulo 5). Recomenda-se ao leitor que estiver interessado em exposições detalhadas do assunto a leitura das referências [38], [39], [40], [41], [42], [43] e [44].

2.1 Notação de Operadores

Esta tese adota a seguinte notação de operadores:

(a) operador atrasador de 1 amostra, denotado por B :

$$B\mathbf{x}_t = \mathbf{x}_{t-1}. \quad (2.1)$$

(b) operador atrasador de m amostras, denotado por B^m , $m \in \mathbb{Z}$:

$$B^m\mathbf{x}_t = \mathbf{x}_{t-m}. \quad (2.2)$$

A resposta impulsiva¹ h_t do sistema atrasador de m amostras é [45]:

$$h_t = \delta_{t-m}, \quad (2.3)$$

e a sua função de transferência² (ou de sistema) é:

$$H(z) = z^{-m}. \quad (2.4)$$

¹Resposta ao pulso unitário δ_t .

²Definida como a transformada z da resposta impulsiva: $H(z) = \sum_{t=-\infty}^{\infty} h_t z^{-t}$.

(c) operador diferença³ ou filtro diferenciador, denotado por Δ :

$$\Delta \mathbf{x}_t = \mathbf{x}_t - \mathbf{x}_{t-1} = (1 - B)\mathbf{x}_t, \quad (2.5)$$

O filtro diferenciador tem a resposta impulsiva

$$h_t = \delta_t - \delta_{t-1}, \quad (2.6)$$

e a função de transferência

$$H(z) = (1 - z^{-1}). \quad (2.7)$$

(d) operador soma ou filtro integrador⁴, denotado por S :

$$\begin{aligned} S\mathbf{x}_t &= \sum_{i=0}^{\infty} \mathbf{x}_{t-i} = \mathbf{x}_t + \mathbf{x}_{t-1} + \mathbf{x}_{t-2} + \dots = \\ &= (1 + B + B^2 + \dots)\mathbf{x}_t = (1 - B)^{-1}\mathbf{x}_t = \Delta^{-1}\mathbf{x}_t. \end{aligned} \quad (2.8)$$

A função de transferência do filtro integrador corresponde à inversa da função de sistema definida por (2.7), ou seja

$$H(z) = (1 - z^{-1})^{-1}. \quad (2.9)$$

2.2 Processos Estocásticos

A análise de séries temporais supõe que o mecanismo gerador dos dados seja um processo estocástico, o qual é definido a seguir.

Definição 2.1 (Processo Estocástico). Seja T um conjunto arbitrário. Um processo estocástico⁵ é uma família $\{\mathbf{x}_t, t \in T\}$, tal que, para cada $t \in T$, \mathbf{x}_t é uma variável aleatória⁶ [48]. ■

³As potenciações dos operadores B e Δ são definidas como $B^j(\mathbf{x}_t) = \mathbf{x}_{t-j}$ e $\Delta^j(\mathbf{x}_t) = \Delta(\Delta^{j-1}(\mathbf{x}_t))$, $j \geq 1$, com $\Delta^0(\mathbf{x}_t) = \mathbf{x}_t$ [40, pág.19]. Polinômios em B e Δ são manipulados da mesma maneira que polinômios de variáveis reais. Por exemplo,

$$\Delta^2 \mathbf{x}_t = \Delta(\Delta \mathbf{x}_t) = (1 - B)(1 - B)\mathbf{x}_t = (1 - 2B + B^2)\mathbf{x}_t = \mathbf{x}_t - 2\mathbf{x}_{t-1} + \mathbf{x}_{t-2}.$$

⁴Também conhecido como sistema acumulador.

⁵Vários autores da área de séries temporais, tais como Tsay[39] e Zivot e Wang [43], utilizam o termo “série temporal” para o que neste trabalho se define como processo estocástico. Nesta tese, o termo “série temporal” é sinônimo de “realização” de processo aleatório, conforme será explicado mais adiante.

⁶Considere um espaço de probabilidade de referência $(\Omega, \mathfrak{F}, \mathbb{P})$, em que Ω denota o espaço amostral de um experimento aleatório \mathcal{H} , \mathfrak{F} é uma álgebra de Borel [46, pág. 23] de eventos definidos em Ω e $\mathbb{P} : \mathfrak{F} \rightarrow \mathbb{R}$, é uma medida de probabilidade [47, pág. 11]. Os elementos ζ de Ω são os resultados aleatórios de \mathcal{H} . A função $\mathbf{x}(\cdot)$, que a cada resultado ζ associa um

Quando o conjunto T da definição 2.1 é o conjunto dos números inteiros \mathbb{Z} , então $\{\mathbf{x}_t\}$ é um processo estocástico de tempo discreto (ou seqüência aleatória); $\{\mathbf{x}_t\}$ é um processo estocástico de tempo contínuo se T é tomado como o conjunto dos números reais \mathbb{R} .

A rigor, a variável aleatória \mathbf{x}_t da definição 2.1 é uma função de dois argumentos $\mathbf{x}(t, \zeta)$, $t \in T$, $\zeta \in \Omega$, uma vez que é definida sobre um espaço amostral Ω . Para cada $\zeta \in \Omega$ tem-se uma realização, trajetória ou série temporal x_t [42, pág. 22]. O conjunto de todas as realizações é denominado *ensemble*. Note-se que cada trajetória é uma função ou seqüência não-aleatória, e que, para cada t fixo, x_t é um número.

O restante desta tese também adota a notação \mathbf{x}_t para um processo estocástico $\{\mathbf{x}_t, t \in T\}$, o que é usual na literatura de engenharia elétrica (vide, por exemplo, os livros [23], [46] e [47]).

Um processo \mathbf{x}_t é completamente especificado por meio de suas **distribuições finito-dimensionais** (ou funções de distribuição de probabilidade⁷ de ordem n)

$$F_{\mathbf{x}}(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = P\{\mathbf{x}(t_1) \leq x_1, \mathbf{x}(t_2) \leq x_2, \dots, \mathbf{x}(t_n) \leq x_n\} \quad (2.10)$$

em que t_1, t_2, \dots, t_n são elementos quaisquer de T e $n \geq 1$.

A **função densidade de probabilidade (PDF) de ordem n** associada à distribuição finito-dimensional (2.10)⁸ é dada por

$$f_{\mathbf{x}}(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = \frac{\partial^n F_{\mathbf{x}}(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n)}{\partial x_1 \partial x_2 \dots \partial x_n}. \quad (2.11)$$

Aplicando-se a fórmula da densidade de probabilidade condicional

$$f_{\mathbf{x}}(x_k | x_{k-1}, \dots, x_1) = \frac{f_{\mathbf{x}}(x_1, \dots, x_{k-1}, x_k)}{f_{\mathbf{x}}(x_1, \dots, x_{k-1})}, \quad (2.12)$$

em que $f_{\mathbf{x}}(x_1, \dots, x_{k-1}, x_k)$ denota $f_{\mathbf{x}}(x_1, \dots, x_{k-1}, x_k; t_1, \dots, t_{k-1}, t_k)$, seguidamente sobre $f_{\mathbf{x}}(x_1, \dots, x_{n-1}, x_n)$ obtém-se a regra da cadeia da probabilidade [47, pág. 362]

$$f_{\mathbf{x}}(x_1, x_2, \dots, x_n) = f_{\mathbf{x}}(x_1) f_{\mathbf{x}}(x_2 | x_1) f_{\mathbf{x}}(x_3 | x_2, x_1) \dots f_{\mathbf{x}}(x_n | x_{n-1}, \dots, x_1). \quad (2.13)$$

número real $\mathbf{x}(\zeta)$, cujo domínio é Ω e a imagem é \mathbb{R} , é uma variável aleatória somente se a imagem inversa sob $\mathbf{x}(\cdot)$ de todos os subconjuntos de Borel em \mathbb{R} (esta coleção de subconjuntos é denominada Álgebra de Borel) são eventos. Isto quer dizer que o evento $\{\zeta : \mathbf{x}(\zeta) \leq x\} \subset \mathfrak{S}$ para qualquer $x \in \mathbb{R}$ [47, pág. 59].

⁷A função de distribuição de probabilidade de primeira ordem $F_{\mathbf{x}}(x)$ também é conhecida como função de distribuição acumulada (*Cumulative Distribution Function* (CDF)).

⁸Supondo-se que a mesma seja diferenciável.

Quando \mathbf{x}_t é uma sequência de variáveis aleatórias **mutuamente independentes**, (2.13) pode ser reescrita como

$$f_{\mathbf{x}}(x_1, x_2, \dots, x_n) = f_{\mathbf{x}}(x_1)f_{\mathbf{x}}(x_2) \dots f_{\mathbf{x}}(x_n). \quad (2.14)$$

Definição 2.2 (Processo Puramente Estocástico). Um processo $\{\mathbf{x}_t, t \in \mathbb{Z}\}$ **puramente estocástico** é uma sequência de variáveis aleatórias mutuamente independentes [49]. ■

Definição 2.3 (Processo Independente e Identicamente Distribuído). Um processo **Independente e Identicamente Distribuído (IID)** $\{\mathbf{x}_t, t \in \mathbb{Z}\}$, denotado por $\mathbf{x}_t \sim \text{IID}$, é um processo puramente estocástico e identicamente distribuído [49]. ■

Os conceitos de **estacionariedade em sentido estrito** e em **sentido amplo** são enunciados a seguir.

Definição 2.4 (Estacionariedade em Sentido Estrito). Um processo aleatório \mathbf{x}_t é estacionário em sentido estrito se [46, pág. 297]

$$F_{\mathbf{x}}(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = F_{\mathbf{x}}(x_1, x_2, \dots, x_n; t_1 + c, t_2 + c, \dots, t_n + c),$$

para qualquer c . ■ (2.15)

De acordo com (2.15), as propriedades estatísticas de um processo estacionário em sentido estrito \mathbf{x}_t não mudam com uma translação do mesmo, ou seja, \mathbf{x}_t e $\mathbf{x}(t + c)$ possuem as mesmas estatísticas para qualquer c . Esta condição é bastante forte e difícil de ser verificada empiricamente, porque muitas vezes não se sabe quais são as distribuições finito-dimensionais que caracterizam um determinado processo aleatório na prática. Sendo assim, adota-se uma caracterização parcial do processo por meio da estimação de momentos de baixa ordem como média, autocorrelação e autocovariância e assume-se uma condição mais fraca de estacionariedade, conhecida como estacionariedade em sentido amplo, que será definida mais adiante (vide definição 2.5).

A média $\mu_{\mathbf{x}}(t)$ de \mathbf{x}_t é o valor esperado da variável aleatória \mathbf{x}_t :

$$\mu_{\mathbf{x}}(t) = E[\mathbf{x}_t] = \int_{-\infty}^{\infty} x f_{\mathbf{x}}(x; t) dx, \quad (2.16)$$

em que $f_{\mathbf{x}}(x; t)$ denota a função densidade de probabilidade de primeira ordem de \mathbf{x}_t .

A autocorrelação $R_{\mathbf{x}}(t_1, t_2)$ de \mathbf{x}_t é o valor esperado do produto $\mathbf{x}_{t_1} \mathbf{x}_{t_2}$ [46,

pág. 288]:

$$R_{\mathbf{x}}(t_1, t_2) = E[\mathbf{x}_{t_1} \mathbf{x}_{t_2}] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_{\mathbf{x}}(x_1, x_2; t_1, t_2) dx_1 dx_2, \quad (2.17)$$

em que $f_{\mathbf{x}}(x_1, x_2; t_1, t_2)$ denota a função densidade de probabilidade de segunda ordem de \mathbf{x}_t .

A autocovariância $C_{\mathbf{x}}(t_1, t_2)$ de \mathbf{x}_t é a covariância das variáveis aleatórias \mathbf{x}_{t_1} e \mathbf{x}_{t_2} [46, pág. 289]:

$$C_{\mathbf{x}}(t_1, t_2) = R_{\mathbf{x}}(t_1, t_2) - \mu_{t_1} \mu_{t_2}. \quad (2.18)$$

O coeficiente de correlação $\rho_{\mathbf{x}}(t_1, t_2)$ de \mathbf{x}_t é a razão [46, pág. 294]:

$$\rho_{\mathbf{x}}(t_1, t_2) = \frac{C_{\mathbf{x}}(t_1, t_2)}{\sqrt{C_{\mathbf{x}}(t_1, t_1) C_{\mathbf{x}}(t_2, t_2)}}. \quad (2.19)$$

A grandeza $\rho_{\mathbf{x}}(t_1, t_2)$ é uma medida do grau de dependência linear entre as variáveis aleatórias \mathbf{x}_{t_1} e \mathbf{x}_{t_2} , ou seja, quantifica o quanto o diagrama de dispersão de \mathbf{x}_{t_1} versus \mathbf{x}_{t_2} se aproxima de uma reta. Observe-se que a maioria dos autores das áreas de séries temporais e de redes utiliza o termo Função de Autocorrelação (FAC) (*AutoCorrelation Function* (ACF)) quando se refere à grandeza definida por (2.19) [1], [2], [39], [43], [42] e [50].

Definição 2.5 (Estacionariedade em Sentido Amplo⁹). Um processo aleatório \mathbf{x}_t é estacionário em sentido amplo [46, pág. 298] se a sua média é constante

$$E[\mathbf{x}_t] = \mu_{\mathbf{x}}, \quad (2.20)$$

e se a autocorrelação depende somente do *lag* $\tau = t_2 - t_1$:

$$R_{\mathbf{x}}(t_1, t_2) = R_{\mathbf{x}}(t_1, t_1 + \tau) = R_{\mathbf{x}}(\tau). \quad \blacksquare \quad (2.21)$$

Observe-se que (2.21) implica uma autocovariância que depende do *lag*, ou seja, $C_{\mathbf{x}}(\tau)$. Neste caso, a variância do processo estacionário é constante e é dada por

$$\sigma_{\mathbf{x}}^2 = C_{\mathbf{x}}(0). \quad (2.22)$$

O restante deste trabalho refere-se aos processos estacionários em sentido amplo simplesmente como “processos estacionários”.

O conceito de **estacionariedade** é fundamental para a análise de séries tem-

⁹A nomenclatura “estacionariedade em sentido amplo” é amplamente adotada pelos autores da área de engenharia elétrica [23], [46], [47], enquanto que os estatísticos preferem usar os termos “estacionariedade fraca” ou “estacionariedade de segunda ordem” [39], [42] e [43].

porais, dado que a maior parte dos procedimentos de análise estatística supõe que as séries sejam estacionárias. Contudo, séries empíricas podem apresentar alguma forma de não-estacionariedade não-explosiva, conforme será explicado a seguir. Portanto, é necessário submeter séries não-estacionárias a algum tipo de transformação que as torne estacionárias. Séries reais geralmente apresentam os seguintes tipos de não-estacionariedade [42]:

- (a) não-estacionariedade quanto ao nível: as séries oscilam ao redor de um nível médio durante algum tempo e depois saltam para outro nível. A primeira diferença torna essas séries estacionárias.
- (b) não-estacionariedade quanto à inclinação: as séries flutuam ao redor de uma reta, com inclinação positiva ou negativa. A segunda diferença desse tipo de série é estacionária.

Definição 2.6 (Ruído Branco Independente). Um processo $\mathbf{w}_t \sim \text{IID}$ é um **Ruído Branco Independente (RBI)** quando possui média $\mu_{\mathbf{w}}$ e variância $\sigma_{\mathbf{w}}^2$, $\mathbf{w}_t \sim \text{RBI}(\mu_{\mathbf{w}}, \sigma_{\mathbf{w}}^2)$. ■

Definição 2.7 (Ruído Branco). Uma sequência $\{\mathbf{w}_t, t \in \mathbb{Z}\}$ de variáveis aleatórias não-correlacionadas¹⁰ com média $\mu_{\mathbf{w}}$ e variância $\sigma_{\mathbf{w}}^2$ é denominada **Ruído Branco (RB)**, $\mathbf{w}_t \sim \text{RB}(\mu_{\mathbf{w}}, \sigma_{\mathbf{w}}^2)$. ■

Observação 2.1 (Ruído Branco Gaussiano). Um RB Gaussiano (RBG) \mathbf{w}_t é um RBI e é denotado por $\mathbf{w}_t \sim \mathcal{N}(\mu_{\mathbf{w}}, \sigma_{\mathbf{w}}^2)$, em que \mathcal{N} denota a distribuição normal (Gaussiana) de probabilidades.

Considere um processo estacionário \mathbf{x}_t . A média amostral de uma realização x_t com N pontos é dada por

$$\bar{x} = \frac{1}{N} \sum_{t=1}^N x_t, \quad (2.23)$$

a autocovariância amostral de lag τ por

$$\hat{C}_{\tau} = \frac{1}{N} \sum_{t=\tau+1}^N (x_t - \bar{x})(x_{t-\tau} - \bar{x}), \quad (2.24)$$

e a autocorrelação amostral (*Sample AutoCorrelation Function* (SACF)) de lag τ por

$$\hat{\rho}_{\tau} = \frac{\hat{C}_{\tau}}{\hat{C}_0}, \quad (2.25)$$

¹⁰Neste caso, $\rho(\tau) = 0$ para $\tau \neq 0$.

em que \hat{C}_0 (também denotada por s_x^2)

$$\hat{C}_0 = s_x^2 = \frac{1}{N} \sum_{t=1}^N (x_t - \bar{x})^2 \quad (2.26)$$

é a variância amostral¹¹ de x_t [43, pág.58].

Definição 2.8 (Ergodicidade). Um processo estacionário \mathbf{x}_t é dito ergódico se os seus momentos amostrais convergem em probabilidade¹² para os momentos da população; isto é, se $\bar{\mathbf{x}} \xrightarrow{p} \mu$, $\hat{C}_\tau \xrightarrow{p} C_\tau$ e $\hat{\rho}_\tau \xrightarrow{p} \rho_\tau$ [43, pág. 58]. ■

2.3 Modelagem de Séries Temporais

A modelagem de uma série temporal x_t consiste na estimação de uma função invertível $h(\cdot)$, denominada **modelo** de \mathbf{x}_t , tal que

$$\mathbf{x}_t = h(\dots, \mathbf{w}_{t-2}, \mathbf{w}_{t-1}, \mathbf{w}_t, \mathbf{w}_{t+1}, \mathbf{w}_{t+2}, \dots), \quad (2.27)$$

em que $\mathbf{w}_t \sim \text{IID}$ e

$$g(\dots, \mathbf{x}_{t-2}, \mathbf{x}_{t-1}, \mathbf{x}_t, \mathbf{x}_{t+1}, \mathbf{x}_{t+2}, \dots) = \mathbf{w}_t, \quad (2.28)$$

em que $g(\cdot) = h^{-1}(\cdot)$. O processo \mathbf{w}_t é a **inovação** no instante t e representa a nova informação sobre a série que é obtida no instante t .

Na prática, o modelo ajustado é **causal**, ou seja,

$$\mathbf{x}_t = h(\mathbf{w}_t, \mathbf{w}_{t-1}, \mathbf{w}_{t-2}, \dots). \quad (2.29)$$

A metodologia de construção de um modelo é baseada no ciclo iterativo ilustrado pela Fig. 2.1 [37], [42]:

- (a) uma classe geral de modelos é considerada para a análise (**especificação**);
- (b) há a **identificação** de um modelo, com base em critérios estatísticos;
- (c) segue-se a fase de **estimação**, na qual os parâmetros do modelo são obtidos.

Na prática, é importante que o modelo seja **parcimonioso**¹³ e

¹¹A variância amostral também pode ser definida como $s_x^2 = \frac{1}{N-1} \sum_{t=1}^N (x_t - \bar{x})^2$ [23, pág.241].

¹²Diz-se que a seqüência $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \dots\}$ converge em probabilidade para x se $\lim_{n \rightarrow \infty} P(|\mathbf{x}_n - x| \geq \epsilon) = 0$ para todo $\epsilon > 0$.

¹³Diz-se que um modelo é parcimonioso quando o mesmo utiliza poucos parâmetros. A utilização de um número excessivo de parâmetros é indesejável porque o grau de incerteza no procedimento de inferência estatística aumenta com o aumento do número de parâmetros.

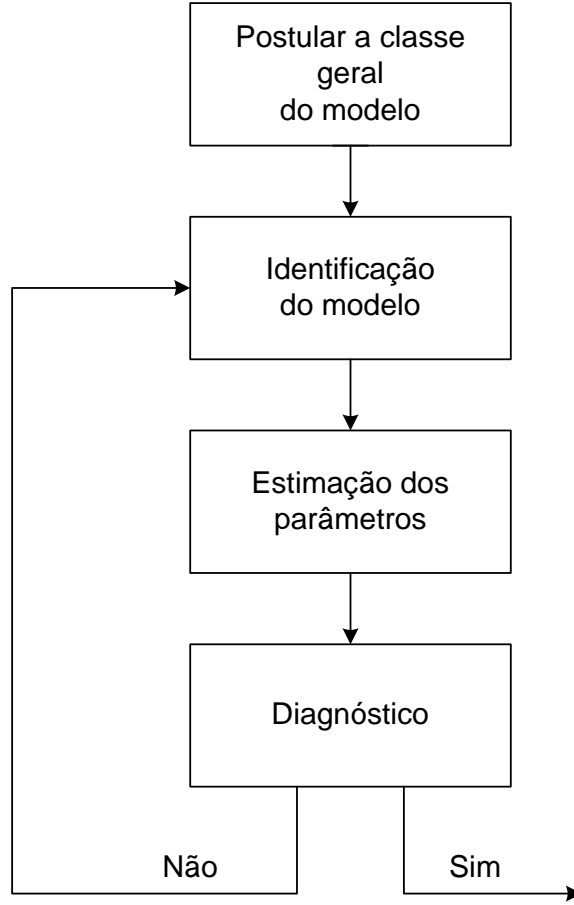


Figura 2.1: Ciclo iterativo Box-Jenkins.

(d) finalmente, há o **diagnóstico** do modelo ajustado por meio de uma análise estatística da série w_t de resíduos (w_t é compatível com um RB?).

O processo \mathbf{x}_t de (2.29) é **linear** quando corresponde à convolução de um processo $\mathbf{w}_t \sim \text{IID}$ com uma sequência determinística h_t [51, pág. 377]

$$\begin{aligned}
 \mathbf{x}_t &= h_t \star \mathbf{w}_t = \sum_{k=0}^{\infty} h_k \mathbf{w}_{t-k} \\
 &= \mathbf{w}_t + h_1 \mathbf{w}_{t-1} + h_2 \mathbf{w}_{t-2} + \dots \\
 &= (1 + h_1 B + h_2 B^2 + \dots) \mathbf{w}_t \\
 &= H(B) \mathbf{w}_t
 \end{aligned} \tag{2.30}$$

em que o símbolo \star denota a operação de convolução e $h_0 = 1$. A Eq. (2.30) também é conhecida como representação de **média móvel de ordem infinita** ($\text{MA}(\infty)$) [41].

A forma geral de filtro linear de (2.30) é¹⁴

$$\mathbf{x}(t) = \sum_{k=1}^p \phi_k \mathbf{x}(t-k) + \mathbf{w}(t) - \sum_{k=1}^q \theta_k \mathbf{w}(t-k). \quad (2.31)$$

A sequência h_t é denominada resposta impulsiva de (2.31), também conhecida como **modelo ARMA** de ordens p e q (ARMA(p, q)).

Numa forma mais compacta, tem-se que

$$\phi(B)\mathbf{x}_t = \theta(B)\mathbf{w}_t, \quad (2.32)$$

em que $\phi(B)$ é o operador auto-regressivo de ordem p

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (2.33)$$

e $\theta(B)$ denota o operador de média móvel de ordem q

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q. \quad (2.34)$$

Se todos os pólos e zeros da função de transferência

$$H(z) = \sum_{k=0}^{\infty} h_k z^{-k} = \frac{\theta(z)}{\phi(z)} \quad (2.35)$$

do filtro (2.32) estiverem dentro do círculo de raio unitário

$$\phi(z) = 0, \quad |z| < 1 \quad (2.36)$$

$$\theta(z) = 0, \quad |z| < 1 \quad (2.37)$$

então o processo \mathbf{x}_t dado por (2.30) é **estacionário** (ou não-explosivo) e **invertível**, respectivamente [40, pág. 537], [51, pág. 377].

Quando as inovações em (2.30) são variáveis aleatórias com distribuição α -estável¹⁵ (α -stable) [51], [53], [54], (2.30) define um **processo não-Gaussiano**

¹⁴Demonstra-se que (2.30) é a única solução do filtro (2.31) [51, pág. 377], [52].

¹⁵Uma variável aleatória X com distribuição α -estável possui **cauda pesada** e é definida pela seguinte função característica [51]:

$$\Phi_X(w) = E[e^{jwX}] = \int_{-\infty}^{\infty} f_X(x) e^{jwx} dx = \exp\{j\mu w - |\sigma w|^\alpha [1 - j\eta \text{sign}(w)\varphi(w, \alpha)]\}, \quad (2.38)$$

em que,

$$\varphi(w, \alpha) = \begin{cases} \tan(\alpha\pi/2) & \text{se } \alpha \neq 1 \\ -\frac{2}{\pi} \ln|w| & \text{se } \alpha = 1, \end{cases} \quad (2.39)$$

e $\text{sign}(\cdot)$ denota a função sinal, α ($0 < \alpha \leq 2$) é o **expoente característico**, μ ($\mu \in \mathbb{R}$) é o **parâmetro de localização**, η ($-1 \leq \eta \leq 1$) é o **parâmetro de assimetria** e $\sigma \geq 0$ é o **parâmetro de dispersão** ou **escala**. A **variância** de X é **infinita** para $0 < \alpha < 2$.

com variância infinita [40, pág. 535], [51, pág. 380]. Este fato é justificado pelo **teorema central do limite generalizado** [54], [55], o qual afirma que, se o limite de uma soma de variáveis aleatórias IID converge, então este limite só pode ser uma variável aleatória com distribuição estável (a distribuição normal é um caso particular da distribuição estável¹⁶). O próximo Capítulo introduz os conceitos de variável aleatória com distribuição de cauda pesada e as distribuições estáveis.

Se as inovações em (2.30) tiverem variância finita, ou seja, $\mathbf{w}_t \sim \text{RBI}(\mu_w, \sigma_w^2)$, então \mathbf{x}_t será Gaussiano quando h_t tiver duração infinita (teorema central do limite) [47, pág. 225]). Note-se que qualquer não-linearidade na função $h(\cdot)$ de (2.29) resulta num processo \mathbf{x}_t **não-linear**. Neste caso, \mathbf{x}_t tem estatísticas necessariamente não-Gaussianas. Por outro lado, processos Gaussianos são necessariamente lineares. O restante deste Capítulo assume que as inovações em (2.30) são do tipo $\mathbf{w}_t \sim \text{RBI}(\mu_w, \sigma_w^2)$.

A autocorrelação do RBI \mathbf{w}_t em (2.30) é

$$R_w(\tau) = \sigma_w^2 \delta_\tau + \mu_w^2, \quad (2.40)$$

em que δ_τ denota o pulso unitário de tempo discreto¹⁷. Logo a sua função densidade espectral ou Densidade Espectral de Potência (DEP)¹⁸ é

$$S_w(f) = \mu_w^2 \delta(f) + \sigma_w^2, \quad -1/2 \leq f \leq 1/2, \quad (2.41)$$

em que f denota a frequência normalizada e $\delta(f)$ é a função generalizada Impulso (ou Delta de Dirac).

A DEP do processo linear \mathbf{x}_t é [45]

$$\begin{aligned} S_x(f) &= |H(f)|^2 S_w(f) \\ &= \mu_w^2 |H(0)|^2 \delta(f) + \sigma_w^2 |H(f)|^2, \quad -1/2 \leq f \leq 1/2, \end{aligned} \quad (2.42)$$

em que $H(f) = H(z)|_{z=e^{j2\pi f}}$ é a resposta em frequência do filtro.

A autocovariância de \mathbf{x}_t é dada por [42, pág. 112]

$$C_x(\tau) = \sigma_w^2 \sum_{t=0}^{\infty} h_t h_{t+\tau}, \quad (2.43)$$

¹⁶Quando $\alpha = 2$.

¹⁷ $\delta_\tau = 1$ para $\tau = 0$, $\delta_\tau = 0$ para $\tau \neq 0$, $\tau \in \mathbb{Z}$.

¹⁸A DEP de \mathbf{w}_t é definida como a *Transformada de Fourier de Tempo Discreto* (TFTD) da sua autocorrelação $R_w(\tau)$, ou seja, $S_w(f) = \sum_{m=-\infty}^{\infty} R_w(m) e^{-j2\pi f m}$ [47, pág. 351].

e a sua variância por

$$\sigma_x^2 = \sigma_w^2 \sum_{t=0}^{\infty} h_t^2. \quad (2.44)$$

O restante desta tese assume, sem perda de generalidade, que a média do processo linear \mathbf{x}_t de (2.30) seja nula ($\mu_x = 0$).

Como deseja-se que os modelos estimados na prática sejam invertíveis (ou seja, a condição (2.37) é válida), pode-se definir o operador inverso $G(B) = H^{-1}(B)$ e reescrever (2.30) na forma auto-regressiva de ordem infinita (AR(∞))

$$\begin{aligned} \mathbf{x}_t &= g_1 \mathbf{x}_{t-1} + g_2 \mathbf{x}_{t-2} + \dots + \mathbf{w}_t \\ &= \sum_{k=1}^{\infty} g_k \mathbf{x}_{t-k} + \mathbf{w}_t. \end{aligned} \quad (2.45)$$

Portanto, \mathbf{x}_t pode ser interpretado como uma soma ponderada de seus valores passados $\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \dots$ mais uma inovação \mathbf{w}_t . O modelo equivalente AR(∞) sugere que pode-se calcular a probabilidade de um valor futuro \mathbf{x}_{t+k} estar entre dois limites especificados, ou seja, (2.45) afirma que é possível fazer inferências ou **previsões** de valores futuros da série.

2.4 Modelos Auto-Regressivos

Um modelo auto-regressivo de ordem p (AR(p)) satisfaz à equação

$$\phi(B)\mathbf{x}_t = \mathbf{w}_t. \quad (2.46)$$

2.4.1 Função de Autocorrelação

Multiplicando-se ambos os membros de (2.46) por \mathbf{x}_{t-k} e tomando-se a esperança obtém-se

$$E[\mathbf{x}_t \mathbf{x}_{t-k}] = \phi_1 E[\mathbf{x}_{t-1} \mathbf{x}_{t-k}] + \phi_2 E[\mathbf{x}_{t-2} \mathbf{x}_{t-k}] + \dots + \phi_p E[\mathbf{x}_{t-p} \mathbf{x}_{t-k}] + E[\mathbf{w}_t \mathbf{x}_{t-k}],$$

como \mathbf{x}_{t-k} não depende de \mathbf{w}_t , mas somente de ruídos até o instante $t - k$, que são não-correlacionados com \mathbf{w}_t , então $E[\mathbf{w}_t \mathbf{x}_{t-k}] = 0$, $k > 0$, e

$$C_x(k) = \phi_1 C_x(k-1) + \phi_2 C_x(k-2) + \dots + \phi_p C_x(k-p), \quad k > 0. \quad (2.47)$$

Dividindo-se (2.47) por $C_x(0) = \sigma_x^2$, obtém-se

$$\rho_x(k) = \phi_1 \rho_x(k-1) + \phi_2 \rho_x(k-2) + \dots + \phi_p \rho_x(k-p), \quad k > 0, \quad (2.48)$$

ou

$$\phi(B)\rho_x(k) = 0. \quad (2.49)$$

Sejam G_i^{-1} , $i = 1, \dots, p$, as raízes de $\phi(B) = 0$ (**equação característica** do modelo [39, pág. 40]). Então pode-se escrever que

$$\phi(B) = \prod_{i=1}^p (1 - G_i B),$$

e demonstra-se que a solução geral de (2.49) é [42, pág. 116]

$$\rho_x(k) = A_1 G_1^k + A_2 G_2^k + \dots + A_p G_p^k, \quad (2.50)$$

em que as constantes A_i , $i = 1, 2, \dots, p$, são determinadas¹⁹ por condições iniciais sobre $\rho_x(0), \rho_x(1), \dots, \rho_x(p-1)$. Como as raízes da equação característica do modelo devem estar fora do círculo unitário²⁰, deve-se ter que $|G_k| < 1$, $k = 1, \dots, p$. Portanto, o gráfico da FAC de um processo AR(p) é, em geral, constituído de uma mistura de exponenciais e senóides amortecidas [39, pág. 40].

2.4.2 Identificação

Na prática, a ordem p de uma série AR é desconhecida e deve ser especificada de forma empírica. Há duas abordagens para se determinar o valor de p : i) uso da **Função de Autocorrelação Parcial (FACP)** e ii) uso de algum **critério de seleção (identificação) de modelo**[39].

2.4.2.1 FACP de modelos AR(p)

Seja ϕ_{mi} o i -ésimo coeficiente de um processo AR(m), de modo que o último coeficiente seja ϕ_{mm} . Fazendo-se $k = 1, \dots, m$ em (2.48) (adota-se a seguir a

¹⁹Por exemplo, considere um modelo AR(2)

$$\mathbf{x}_t = \phi_1 \mathbf{x}_{t-1} + \phi_2 \mathbf{x}_{t-2} + \mathbf{w}_t$$

e a sua FAC, a qual satisfaz a equação de diferenças de segunda ordem

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad k > 0$$

com valores iniciais $\rho_0 = 1$ e $\rho_1 = \frac{\phi_1}{1-\phi_2}$. De (2.50), tem-se que a solução geral dessa Eq. de diferenças de segunda ordem é [37, pág.59]

$$\rho_k = \frac{G_1(1 - G_2^2)G_1^k - G_2(1 - G_1^2)G_2^k}{(G_1 - G_2)(1 + G_1 G_2)}.$$

²⁰Como $B = z^{-1}$, então as Eqs. 2.36 e 2.37 podem ser reescritas como $\phi(B) = 0$ para $|B| > 1$ e $\theta(B) = 0$ para $|B| > 1$.

notação simplificada $\rho_x(k) = \rho_k$) e levando-se em conta que $\rho_k = \rho_{-k}$ (simetria par da FAC), obtêm-se as **equações de Yule-Walker** [37, pág.64], [42, pág.134], [43, pág.69]

$$\begin{aligned}\rho_1 &= \phi_{m1} + \phi_{m2}\rho_1 + \dots + \phi_{mm}\rho_{m-1}, \\ \rho_2 &= \phi_{m1}\rho_1 + \phi_{m2} + \dots + \phi_{mm}\rho_{m-2}, \\ &\vdots \\ \rho_m &= \phi_{m1}\rho_{m-1} + \phi_{m2}\rho_{m-2} + \dots + \phi_{mm},\end{aligned}\tag{2.51}$$

que podem ser reescritas em forma matricial

$$\begin{bmatrix} 1 & \rho_1 & \dots & \rho_{m-1} \\ \rho_1 & 1 & \dots & \rho_{m-2} \\ \vdots & \vdots & \dots & \vdots \\ \rho_{m-1} & \rho_{m-2} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{m1} \\ \phi_{m2} \\ \vdots \\ \phi_{mm} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_m \end{bmatrix}\tag{2.52}$$

ou na forma compacta

$$R_m \phi_m = \tilde{\rho}_m,\tag{2.53}$$

em que R_m denota a matriz de autocorrelações de ordem m , ϕ_m é o vetor de parâmetros do modelo e $\tilde{\rho}_m$ é o vetor de autocorrelações.

Resolvendo-se (2.53) sucessivamente para $m = 1, 2, \dots$, obtêm-se

$$\begin{aligned}\phi_{11} &= \rho_1 \\ \phi_{22} &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} \\ \phi_{33} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}\end{aligned}\tag{2.54}$$

e, em geral,

$$\phi_{mm} = \frac{|R_m^*|}{|R_m|},\tag{2.55}$$

em que $|R_m|$ denota o determinante da matriz de autocorrelações de ordem m e R_m^* é a matriz R_m com a última coluna substituída pelo vetor de autocorrelações.

A sequência $\{\phi_{mm}, m = 1, 2, \dots\}$ é a FACP. Demonstra-se que um modelo $\text{AR}(p)$ tem $\phi_{mm} \neq 0$ para $m \leq p$ e $\phi_{mm} = 0$ para $m > p$ [37].

A FACP pode ser estimada por meio do ajuste da sequência de modelos $\text{AR}(m)$, $m = 1, 2, \dots$

$$\begin{aligned} \mathbf{x}_t &= \phi_{11}\mathbf{x}_{t-1} + \mathbf{w}_{1t} \\ \mathbf{x}_t &= \phi_{21}\mathbf{x}_{t-1} + \phi_{22}\mathbf{x}_{t-2} + \mathbf{w}_{2t} \\ &\vdots \\ \mathbf{x}_t &= \phi_{m1}\mathbf{x}_{t-1} + \phi_{m2}\mathbf{x}_{t-2} + \dots + \phi_{mm}\mathbf{x}_{t-m} + \mathbf{w}_{mt}, \\ &\vdots \end{aligned} \tag{2.56}$$

pelo método dos mínimos quadrados [39, pág. 40], [43, pág. 70]. A sequência $\{\hat{\phi}_{mm}, m = 1, 2, \dots\}$ obtida é a Função de Autocorrelação Parcial Amostral (FACPA).

2.4.2.2 Identificação do Modelo

A idéia básica de um critério de seleção (ou critério de informação) de modelo ARMA é escolher as ordens k e l que minimizam a quantidade [48]

$$P(k, l) = \ln \hat{\sigma}_{k,l}^2 + (k + l) \frac{C(N)}{N}, \tag{2.57}$$

em que $\hat{\sigma}_{k,l}^2$ é uma estimativa da variância residual obtida ajustando-se um modelo $\text{ARMA}(k, l)$ às N observações da série e $C(N)$ é uma função do tamanho da série. A quantidade $(k + l) \frac{C(N)}{N}$ é denominada termo penalizador e aumenta quando o número de parâmetros aumenta, enquanto que $\hat{\sigma}_{k,l}^2$ diminui.

Akaike [6], [56] propôs o critério de informação

$$AIC(k, l) = \ln \hat{\sigma}_{k,l}^2 + \frac{2(k + l)}{N}, \tag{2.58}$$

conhecido como AIC, em que $\hat{\sigma}_{k,l}^2$ é o estimador de máxima verossimilhança de σ_w^2 para um modelo $\text{ARMA}(k, l)$. Deve-se especificar valores limites superiores K e L para k e l e calcular (2.58) para todas as possíveis combinações (k, l) com $0 \leq k \leq K$ e $0 \leq l \leq L$. Em geral, K e L são funções de N , por exemplo, $K = L = \ln N$ [48, pág. 85].

Para o caso de modelos $\text{AR}(p)$, (2.58) reduz-se a

$$AIC(k) = \ln \hat{\sigma}_k^2 + \frac{2k}{N}, \quad k \leq K. \tag{2.59}$$

Outro critério sistemático bastante utilizado é o (Schwarz) *Bayesian Information Criteria* (BIC) [43, pág. 77]

$$BIC(k, l) = \ln \hat{\sigma}_{k,l}^2 + \frac{\ln N}{N}(k + l). \quad (2.60)$$

Para o caso de modelos $AR(p)$, (2.60) reduz-se a

$$BIC(k) = \ln \hat{\sigma}_k^2 + \frac{k \ln N}{N}. \quad (2.61)$$

2.4.3 Estimação de Modelos AR

Tendo-se identificado a ordem p do modelo AR, pode-se partir para a fase de estimação dos parâmetros. Os métodos dos momentos, Mínimos Quadrados (MQ) e Máxima Verossimilhança (MV) podem ser usados para tal [42], [44], [48]. Como em geral os estimadores de momentos não são bons [42], pacotes estatísticos como S-PLUS, E-VIEWS, etc., utilizam algum Estimador de Mínimos Quadrados (EMQ) ou Estimador de Máxima Verossimilhança (EMV). Os livros [42] e [48] de Morettin fornecem maiores detalhes sobre o assunto.

2.5 Modelagem de Séries Não-Estacionárias

Um processo não-estacionário tem momentos que dependem do tempo. As formas mais comuns de não-estacionariedade são causadas pela variação da média e da variância com o tempo.

2.5.1 Séries Estacionárias com Relação a Tendências

Um processo \mathbf{y}_t é dito estacionário com relação a tendências se for da forma

$$\mathbf{y}_t = TD_t + \mathbf{x}_t, \quad (2.62)$$

em que TD_t denota o termo de tendências determinísticas (constante, tendência, sazonalidade, etc.) que depende de t e \mathbf{x}_t é um processo estacionário. Por exemplo, o processo \mathbf{y}_t dado por

$$\begin{aligned} \mathbf{y}_t &= \mu + \delta t + \mathbf{x}_t, & \mathbf{x}_t &= \phi \mathbf{x}_{t-1} + \mathbf{w}_t \\ \mathbf{y}_t - \mu - \delta t &= \phi(\mathbf{y}_{t-1} - \mu - \delta(t-1)) + \mathbf{w}_t \\ \mathbf{y}_t &= c + \beta t + \phi \mathbf{y}_{t-1} + \mathbf{w}_t \end{aligned} \quad (2.63)$$

em que $|\phi| < 1$, $c = \mu(1 - \phi) + \delta$, $\beta = \delta(1 - \phi)t$ e \mathbf{w}_t é um RBG com média nula e potência σ^2 , é um processo AR(1) estacionário com relação a tendências [43].

2.5.2 Modelo ARIMA

Se o processo que corresponde à diferença de ordem $d = 1, 2, \dots$ de \mathbf{x}_t

$$\mathbf{y}_t = (1 - B)^d \mathbf{x}_t = \Delta^d \mathbf{x}_t \quad (2.64)$$

for estacionário, então pode-se representar \mathbf{y}_t por meio de um modelo ARMA(p, q)

$$\phi(B)\mathbf{y}_t = \theta(B)\mathbf{w}_t. \quad (2.65)$$

Neste caso,

$$\phi(B)\Delta^d \mathbf{x}_t = \theta(B)\mathbf{w}_t \quad (2.66)$$

é um modelo ARIMA(p, d, q) e diz-se que \mathbf{x}_t é uma “integral” de \mathbf{y}_t [42], pois

$$\mathbf{x}_t = S^d \mathbf{y}_t. \quad (2.67)$$

É daí que surge o termo “integrado” do acrônimo ARIMA, indicando que (2.66) é um modelo integrado de ordem d , denotado por $\mathbf{x}_t \sim I(d)$.

Como o filtro ARIMA(p, d, q)

$$H(z) = \frac{\theta(z)}{\phi(z)(1 - z^{-1})^d} \quad (2.68)$$

é marginalmente estável [52], pois possui d pólos sobre o círculo unitário, \mathbf{x}_t de (2.66) é um processo **não-estacionário homogêneo** (no sentido de ser **não-explosivo**) ou portador de **raízes unitárias** [39], [42], [43]. Observe-se que [42, pág.139]:

- (a) $d = 1$ corresponde ao caso de séries não-estacionárias homogêneas quanto ao nível (oscilam ao redor de um nível médio durante algum tempo e depois saltam para outro nível temporário);
- (b) $d = 2$ corresponde ao caso de séries não-estacionárias homogêneas quanto à inclinação (oscilam numa direção por algum tempo e depois mudam para outra direção temporária).

O modelo ARIMA (2.66) pode ser representado de três formas:

(a) ARMA($p + d, q$) (similar à Eq. (2.31))

$$\mathbf{x}(t) = \sum_{k=1}^{p+d} \varphi_k \mathbf{x}(t-k) + \mathbf{w}(t) - \sum_{k=1}^q \theta_k \mathbf{w}(t-k); \quad (2.69)$$

(b) AR(∞) (forma invertida), dada por (2.45) ou

(c) MA(∞), conforme (2.30).

2.5.3 Passeio Aleatório

Considere o modelo $\mathbf{y}_t \sim I(1)$

$$\mathbf{y}_t = \mathbf{y}_{t-1} + \mathbf{x}_t, \quad (2.70)$$

em que \mathbf{x}_t é um processo estacionário. Supondo-se a condição inicial y_0 , tem-se que (2.70) pode ser reescrita como uma soma integrada

$$\mathbf{y}_t = \mathbf{y}_0 + \sum_{j=1}^t \mathbf{x}_j. \quad (2.71)$$

A soma integrada $\sum_{j=1}^t \mathbf{x}_j$ é denominada tendência estocástica e é denotada por TS_t . Observe-se que

$$TS_t = TS_{t-1} + \mathbf{x}_t, \quad (2.72)$$

em que $TS_0 = 0$.

Se $\mathbf{x}_t \sim \mathcal{N}(0, \sigma_x^2)$ em (2.70), então \mathbf{y}_t é conhecido como **passeio aleatório** ou **casual**. Incluindo-se uma constante no lado direito de (2.70), tem-se um passeio aleatório com *drift*,

$$\mathbf{y}_t = \theta_0 + \mathbf{y}_{t-1} + \mathbf{x}_t. \quad (2.73)$$

Dada a condição inicial y_0 , pode-se escrever

$$\begin{aligned} \mathbf{y}_t &= y_0 + \theta_0 t + \sum_{j=1}^t \mathbf{x}_j \\ &= TD_t + TS_t \end{aligned} \quad (2.74)$$

e demonstra-se que a média, variância, autocovariância e ACF de \mathbf{y}_t são dadas

respectivamente por [57]

$$\mu_t = y_0 + t\theta_0 \quad (2.75)$$

$$\sigma^2(t) = t\sigma_x^2 \quad (2.76)$$

$$C_k(t) = (t - k)\sigma_x^2 \quad (2.77)$$

$$\rho_k(t) = \frac{t - k}{t}. \quad (2.78)$$

Observe que $\rho_k(t) \approx 1$ quando $t \gg k$ e é por isso que a literatura afirma que um passeio aleatório tem “*strong memory*” (“memória forte”) [39]. Note também que a SACF do passeio aleatório decai linearmente para grandes *lags*.

3 A Natureza Fractal do Teletráfego

Este Capítulo introduz a noção de fractal e formula os conceitos de LRD e auto-similaridade de segunda ordem. Também são abordados os conceitos de variável aleatória com distribuição de cauda pesada, distribuições estáveis e impulsividade [30]. Uma discussão detalhada dos conceitos de fractais e multifractais não faz parte do escopo desta tese. Recomenda-se ao leitor interessado a leitura das seguintes referências: [58], [59], [60], [61], [62], [63], [11], [64] e [65].

3.1 Fractais

As formas da geometria clássica - triângulos, círculos, esferas, etc. - perdem suas estruturas quando ampliadas (isto é, quando se dá um *zoom* numa determinada região da figura). Por exemplo, uma pessoa na superfície da Terra tem a impressão de que a mesma é plana (já um astronauta em órbita vê uma Terra redonda). Suponha que um indivíduo não tenha sido informado de que foi posicionado num ponto de um círculo com raio de curvatura da ordem de centenas de quilômetros. Esse observador percebe o círculo como sendo uma reta, apesar disto não ser verdade.

Benoit B. Mandelbrot propôs em 1975 [59] o termo **fractal** (do latim *fractus*, que tem o significado de fraturado, quebrado) para designar objetos matemáticos que possuem uma estrutura rica em detalhes ao longo de muitas escalas de observação. O conjunto de Mandelbrot ou o “boneco de pão-de-mel” da Fig. 3.1 é um fractal matemático com estrutura detalhada (ou seja, é altamente irregular) ao longo de uma série infinita de escalas. Note-se que a Fig. 3.1 (b), obtida por meio de um *zoom* de uma determinada região da Fig. 3.1 (a), mostra que existe um sub-boneco de pão-de-mel embebido no boneco original. Esta característica ilustra uma outra propriedade dos fractais conhecida como **auto-similaridade**: um objeto auto-similar contém réplicas menores de si mesmo em todas as esca-

las. O conjunto de Mandelbrot é um exemplo de fractal determinístico, pois é exatamente auto-similar.

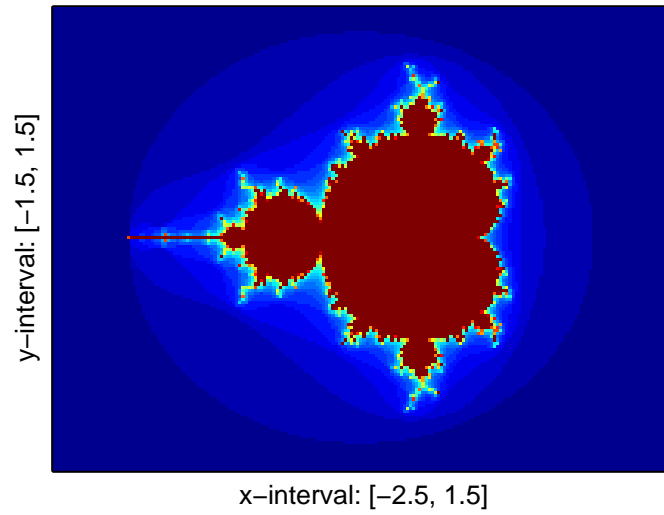
As ciências físicas e humanas nos dão vários exemplos de fractais aleatórios, em que a auto-similaridade ocorre num sentido estatístico, tais como séries temporais nas áreas de climatologia e hidrologia, séries de imagens de ressonância magnética funcional do cérebro humano, movimento turbulento de fluidos e séries de dados financeiros, dentre outros, que envolvem uma extensa faixa de escalas de medição [58], [67], ou seja, que não ocorrem numa escala temporal ou espacial característica. O contorno de um litoral (vide Fig. 3.2) e a couve-flor (vide Fig. 3.3) são exemplos de fractais encontrados na natureza [3].

Um fractal matemático bastante conhecido é o conjunto de Cantor, descrito em 1883 pelo matemático alemão Georg Cantor [68] apud [69], o qual é constituído por um número infinito não-enumerável de pontos no segmento $[0, 1]$ da reta real. O conjunto de Cantor pode ser obtido de acordo com os seguintes passos: 1) comece com o intervalo $[0, 1]$, 2) remova o intervalo aberto $(1/3, 2/3)$, ou seja, remova o terço do meio de $[0, 1]$, mas não os números $1/3$ e $2/3$ e 3) a cada nova iteração, remova o terço do meio (intervalo aberto) dos segmentos resultantes do passo anterior. A Fig. 3.4 mostra as cinco primeiras iterações do procedimento acima descrito.

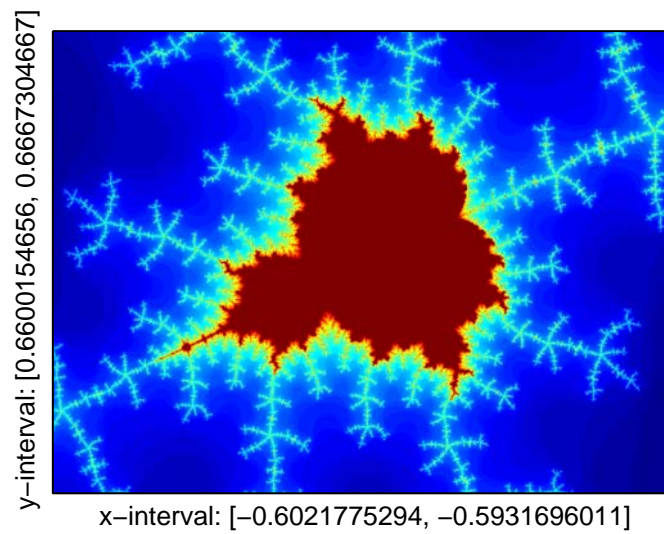
A Fig. 3.5 ilustra que o tráfego Ethernet coletado numa rede local da Universidade de Drexel é auto-similar e altamente impulsivo em quatro escalas temporais de agregação (10 ms, 100 ms, 1 s e 10 s). A série na escala de 100 ms foi obtida por meio de uma agregação da série na escala de 10 ms, ou seja, uma ordenada da série de 100 ms corresponde à soma dos *bytes* em 10 *bins* consecutivos da série na escala de 10 ms. O mesmo procedimento foi aplicado para obtenção das séries nas escalas de 1 s e 10 s, que são agregações das séries nas escalas de 100 ms e 1 s, respectivamente. É surpreendente observar que agregações sucessivas não suavizam o tráfego (a suavização aconteceria se o tráfego fosse bem modelado pelo processo de Poisson¹ [2]). Note-se que a alternância de períodos de surtos e de suavidade é preservada nas quatro escalas de tempo em questão. Isto permite afirmar que não há uma escala de tempo característica para a ocorrência de surtos. Esta propriedade de invariância com relação à mudança de escala temporal é conhecida na literatura como “*scaling invariance*” ou “*scaling behavior*” (que devem ser entendidos como sinônimos de auto-similaridade) [61], [70].

Na Fig. 3.6, pode-se observar, qualitativamente, as diferenças entre o tráfego

¹Agregações sucessivas de uma série temporal de Poisson tendem a produzir séries do tipo RB, conforme ilustrado pela Fig. 3.6 (vide séries de Poisson da coluna central).



(a)



(b)

Figura 3.1: A Fig. (a) é o conjunto de Mandelbrot, também conhecido como “boneco de pão-de-mel”. A Fig (b) mostra um sub-boneco de pão-de-mel que está embebido na Fig. (a). As figuras foram obtidas via **MATLAB** utilizando-se o código de A. Klimke [66], da Universidade de Stuttgart.

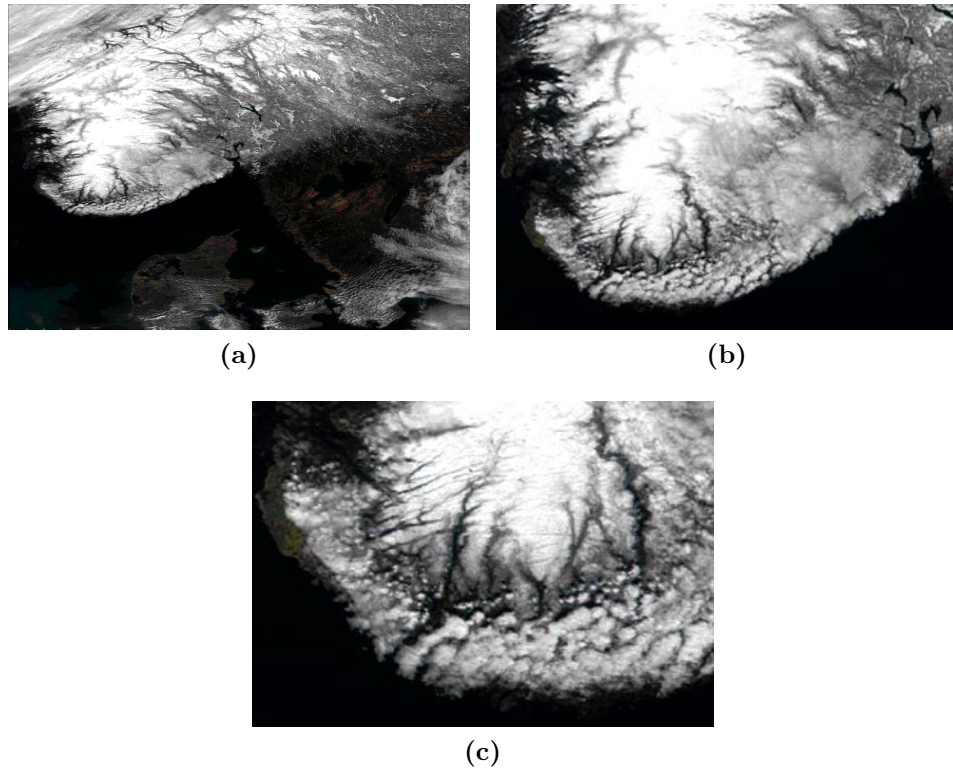


Figura 3.2: As Figs. (a), (b) e (c) são fotos do litoral da Suécia (com resoluções de 1,4 km, 500 m e 250 m, respectivamente) que foram tiradas pelo satélite Terra em 2000 [3]. Essas fotos mostram que o contorno de um litoral é fractal: quando ampliado, novos detalhes são revelados.



Figura 3.3: Foto que ilustra a auto-similaridade de uma couve-flor [3]. Note que as partes menores se parecem com a couve-flor original.

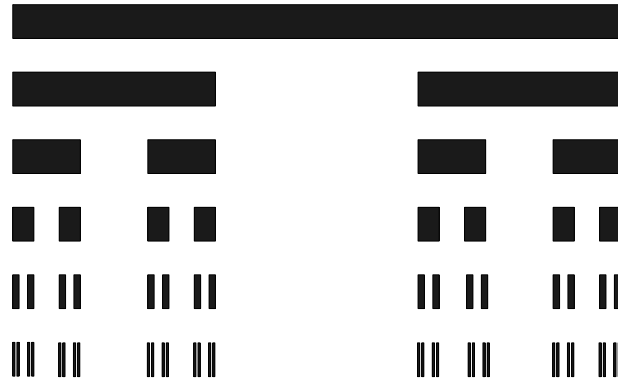


Figura 3.4: Ilustração das cinco primeiras iterações para obtenção do conjunto de Cantor. Repare-se que o formato de uma determinada parte do objeto, quando magnificada, se assemelha ao formato do todo (auto-similaridade).

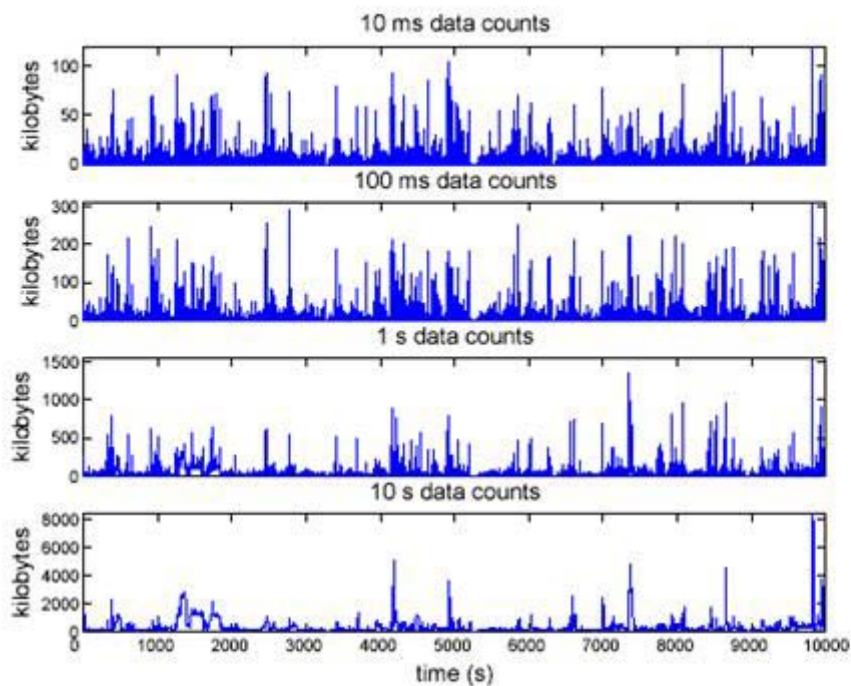


Figura 3.5: Visualização do tráfego *Fast Ethernet* (100 Mbps) coletado num servidor WWW/Email/FTP da Universidade de Drexel [4] em quatro diferentes níveis de agregação: 10 ms, 100 ms, 1 s e 10 s (de cima para baixo). Este *trace* tem um tamanho total de 3 h.

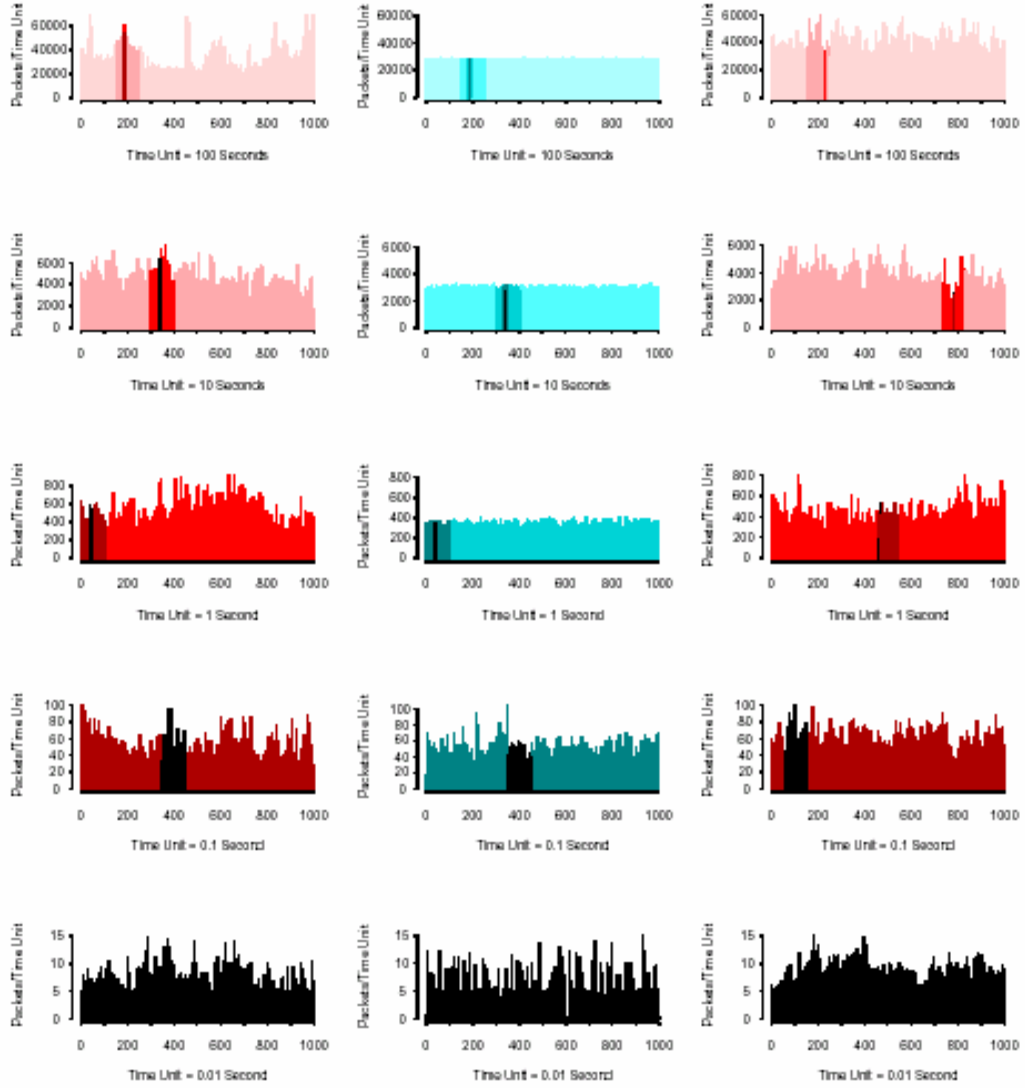


Figura 3.6: Comparação de tráfego Ethernet real e sintetizado [1]. À esquerda tem-se o tráfego real, ao centro tráfego simulado utilizando-se o modelo de Poisson e à direita tráfego gerado por meio de um modelo auto-similar.

Ethernet real e o tráfego simulado por meio do modelo de Poisson [1]. Para o tráfego Ethernet (à esquerda), note-se que a alternância de períodos de surtos e de suavidade também é preservada em várias escalas de tempo, como na Fig. 3.5. Para o tráfego Poisson (ao centro), note-se que períodos de surto de tráfego ocorrem em escalas de tempo menores (10 milissegundos e 0,1 segundo) e que a agregação resulta numa série de nível aproximadamente constante (vide o gráfico em que a unidade de tempo é igual a 10 segundos), o que é uma clara indicação de que as propriedades estatísticas do tráfego Poisson não são mantidas ao longo de várias escalas de agregação. Para o tráfego sintetizado por meio de um modelo auto-similar (à direita), pode-se observar que também há alternância de períodos de surtos e de suavidade em várias escalas de tempo.

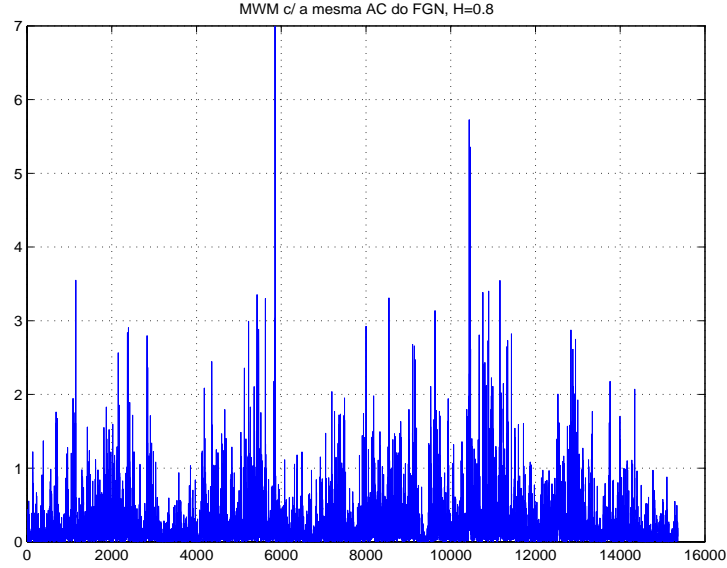


Figura 3.7: Tráfego multifractal simulado por meio do modelo MWM. Note-se que a impulsividade da série varia no tempo (heterocedasticidade).

3.2 O Expoente de Hurst

Por razões históricas, o grau de persistência (LRD) de uma série temporal é caracterizado pelo parâmetro H , $0 < H < 1$, de Hurst² [71]. Uma série temporal é LRD (auto-similar) quando $1/2 < H < 1$ e tem memória curta ou dependência de curta duração (*Short-Range Dependence* (SRD)) para $0 < H \leq 1/2$ [58]. Quanto mais próximo H estiver de 1, maior será o grau de persistência da série. Informalmente, diz-se que uma série é monofractal quando H é invariante no tempo e multifractal quando H varia no tempo de forma determinística ou aleatória. Os artigos [61], [72], [73], [74] mostraram que o tráfego WAN pode ser multifractal (com distribuição marginal não-Gaussiana) em escalas refinadas de tempo (até centenas de milissegundos), em contraste com o comportamento monofractal (ou auto-similar) que foi observado para o tráfego LAN [1]. A Fig. 3.7 mostra uma realização produzida pelo modelo *Multifractal Wavelet Model* (MWM) de Riedi *et al* [11].

Vários pesquisadores notaram que a FAC de várias séries temporais empíricas em hidrologia [71], economia [75], [76] e telecomunicações [1], [2] apresenta um decaimento extremamente lento, do tipo hiperbólico, para grandes valores de *lag*. Isto quer dizer que a dependência entre observações distantes daquelas séries, apesar de serem pequenas, não são de maneira alguma desprezíveis. Em contraste, séries SRD têm FACs com decaimento rápido, do tipo exponencial. A Fig. 3.8 mostra a FAC de uma série LRD com $H = 0,9$ e $N = 4096$ amostras, que foi

²Hurst, um hidrologista, observou que a série histórica do nível mínimo anual do rio Nilo possui memória longa. Esta série tem sido registrada há centenas de anos.

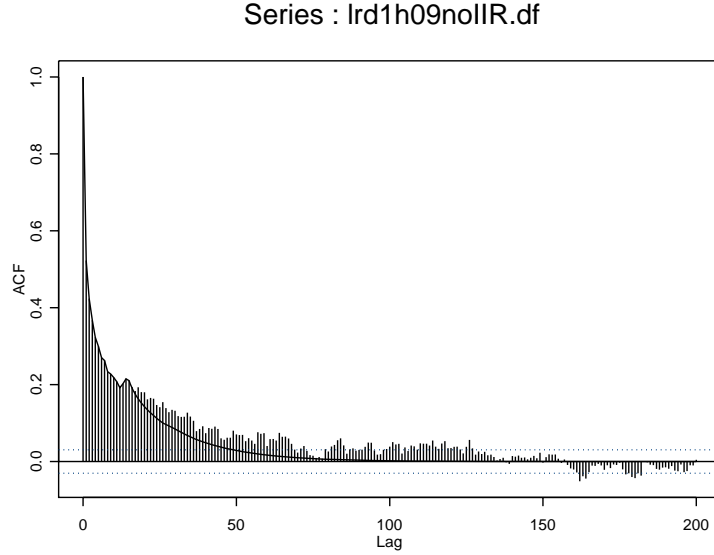


Figura 3.8: FAC de uma série LRD com $H = 0,9$ e $N = 4096$ amostras. A linha contínua mostra a função de autocorrelação do melhor modelo auto-regressivo (AR) que pôde ser ajustado pela função `ar` do *software* **S+FinMetrics** [5] à série segundo o critério AIC [6]. No caso, obteve-se um modelo AR(15). Note que o decaimento da autocorrelação do modelo AR(15) ajustado decai mais rapidamente (exponencialmente) para zero do que o da série LRD.

simulada pelo gerador *wavelet* de tráfego desenvolvido por Mello, Lima, Lipas e Amazonas [77], [78]. Observe que a FAC é altamente persistente, apresenta um decaimento lento para zero e tem valores significativos até o *lag* 200. Tráfego agregado também pode apresentar uma FAC que mistura as características de memória longa e de dependência de curta duração (SRD) [50], [79], [80], comportamento típico da classe de modelos *Fractional Autoregressive Integrated Moving Average* (FARIMA) (ou ARFIMA) [81], [82], que será abordada no Capítulo 4.

A memória longa é caracterizada no domínio da frequência pela existência de uma singularidade do tipo $1/f^\alpha$, $0 < \alpha < 1$ ($\alpha = 2H - 1$), para $f \rightarrow 0$. A Fig. 3.9 mostra que a DEP de um modelo LRD da classe *Fractionally Differenced* (FD)³ [81], [82] com parâmetro $d = 0,4$ ($d = H - 1/2$) apresenta o comportamento $1/f^\alpha$ na origem do espectro, ao passo que um modelo AR de ordem 4 (AR(4)) não. A literatura costuma referir-se aos processos LRD como “ruídos $1/f$ ”.

³O modelo FD(d) corresponde ao processo FARIMA(p, d, q) com $p = q = 0$, em que p denota o parâmetro autoregressivo e q o de média móvel.

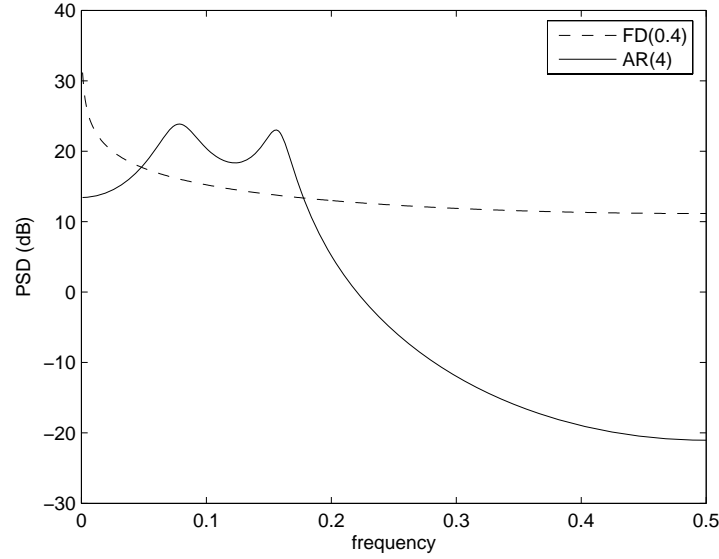


Figura 3.9: DEPs para modelos AR(4) e FD(0,4) de mesma potência.

3.3 LRD e Auto-Similaridade

3.3.1 Dependência de Longa Duração

Considere um processo aleatório estacionário \mathbf{x}_t , $t \in \mathbb{Z}$, com média $\mu_{\mathbf{x}}$ e variância $\sigma_{\mathbf{x}}^2$. Sejam x_1, x_2, \dots, x_N observações de uma realização x_t . Se as variáveis aleatórias $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ forem independentes ou não-correlacionadas, então, a variância de \bar{x} (média amostral) será dada por

$$\sigma_{\bar{x}}^2 = \frac{\sigma_{\mathbf{x}}^2}{N}. \quad (3.1)$$

Se a amostra for suficientemente grande, a distribuição amostral do estimador \bar{x} será normal. A expressão do intervalo de confiança para $\mu_{\mathbf{x}}$, ao nível de confiança $(1 - \beta)$, é dada por⁴

$$\bar{x} - z_{\beta/2} \frac{\sigma_{\mathbf{x}}}{\sqrt{N}} \leq \mu_{\mathbf{x}} \leq \bar{x} + z_{\beta/2} \frac{\sigma_{\mathbf{x}}}{\sqrt{N}}, \quad (3.2)$$

em que $z_{\beta/2}$ denota o quantil⁵ $q_{(1-\beta/2)}$ da distribuição normal padrão [58].

Definição 3.1 (Dependência de Longa Duração). \mathbf{x}_t é um processo com dependência de longa duração ou memória longa se existirem constantes α e C_P ,

⁴Dada uma probabilidade $1 - \beta$, encontra-se um valor $z_{\beta/2}$ tal que $P\{-z_{\beta/2} < Z < z_{\beta/2}\} = 1 - \beta$ ($z_{\beta/2} = 1,96$ para $1 - \beta = 95\%$).

⁵O quantil q_{α} de uma função de distribuição $F_{\mathbf{x}}$ é o valor para o qual se tem $F_{q_{\alpha}} = \alpha$ [83, pág.181]. A mediana, por exemplo, corresponde ao quantil $q_{0,5}$.

satisfazendo $0 < \alpha < 1$ e $C_P > 0$, tais que [58], [67], [84]

$$\lim_{f \rightarrow 0} \frac{P_x(f)}{C_P |f|^{-\alpha}} = 1, \quad (3.3)$$

em que $P_x(f)$ denota a DEP de \mathbf{x}_t e f representa a frequência normalizada ($-1/2 \leq f \leq 1/2$), em ciclos/amostra.

Portanto, a DEP de processos LRD tende a infinito na frequência zero (singularidade do tipo $1/f^\alpha$ na origem do espectro). Note-se que a definição 3.1 é assintótica, pois o formato da DEP em frequências afastadas da origem não é especificado. É mais comum caracterizar-se a memória longa pelo parâmetro H de Hurst:

$$H = \frac{\alpha + 1}{2}, \quad 1/2 < H < 1. \quad (3.4)$$

Quanto maior o valor de H , maior é o grau de memória longa do processo.

Uma definição alternativa pode ser dada no domínio do tempo, em termos da autocorrelação $R_x(\tau)$. \mathbf{x}_t é um processo do tipo $1/f^\alpha$ se a sua autocorrelação $R_x(\tau)$, para valores suficientemente grandes do *lag* τ , decresce segundo uma função potência (isto é, o decaimento para zero é extremamente lento e do tipo hiperbólico):

$$\lim_{\tau \rightarrow \infty} \frac{R_x(\tau)}{C_R \tau^{-(1-\alpha)}} = 1, \quad (3.5)$$

em que $C_R > 0$.

A singularidade na origem do espectro implica a validade da relação

$$\sum_{\tau=-\infty}^{\infty} R_x(\tau) = \infty, \quad (3.6)$$

para $1/2 < H < 1$, ou seja, as autocorrelações decaem para zero tão lentamente que não são somáveis. A Eq. (3.6) afirma que a dependência estatística entre eventos distantes diminui muito lentamente com o aumento do *lag* τ . Deste modo, a razão entre autocorrelações para valores suficientemente grandes do *lag* não se altera de modo apreciável. Este comportamento é radicalmente diferente do apresentado por um processo da classe ARMA [38],[39], em que a autocorrelação decresce rapidamente (de forma exponencial) para zero:

$$|R_x(\tau)| \leq C r^{-\tau}, \quad \tau = 1, 2, \dots, \quad (3.7)$$

em que $C > 0$ e $0 < r < 1$. Processos ARMA possuem dependência de curta duração.

Se \mathbf{x}_t for LRD, a variância de \bar{x} decresce com o tamanho N da amostra

mais lentamente do que no caso tradicional (variáveis independentes ou não-correlacionadas) e da seguinte maneira [58, pág.6]

$$\sigma_{\bar{x}}^2 \approx \sigma_x^2 c(\rho_x) N^{\alpha-1}, \quad (3.8)$$

em que $c(\rho_x)$ é definido por

$$\lim_{N \rightarrow \infty} N^{-(1+\alpha)} \sum_{i \neq j} \rho_x(i, j). \quad (3.9)$$

Neste caso, a distribuição de \bar{x} é assintoticamente Gaussiana, com $E[\bar{x}] = \mu_x$. Por outro lado, a variância da média amostral de séries ARMA e de Markov tem o seguinte comportamento assintótico [58, pág.5]:

$$\sigma_{\bar{x}}^2 \approx \sigma_x^2 c(\rho_x) N^{-1}. \quad (3.10)$$

O comportamento LRD de x_t faz com que a estimação de parâmetros como \bar{x} seja mais difícil do que quando as observações são não-correlacionadas. Neste caso, a equação do intervalo de confiança para μ_x (dada por (3.2)) não é aplicável. De fato, para um determinado nível de confiança $(1 - \beta)$, o intervalo de confiança dado deve ser “esticado” por meio da multiplicação do mesmo por um fator F dado por [58, pág.9]

$$F = N^{\alpha/2} \sqrt{c(\rho_x)}. \quad (3.11)$$

Observe-se que este fator de correção F cresce com N , sendo que no limite diverge para o infinito.

Pode-se adquirir uma certa intuição das propriedades de um processo LRD por meio da análise dos seus processos agregados. O processo agregado de \mathbf{x}_t de grau M , denotado por $X_t^{(M)}$, corresponde a uma média móvel de blocos não-sobrepostos de \mathbf{x}_t de tamanho M , ou seja,

$$X_i^{(M)} = \frac{1}{M} \sum_{t=M(i-1)+1}^{Mi} \mathbf{x}_t. \quad (3.12)$$

A seguinte propriedade é válida para um processo \mathbf{x}_t com memória longa [30]:

$$\lim_{M \rightarrow \infty} \frac{\text{Var } X_t^{(M)}}{M^{(2H-2)}} = c, \quad (3.13)$$

em que c é uma constante. Observe-se que a agregação equivale a uma operação de mudança de escala temporal.

Conforme visto anteriormente, a Fig. 3.6 ilustra que agregações sucessivas de um processo aleatório SRD (como o de Poisson) tende a suavizar o pro-

cesso agregado resultante depois de algumas mudanças de escala. Por outro lado, (3.13) afirma que o grau de suavidade da seqüência de processos agregados $\{X_t^{(M)}\}_{M=2,3,\dots}$, associados a um processo LRD \mathbf{x}_t , aumenta com o aumento de M muito mais lentamente do que no caso de processos SRD, em que a variância do processo agregado decai proporcionalmente a M^{-1} para $M \rightarrow \infty$. Sendo assim, depreende-se que o processo agregado é estatisticamente similar ao processo original, no sentido de que um número finito de agregações sucessivas não destrói o caráter impulsivo do processo original. Portanto, (3.13) sugere que as propriedades de dependência de longa duração e auto-similaridade (que será definida a seguir) estejam intimamente relacionadas.

3.3.2 Auto-Similaridade

Definição 3.2 (Processo H -ss). Um processo estocástico $\{\mathbf{y}_t\}_{t \in \mathbb{R}}$ é auto-similar com parâmetro $0 < H < 1$, ou seja, é H -ss (*self-similar with parameter H*) se, para qualquer $a > 0$,

$$\{\mathbf{y}(t)\} \stackrel{d}{=} \{a^{-H} \mathbf{y}(at)\}, \quad (3.14)$$

em que $\stackrel{d}{=}$ denota igualdade entre as distribuições finito-dimensionais [84].

Um processo H -ss é LRD se $1/2 < H < 1$. O processo Movimento Browniano (de tempo contínuo) [85], também conhecido como processo de Wiener, satisfaz a definição 3.2, sendo auto-similar com $H = 1/2$ (mas não é LRD). Se o processo $\mathbf{x}_t = \Delta \mathbf{y}_t$, denominado processo de incrementos de \mathbf{y}_t ou primeira diferença de \mathbf{y}_t , for estacionário, então \mathbf{y}_t é denominado H -sssi (*H self-similar with stationary increments* - auto-similar com incrementos estacionários). Neste caso, o processo H -sssi \mathbf{y}_t é um processo integrado de ordem 1, $\mathbf{y}_t \sim I(1)$.

Se os momentos de $\mathbf{y}(t)$ de ordem menor ou igual a q existirem, pode-se concluir a partir de (3.14) que [30]

$$E|\mathbf{y}_t|^q = E|\mathbf{y}_1|^q |t|^{qH}. \quad (3.15)$$

Portanto, o processo $\mathbf{y}_t \sim I(1)$ não pode ser estacionário.

Assumindo-se $E[\mathbf{y}_t] = 0$ com o intuito de simplificar a notação⁶, demonstra-se que a autocovariância de \mathbf{y}_t é dada por [58]

$$C_y(t, s) = E[\mathbf{y}_t \mathbf{y}_s] = \frac{\sigma_x^2}{2} [t^{2H} + s^{2H} - (t - s)^{2H}]. \quad (3.16)$$

em que $\sigma_x^2 = E[(\mathbf{y}_t - \mathbf{y}_{t-1})^2] = E[\mathbf{y}^2(1)]$ é a variância do processo de incrementos

⁶O que implica $E[\mathbf{x}_t] = E[\mathbf{y}_t - \mathbf{y}_{t-1}] = 0$

\mathbf{x}_t .

Considere a versão amostrada \mathbf{y}_t , $t \in \mathbb{Z}$, de um processo \mathbf{y}_t H -sssi, em que o período de amostragem é unitário. Existem vários processos não-Gaussianos \mathbf{y}_t H -sssi. Entretanto, para cada valor de $H \in (0, 1)$ há exatamente um único processo Gaussiano \mathbf{y}_t H -sssi, denominado movimento Browniano fracionário de tempo discreto (*Discrete-time Fractional Brownian Motion* (DFBM)) [58], [67]. O ruído Gaussiano fracionário (*Fractional Gaussian Noise* (FGN)), proposto por Mandelbrot e van Ness em 1968 [85], corresponde ao processo de incrementos do DFBM. O FGN é um modelo bastante utilizado em simulações de tráfego LRD [77], [50], [86] ⁷.

Definição 3.3 (Auto-Similaridade Exata de Segunda Ordem). Seja o processo estacionário de tempo discreto $\mathbf{x}_t = \mathbf{y}_t - \mathbf{y}_{t-1}$. \mathbf{x}_t é um processo exatamente auto-similar de segunda ordem com parâmetro de Hurst H ($1/2 < H < 1$) se a sua autocovariância existe e for dada por [58]

$$C_x(\tau) = \frac{\sigma_x^2}{2} [|\tau + 1|^{2H} - 2|\tau|^{2H} + |\tau - 1|^{2H}], \quad \tau = \dots, -1, 0, 1, \dots \quad (3.17)$$

Pode-se demonstrar que a autocovariância dada por (3.17) satisfaz [58]

$$\lim_{\tau \rightarrow \infty} \frac{C_x(\tau)}{\sigma_x^2 \tau^{2H-2} H(2H-1)} = 1, \quad (3.18)$$

ou seja, $C_x(\tau)$ tem um decaimento hiperbólico. Portanto, auto-similaridade de segunda ordem implica LRD quando $1/2 < H < 1$.

Considere o processo agregado $X_t^{(M)}$ de um processo exatamente auto-similar de segunda ordem \mathbf{x}_t , ao nível de agregação M . Demonstra-se que

$$C_x^{(M)}(\tau) = C_x(\tau), \quad M = 2, 3, \dots \quad (3.19)$$

A Eq. (3.19) diz que as estatísticas de segunda ordem do processo original não mudam com a mudança de escala, o que justifica o termo “exatamente auto-similar de segunda ordem”.

Definição 3.4 (Auto-Similaridade Assintótica de Segunda Ordem). Um processo \mathbf{x}_t é assintoticamente auto-similar de segunda ordem com parâmetro de Hurst H ($1/2 < H < 1$) se a sua autocovariância e a autocovariância do seu processo

⁷Os modelos DFBM e FGN são não-paramétricos e não são utilizados para se fazer previsão de valores futuros de tráfego. O processo FGN é definido no domínio da frequência a partir de uma prescrição específica de DEP [50], ao passo que o MWM utiliza a análise de multirresolução de Haar e é baseado numa cascata binomial multiplicativa no domínio *wavelet* (vide seção 4.2) [11].

agregado estão relacionadas por [30]

$$\lim_{M \rightarrow \infty} C_{\mathbf{x}}^{(M)}(\tau) = C_{\mathbf{x}}(\tau). \quad (3.20)$$

Tsybarov e Georganas [87] demonstram que (3.13) implica (3.20). Portanto, um processo LRD também é assintoticamente auto-similar de segunda ordem.

3.4 Impulsividade

Vários fenômenos exibem comportamento impulsivo, tais como ruído atmosférico de baixa frequência, ruído produzido pelo homem, ruído acústico submarino, transitórios em linhas de transmissão, atividade sísmica [85], [88], [89], [90], [91], séries financeiras [24], *clusters* de erros em circuitos telefônicos [92] e o tráfego em redes de computadores [2]. Neste contexto, a distribuição de probabilidades do tipo *stable* (estável) [93] é uma ferramenta básica de modelagem estatística de sinais impulsivos [55]. O uso das distribuições estáveis é justificado pelo teorema central do limite generalizado [54], o qual afirma que, se o limite da soma de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) converge, então este limite só pode ser uma variável aleatória com distribuição estável.

3.4.1 Distribuições Estáveis

Definição 3.5 (Distribuição de Cauda Pesada). Uma variável aleatória \mathbf{x} possui uma distribuição de cauda pesada com índice α se [51]

$$P(\mathbf{x} \geq x) \sim cx^{-\alpha}L(x), \quad x \rightarrow \infty, \quad (3.21)$$

para $c > 0$ e $0 < \alpha < 2$, em que $L(x)$ é uma função positiva que varia lentamente para grandes valores de x , ou seja, $\lim_{x \rightarrow \infty} L(bx)/L(x) = 1$ para qualquer b positivo.

A Eq. (3.21) afirma que observações de uma variável aleatória com distribuição de cauda pesada podem ocorrer, com probabilidades não-desprezíveis, com valores bastante distantes da média (*outliers*). Portanto, esse tipo de variável aleatória apresenta alta variabilidade.

Um exemplo simples de distribuição de cauda pesada é a distribuição de Pareto I [23],[94], que é definida em termos de sua função distribuição complementar

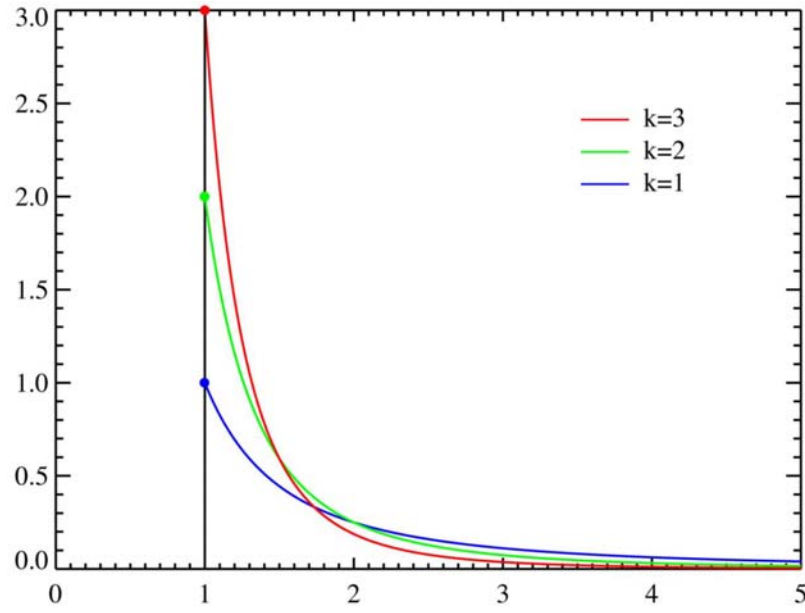


Figura 3.10: Funções densidade de probabilidade de Pareto I para vários valores de $k = \alpha$ com $x_m = 1$. Observe que a função decai lentamente para zero para valores grandes de x . A função tende para o impulso $\delta(x - x_m)$ quando $k \rightarrow \infty$.

de probabilidade (função de sobrevivência):

$$\bar{F}(x) = P(\mathbf{x} \geq x) = \begin{cases} \left(\frac{x}{x_m}\right)^{-\alpha}, & x \geq x_m, \\ 1, & x < x_m, \end{cases} \quad (3.22)$$

As Figs. 3.10 e 3.11 ilustram as funções densidade e distribuição de probabilidade Pareto I.

As estatísticas de ordem p das distribuições de cauda pesada são finitas se, e somente se, $p \leq \alpha$. É por essa razão que essas distribuições têm variância infinita (a média é infinita se $\alpha < 1$). As distribuições de cauda pesada são também conhecidas como distribuições de probabilidade com variância infinita.

Um membro importante da classe das distribuições de cauda pesada é a **distribuição estável**, descoberta por Lévy na década de 1920 [93]. A distribuição estável não possui expressão analítica⁸, mas pode ser definida em termos da sua função característica⁹. Uma variável aleatória \mathbf{x} estável é definida pela seguinte função característica [51]:

$$\Phi_{\mathbf{x}}(w) = E[e^{jw\mathbf{x}}] = \int_{-\infty}^{\infty} f_{\mathbf{x}}(x)e^{jwx} dx = \exp\{j\mu w - |\sigma w|^{\alpha}[1 - j\eta \text{sign}(w)\varphi(w, \alpha)]\}, \quad (3.23)$$

⁸Com exceção dos casos-limite $\alpha = 1$ (Cauchy) e $\alpha = 2$ (Gaussiana).

⁹Definida como a transformada de Fourier da função densidade de probabilidade[47].

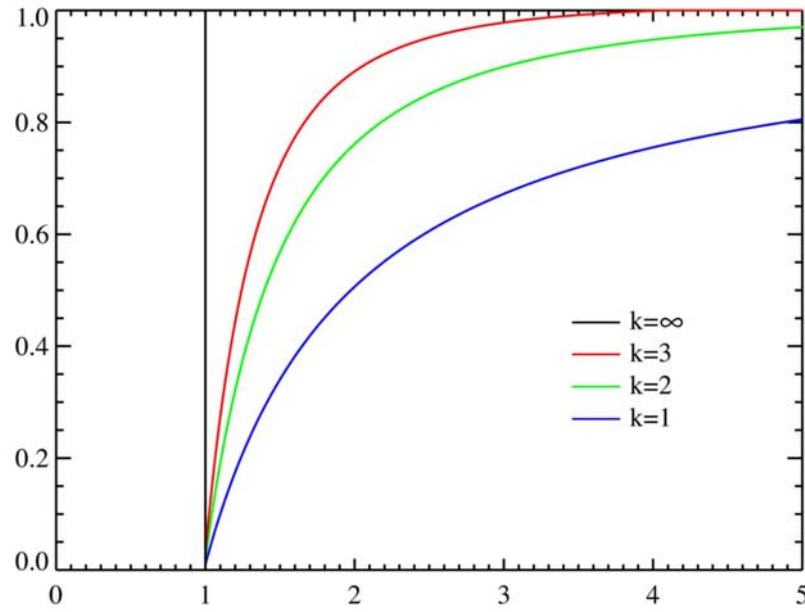


Figura 3.11: Funções distribuição de probabilidade de Pareto I para vários valores de $k = \alpha$ com $x_m = 1$.

em que,

$$\varphi(w, \alpha) = \begin{cases} \tan(\alpha\pi/2) & \text{se } \alpha \neq 1 \\ -\frac{2}{\pi} \ln |w| & \text{se } \alpha = 1, \end{cases} \quad (3.24)$$

e $\text{sign}(\cdot)$ denota a função sinal, α ($0 < \alpha \leq 2$) é o **expoente característico**, μ ($\mu \in \mathbb{R}$) é o **parâmetro de localização**, η ($-1 \leq \eta \leq 1$) é o **parâmetro de assimetria** e $\sigma \geq 0$ é o **parâmetro de dispersão** ou **escala**. A notação $S_\alpha(\sigma, \eta, \mu)$ denota a distribuição definida por (3.23). A Fig. 3.12 ilustra uma realização α -stable, simulada por meio da função `stablernd`, da biblioteca de funções `STABLE4.0` [95] para `MATLAB`. As Figs. 3.13, 3.14, 3.15 e 3.16 ilustram algumas funções densidade e distribuição de probabilidade de variáveis aleatórias estáveis.

O parâmetro α especifica o grau de impulsividade da variável aleatória (o peso da cauda da distribuição aumenta conforme α diminui). Para qualquer constante real a , $\mathbf{x} + a$ tem distribuição $S_\alpha(\sigma, \eta, \mu + a)$. Para $a > 0$ e $\alpha \neq 1$, $a\mathbf{x}$ tem distribuição $S_\alpha(a\sigma, \eta, a\mu)$. A distribuição é simétrica em relação a μ se $\eta = 0$. A distribuição Gaussiana é um caso particular da estável para $\alpha = 2$, pois, neste caso, tem-se que $\Phi_{\mathbf{x}}(w) = \exp(-\sigma^2 w^2 + j\mu w)$ corresponde à função característica de uma variável aleatória Gaussiana com média μ e variância $2\sigma^2$. Quando $\eta = \mu = 0$ diz-se que \mathbf{x} é do tipo estável simétrica (*symmetric stable* (S α S)) com $\Phi_{\mathbf{x}}(w) = \exp(-\sigma^\alpha |w|^\alpha)$.

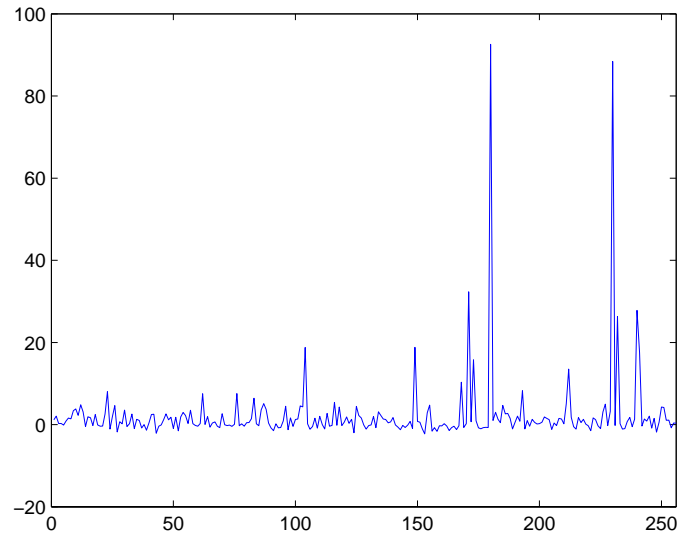


Figura 3.12: Simulação de uma realização com distribuição $S_{1,2}(1, 1, 0)$ (256 amostras).

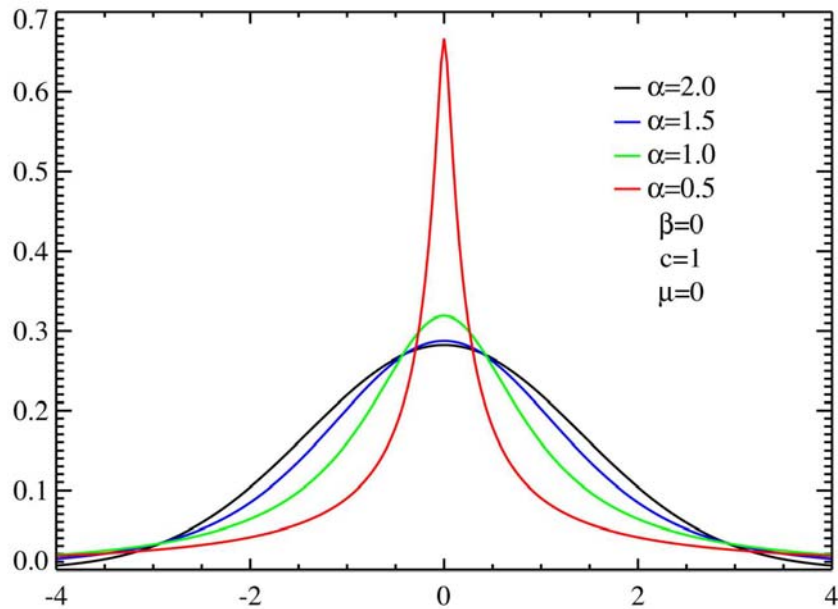


Figura 3.13: Funções densidade de probabilidade estáveis simétricas centradas em zero [7]. Os parâmetros β e c correspondem aos parâmetros η e σ , respectivamente, de (3.23).

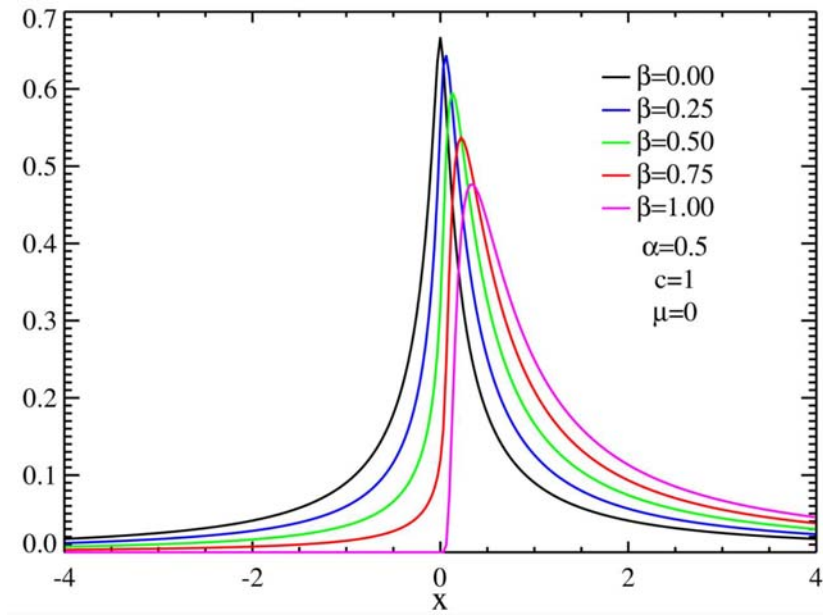


Figura 3.14: Funções densidade de probabilidade estáveis assimétricas [7].

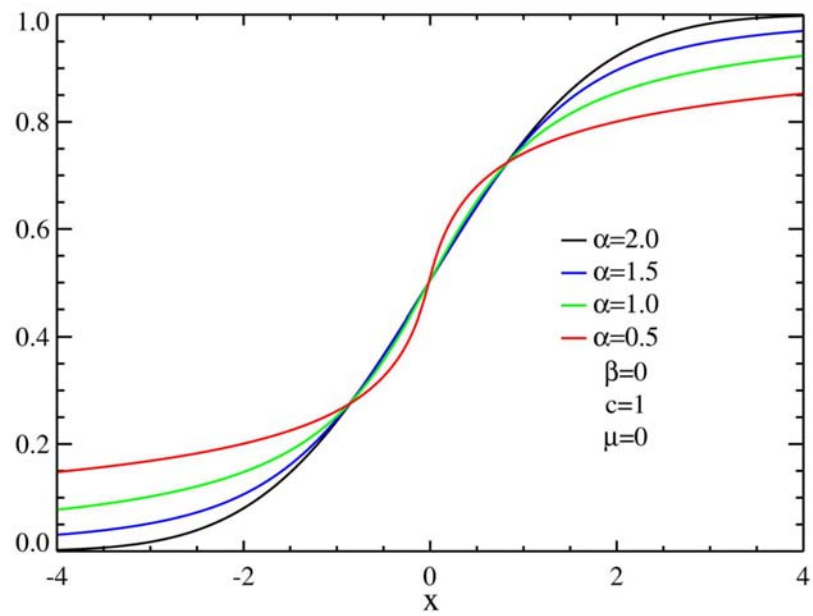


Figura 3.15: Funções distribuição de probabilidade simétricas [7].

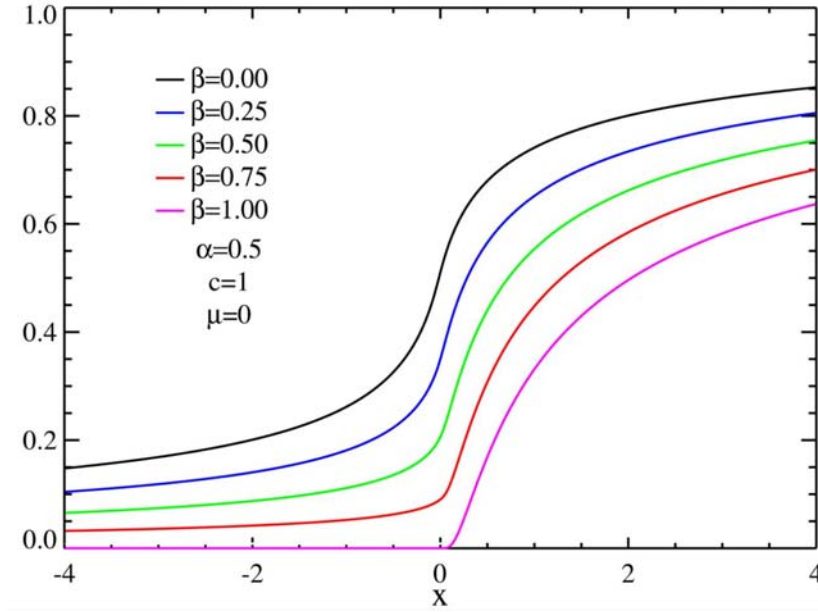


Figura 3.16: Funções distribuição de probabilidade assimétricas [7].

Demonstra-se que, se $\mathbf{x} \sim S_\alpha(\sigma, \eta, \mu)$ com $0 < \alpha < 2$, então,

$$\begin{cases} \lim_{\lambda \rightarrow \infty} \frac{P(\mathbf{x} > \lambda)}{\lambda^{-\alpha}} = C_\alpha \frac{1+\eta}{2} \sigma^\alpha, \\ \lim_{\lambda \rightarrow \infty} \frac{P(\mathbf{x} < -\lambda)}{\lambda^{-\alpha}} = C_\alpha \frac{1-\eta}{2} \sigma^\alpha, \end{cases} \quad (3.25)$$

em que

$$C_\alpha = \left(\int_0^\infty x^{-\alpha} \sin x \, dx \right)^{-1} = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha) \cos(\pi\alpha/2)} & \text{se } \alpha \neq 1 \\ 2/\pi & \text{se } \alpha = 1. \end{cases} \quad (3.26)$$

Portanto, (3.25) mostra que a função de sobrevivência de \mathbf{x} decresce segundo uma função potência para grandes valores de λ .

Distribuições estáveis possuem duas propriedades importantes: estabilidade e o teorema central do limite generalizado [55], [54], que serão enunciadas a seguir.

Teorema 3.1 (Propriedade de Estabilidade). Uma variável aleatória \mathbf{x} é estável se e somente se para quaisquer variáveis aleatórias independentes \mathbf{x}_1 e \mathbf{x}_2 que tenham a mesma distribuição de \mathbf{x} , e para constantes arbitrárias a_1, a_2 , existem constantes a e b tais que

$$a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 \stackrel{d}{=} a \mathbf{x} + b, \quad (3.27)$$

em que $\stackrel{d}{=}$ denota igualdade de distribuição de probabilidades [51].

Utilizando-se a função característica da distribuição estável, pode-se demonstrar uma propriedade mais geral: se $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ são independentes e seguem distribuições estáveis de mesmo (α, η) , então todas as combinações lineares da

forma $\sum_{j=1}^N a_j \mathbf{x}_j$ são estáveis com os mesmos parâmetros α e η .

O teorema central do limite diz que a soma normalizada de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com variância σ^2 e média μ finitas converge para uma distribuição Gaussiana. Formalmente, o teorema afirma que

$$\frac{\bar{\mathbf{x}} - \mu}{\sigma/\sqrt{N}} \xrightarrow{d} \mathbf{x} \sim N(0, 1) \quad \text{para } N \rightarrow \infty. \quad (3.28)$$

A relação (3.28) pode ser reescrita como

$$a_N(\mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_N) - b_N \xrightarrow{d} \mathbf{x} \sim N(0, 1) \quad \text{para } \mathbf{x} \rightarrow \infty, \quad (3.29)$$

em que $a_N = 1/(\sigma\sqrt{N})$ e $b_N = \sqrt{N}\mu/\sigma$.

Teorema 3.2 (Teorema Central do Limite Generalizado). Seja $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \dots\}$ uma seqüência de variáveis aleatórias i.i.d. Há constantes $a_N > 0$, $b_N \in \mathbb{R}$ e uma variável aleatória \mathbf{x} com

$$a_N(\mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_N) - b_N \xrightarrow{d} \mathbf{x}$$

se e somente se \mathbf{x} é α -stable com $0 < \alpha \leq 2$ [51].

Pode-se mostrar que o teorema (3.2) é consequência do teorema (3.1). O teorema central do limite generalizado afirma que se a premissa de variância finita for descartada, então o único limite possível é a distribuição estável.

A definição de distribuição estável pode se estendida para vetores aleatórios [30], [51], [55]. O vetor $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ é um vetor aleatório SaS se, e somente se, a combinação linear $a_1 \mathbf{x}_1 + \dots + a_N \mathbf{x}_N$ é SaS para quaisquer a_1, \dots, a_N reais. Um processo aleatório $\{\mathbf{x}_t\}_{t \in \mathbb{Z}}$ é dito α -stable se para qualquer $N \geq 1$ e índices distintos t_1, \dots, t_N , as variáveis aleatórias $\mathbf{x}_{t_1}, \dots, \mathbf{x}_{t_N}$ são conjuntamente α -stable com o mesmo valor de α .

Definição 3.6 (Processo Estocástico Impulsivo). Um processo estocástico \mathbf{x}_t é impulsivo se a sua distribuição marginal de probabilidades possuir cauda pesada [51].

3.4.2 Medidas de Dependência para Processos Impulsivos

Covariâncias (ou correlações) não podem ser definidas no espaço de variáveis aleatórias estáveis, dado que a variância de uma variável aleatória estável é infinita. Isto não quer dizer que algum tipo de medida de dependência não possa

ser definida. De fato, dois tipos de medidas foram propostos na literatura: a **covariação** e a **codiferença**. A medida mais utilizada na prática é a codiferença, a qual será definida a seguir. O conceito de covariação não será abordado neste trabalho (a referência [55] traz uma exposição detalhada da covariação).

A codiferença de duas variáveis aleatórias conjuntamente SαS, $0 < \alpha \leq 2$, \mathbf{x}_1 e \mathbf{x}_2 é dada por

$$\gamma_{\mathbf{x}_1, \mathbf{x}_2} = (\sigma_{\mathbf{x}_1})^\alpha + (\sigma_{\mathbf{x}_2})^\alpha - (\sigma_{\mathbf{x}_1 - \mathbf{x}_2})^\alpha, \quad (3.30)$$

em que $\sigma_{\mathbf{x}}$ é o parâmetro de escala da variável SαS \mathbf{x} . A codiferença é simétrica, ou seja, $\gamma_{\mathbf{x}_1, \mathbf{x}_2} = \gamma_{\mathbf{x}_2, \mathbf{x}_1}$ e reduz-se à covariância quando $\alpha = 2$. Se \mathbf{x}_1 e \mathbf{x}_2 são independentes, então $\gamma_{\mathbf{x}_1, \mathbf{x}_2} = 0$.

A **codiferença generalizada** [96] é uma extensão da codiferença e é definida para processos com distribuições marginais de cauda pesada genéricas.

Definição 3.7 (Codiferença Generalizada).

$$I(w_1, w_2; \mathbf{x}_1, \mathbf{x}_2) = -\ln E[e^{j(w_1 \mathbf{x}_1 + w_2 \mathbf{x}_2)}] + \ln E[e^{jw_1 \mathbf{x}_1}] + \ln E[e^{jw_2 \mathbf{x}_2}], \quad (w_1, w_2) \in \mathbb{R}^2 \quad (3.31)$$

Se \mathbf{x}_1 e \mathbf{x}_2 são independentes, então $I(w_1, w_2; \mathbf{x}_1, \mathbf{x}_2) = 0$. Para variáveis aleatórias conjuntamente Gaussianas vale

$$I(w_1, w_2; \mathbf{x}_1, \mathbf{x}_2) = -w_1 w_2 C(\mathbf{x}_1, \mathbf{x}_2),$$

em que $C(\mathbf{x}_1, \mathbf{x}_2)$ denota a covariância entre \mathbf{x}_1 e \mathbf{x}_2 . Para o caso dos processos aleatórios estacionários tem-se que

$$I(w_1, w_2; \tau) = I(w_1, w_2; \mathbf{x}_{t+\tau}, \mathbf{x}_t).$$

Definição 3.8 (Memória Longa em Sentido Generalizado). Seja $\{\mathbf{x}_t\}_{t \in \mathbb{R}}$ um processo estacionário. Diz-se que \mathbf{x}_t possui **memória longa no sentido generalizado**, se a sua codiferença generalizada, $I(w_1, w_2; \tau)|_{w_1=-w_2=1}$ satisfaz

$$\lim_{\tau \rightarrow \infty} I(1, -1; \tau)/\tau^{-\beta} = L(\tau), \quad (3.32)$$

em que $L(\tau)$ é uma função que varia lentamente para $\tau \rightarrow \infty$ e $0 < \beta < 1$.

Note-se que para processos gaussianos LRD, a definição acima reduz-se à definição clássica de processo LRD (vide (3.5))¹⁰.

¹⁰Os parâmetros H , α (não se trata do expoente característico de uma distribuição estável, mas do expoente de escala de um processo auto-similar) e β estão relacionados entre si pelas seguintes relações: $\alpha = 2H - 1$ e $\beta = 2 - 2H$.

3.5 Por que o tráfego das redes de dados é fractal?

A investigação das causas do comportamento fractal do teletráfego não faz parte do escopo desta pesquisa, conforme delineado na seção 1.2 desta tese. Não obstante, o autor deste trabalho seria omissos se não fizesse alguns breves comentários sobre essa importante questão, que, como será apontado a seguir, tem sido pesquisada desde o artigo seminal de Leland *et al* [1].

Vários autores [1], [61], [73], [97], [98], [99] têm afirmado que a auto-similaridade do tráfego Internet seria causada pela grande variabilidade do tamanho (em *bytes* ou pacotes) das sessões individuais (*File Transfer Protocol* (FTP), *Hyper Text Transfer Protocol* (HTTP), etc.) que compõem o tráfego agregado. Mais especificamente, aqueles artigos conjecturam que o tráfego IP é auto-similar porque os tamanhos das sessões individuais que compõem o tráfego Internet são gerados por uma distribuição de probabilidades de cauda pesada. Por outro lado, o recente estudo de Gong *et al* [100] reexamina esta questão¹¹ e advoga que há pouca evidência de que a cauda pesada da distribuição tenha algum impacto sobre o projeto dos algoritmos e a infraestrutura da Internet. Gong *et al* propõem um modelo hierárquico *on-off* Markoviano que explica a LRD do tráfego IP e sustentam que as múltiplas escalas de tempo envolvidas no mecanismo de geração do tráfego¹² e os protocolos de transporte (como o TCP) fazem com que a observação da LRD seja inevitável.

¹¹Dentre outros argumentos apresentados, esses autores afirmam que “a inferência estatística de dados provenientes de uma distribuição de cauda pesada a partir de uma amostra finita de dados é extremamente difícil, senão impossível”.

¹²Por exemplo, um protocolo de transporte como o TCP atua numa escala de tempo de milissegundos, ao passo que eventos associados à camada de aplicação (*browsing*, *downloading*, etc.) acontecem numa escala de tempo da ordem de segundos a minutos.

4 Modelagem de Teletráfego com Memória Longa

Este capítulo apresenta os modelos FGN, MWM e ARFIMA, de grande importância em hidrologia, economia, finanças, telecomunicações e outras áreas. O modelo ARFIMA é adequado para esquemas de previsão de teletráfego com LRD porque é paramétrico. Os processos FGN e MWM são não-paramétricos e têm sido amplamente utilizados para simulação/geração de teletráfego [101], [50], [74], [77], [86], [102], [103].

4.1 Introdução

Em geral, modelos de tráfego podem ser classificados como heterogêneos ou homogêneos. Modelos heterogêneos simulam o tráfego agregado (tráfego gerado por vários usuários, aplicações e protocolos) num enlace da rede. Por outro lado, modelos homogêneos referem-se a um tipo específico de tráfego, tal como tráfego de vídeo *Moving Picture Experts Group* (MPEG) [104], [105].

Modelos heterogêneos podem ser subdivididos em duas classes: comportamentais ou estruturais. Modelos comportamentais modelam as estatísticas do tráfego, tais como correlação, distribuição marginal ou até mesmo estatísticas de ordem mais alta (terceira e quarta ordens, por exemplo), sem levar em conta o mecanismo físico de geração de tráfego (ou seja, parâmetros de modelos comportamentais não estão diretamente relacionados aos parâmetros da rede de comunicação) [50], [11], [85], [80], [106]. Por outro lado, modelos estruturais [33], [98], [107], estão relacionados aos mecanismos de geração de pacotes e seus parâmetros podem ser mapeados para parâmetros da rede, tais como número de usuários e banda passante. Observe que os modelos FGN, ARFIMA e MWM são modelos comportamentais de tráfego agregado e que processos do tipo *On/Off*¹ podem ser

¹Modelos deste tipo assumem que uma determinada fonte de tráfego alterne os estados *On*, em que há um fluxo de dados entre o remetente e o destinatário, e *Off*, período de silêncio, em que nenhuma informação é transmitida.

utilizados como modelos estruturais de tráfego (veja [30] para maiores detalhes).

4.2 Modelagem Não-Paramétrica com FGN e MWM

4.2.1 Processo FGN

Em termos históricos, o processo FGN, proposto por Mandelbrot e Van Ness em 1968 [85] para modelagem de séries hidrológicas LRD, é o primeiro modelo importante de memória longa que aparece na literatura. Por definição, se $\{\mathbf{x}_t\}_{t \in \mathbb{Z}}$ é um FGN, então \mathbf{x}_t é um processo estacionário com autocovariância dada por (3.17) (ou seja, $C_{\mathbf{x}}(\tau) = \frac{\sigma_{\mathbf{x}}^2}{2}[|\tau+1|^{2H} - 2|\tau|^{2H} + |\tau-1|^{2H}]$, $\tau = \dots, -1, 0, 1, \dots$). O FGN corresponde à primeira diferença de um processo estocástico de tempo contínuo conhecido como movimento Browniano fracionário (*Fractional Brownian Motion* (FBM)) $\{\mathbf{B}_H(t) : 0 \leq t \leq \infty\}$ com parâmetro de Hurst $0 < H < 1$ [67, pág.279], ou seja,

$$\mathbf{x}_t = \Delta \mathbf{B}_H(t) = \mathbf{B}_H(t+1) - \mathbf{B}_H(t), \quad t = 0, 1, 2, \dots \quad (4.1)$$

Beran fornece uma definição formal do processo FBM em [58]. O FBM tem uma denominação especial quando $H = 1/2$: movimento Browniano (*Brownian motion*) e é designado por $\mathbf{B}_{1/2}(t)$. Neste caso, $\mathbf{x}_1, \mathbf{x}_2, \dots$ são variáveis aleatórias Gaussianas independentes. A Fig. 4.1 ilustra realizações de processos FBM para vários valores do parâmetro de Hurst.

Pode-se criar um FBM de tempo discreto (DFBM), denotado por \mathbf{B}_t , por meio da soma cumulativa de amostras do FGN $\{\mathbf{x}_t\}$:

$$\mathbf{B}_t \equiv \mathbf{B}_H(t) = \sum_{u=0}^{t-1} \mathbf{x}_u, \quad t = 1, 2, \dots \quad (4.2)$$

A DEP do DFBM é dada pela fórmula [67, pág. 280]

$$P_{\mathbf{B}_t}(f) = \sigma_x^2 C_H \sum_{j=-\infty}^{\infty} \frac{1}{|f+j|^{2H+1}}, \quad -\frac{1}{2} \leq f \leq \frac{1}{2}, \quad (4.3)$$

em que σ_x^2 é a potência de um FGN de média nula, $C_H = \frac{\Gamma(2H+1) \sin(\pi H)}{2\pi^{2H+1}}$ e $0 < H < 1$. De acordo com (4.3), a DEP do DFBM possui uma singularidade do tipo $|f|^{-\alpha}$, $0 < \alpha < 1$, na origem, pois

$$P_{\mathbf{B}_t} \propto |f|^{1-2H}, \quad f \rightarrow 0. \quad (4.4)$$

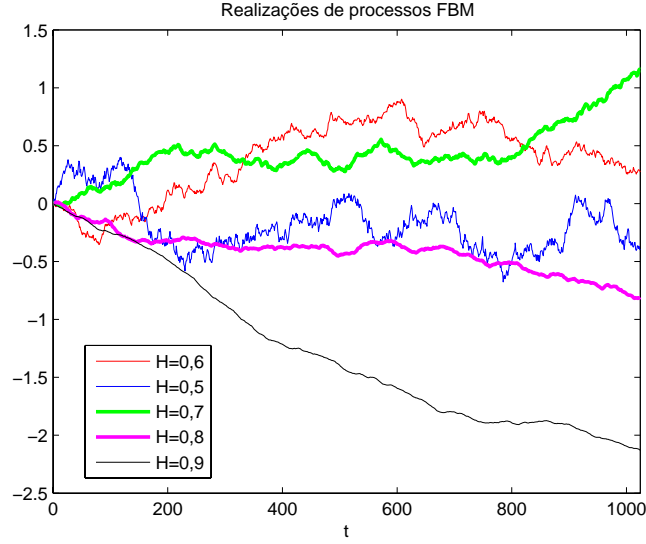


Figura 4.1: Realizações de processos FBM para vários valores do parâmetro de Hurst. As simulações foram feitas com o código MATLAB de geração e identificação de processos FBM desenvolvido por Coeurjolly [8],[9].

O FGN e o DFBM estão relacionados pela função de transferência

$$H(z) = \frac{X(z)}{B(z)} = 1 - z^{-1}, \quad (4.5)$$

em que $X(z)$ e $B(z)$ denotam as transformadas z de x_t e B_t , respectivamente. A resposta em frequência associada a (4.5) é

$$H(f) = H(z)|_{z=e^{j2\pi f}} = 1 - e^{-j2\pi f}. \quad (4.6)$$

Como a relação saída/entrada em termos das DEPs é igual a [47, pág. 351]

$$P_x(f) = |H(f)|^2 P_{B_t}(f), \quad (4.7)$$

em que $|H(f)|^2$ é dado por,

$$|H(f)|^2 = G(f) = 4 \sin^2(\pi f), \quad (4.8)$$

então a DEP do FGN é igual a

$$P_x(f) = 4 \sin^2(\pi f) P_{B_t}(f). \quad (4.9)$$

Assim, (4.3) e (4.9) mostram que a DEP do FGN é caracterizada por somente dois parâmetros: σ_x^2 e H (responsável pela forma do espectro). Além disso, é importante se ter em mente que o FGN é completamente especificado pela sua média e pela sua DEP, pois é Gaussiano.

Em [50], é mostrado que (4.9) pode ser reescrita na forma:

$$P_x(f) = A(f, H)[|2\pi f|^{-2H-1} + B(f, H)], \quad (4.10)$$

em que $A(f, H) = 2 \sin(\pi H) \Gamma(2H + 1) (1 - \cos(2\pi f))$ e $B(f, H) = \sum_{j=1}^{\infty} [(2\pi j + 2\pi f)^{-2H-1} + (2\pi j - 2\pi f)^{-2H-1}]$. Para pequenos valores de f tem-se que $P_x(f) \propto |f|^{1-2H}$.

4.2.2 Transformada *Wavelet*

4.2.2.1 A Transformada *Wavelet* Contínua

A transformada de Fourier de um sinal $x(t)$, caso exista, é definida como

$$X(\nu) = \text{TF}\{x(t)\} = \int_{-\infty}^{\infty} x(t) e^{-j2\pi\nu t} dt, \quad (4.11)$$

em que ν denota a frequência em ciclos/segundo [Hz].

Gabor [108] mostrou que é possível representar o conteúdo espectral **local** de um sinal $x(t)$ em torno de um instante de tempo τ através da transformada janelada de Fourier (*Windowed Fourier Transform* (WFT))

$$X_T(\nu, \tau) = \int_{-\infty}^{\infty} x(t) g_T(t - \tau) e^{-j2\pi\nu t} dt, \quad (4.12)$$

em que $g_T(t)$ é uma janela de suporte finito de duração T e ν denota frequência. A função janela é atrasada no tempo e depois transladada na frequência (modulada no tempo). A WFT é uma representação bidimensional definida no plano ou domínio tempo-frequência porque depende dos parâmetros ν e τ . A WFT equivaleria a uma descrição do tipo “partitura” musical contínua de $x(t)$.

De acordo com o princípio da incerteza de Heisenberg [109, pág.52], um sinal cujo conteúdo de energia esteja bem localizado no tempo tem esta energia bastante espalhada no domínio da frequência. Como a janela de 4.12 tem um certo tamanho T fixo, conclui-se que a WFT não é boa para analisar (ou identificar) comportamentos de $x(t)$ que aconteçam em tempos muito menores ou muito maiores do que T , como, por exemplo, fenômenos transitórios de duração $\Delta t \ll T$ ou ciclos que existam em períodos maiores do que T . A transformada *wavelet* resolve este problema substituindo a modulação pela **mudança de escala**, conforme será descrito a seguir.

Uma *wavelet* $\psi_0(t)$ (também chamada às vezes de “*wavelet* mãe”), $t \in \mathbb{R}$, é uma função que satisfaz três condições [67], [110].

1. A sua transformada de Fourier $\Psi(\nu)$, $-\infty < \nu < \infty$, é tal que existe uma constante finita C_ψ que obedece à **condição de admissibilidade**

$$0 < C_\psi = \int_0^\infty \frac{|\Psi(\nu)|^2}{\nu} d\nu < \infty. \quad (4.13)$$

2. A integral de $\psi_0(t)$ é nula:

$$\int_{-\infty}^\infty \psi_0(t) dt = 0. \quad (4.14)$$

3. A sua energia é unitária:

$$\int_{-\infty}^\infty |\psi_0(t)|^2 dt = 1. \quad (4.15)$$

A condição de admissibilidade (1) impõe que $\Psi(0) = 0$ (tenha valor DC nulo) e que o decaimento de $|\Psi(\nu)|^2$ para $\nu \rightarrow \infty$ seja pelo menos tão rápido quanto o de uma função $\text{sinc}(\nu) \equiv \sin(\pi\nu)/(\pi\nu)$, que é do tipo $1/\nu$ (vide demonstração em [111]). A condição (2) garante que $\Psi(0) = 0$ e impõe que $\psi_0(t)$ tenha cruzamentos por zeros (não necessariamente equiespaçados) de tal modo que excursões positivas da função sejam compensadas por excursões negativas. A condição (3) exige que o suporte efetivo de $\psi_0(t)$ seja finito², o que faz com que a *wavelet* seja uma função localizada no tempo. Sendo assim, $\psi_0(t)$ deve se parecer com uma “pequena onda” ou onda de curta duração (este é o significado do termo *wavelet*). A função $\sin(t)$ (senóide), por exemplo, não é uma *wavelet*, porque oscila entre 1 e -1 para $-\infty \leq t \leq \infty$ (o seno então seria um exemplo de “onda grande” ou onda de duração infinita).

As Figs. 4.2, 4.3 e 4.4 ilustram alguns exemplos de *wavelets*. A *wavelet* de Haar (vide canto superior esquerdo da Fig. 4.2) é o exemplo mais antigo (e também o mais simples) de *wavelet*³.

A transformada *wavelet* foi originalmente desenvolvida como uma ferramenta de análise e síntese de sinais de energia de tempo contínuo⁴ [112], [113], [114], [115], [116], [117]. Um sinal de energia $x(t)$, $t \in \mathbb{R}$ (t denota tempo), obedece à restrição

$$\|x\|^2 = \langle x, x \rangle \equiv \int_{-\infty}^\infty |x(t)|^2 dt < \infty, \quad (4.16)$$

ou seja, $x(t)$ que obedece à restrição (4.16) pertence ao espaço das funções de

²Ou seja, grande parte da energia da função está concentrada num determinado intervalo de tempo.

³O matemático húngaro Alfred Haar propôs esta *wavelet* em 1909 na sua tese de doutoramento, que foi supervisionada por Hilbert, sobre sistemas ortogonais de funções.

⁴As *wavelets* foram introduzidas por Grossmann e Morlet [112].

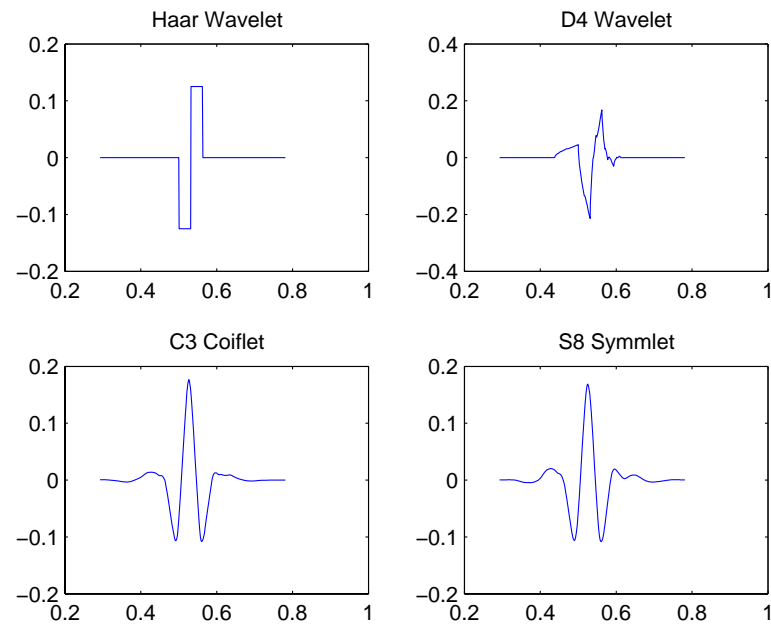


Figura 4.2: Quatro exemplos de funções *wavelet*. Esta figura foi gerada pela função `toon0111.m` do *toolbox* WaveLab850 para MATLAB [10].

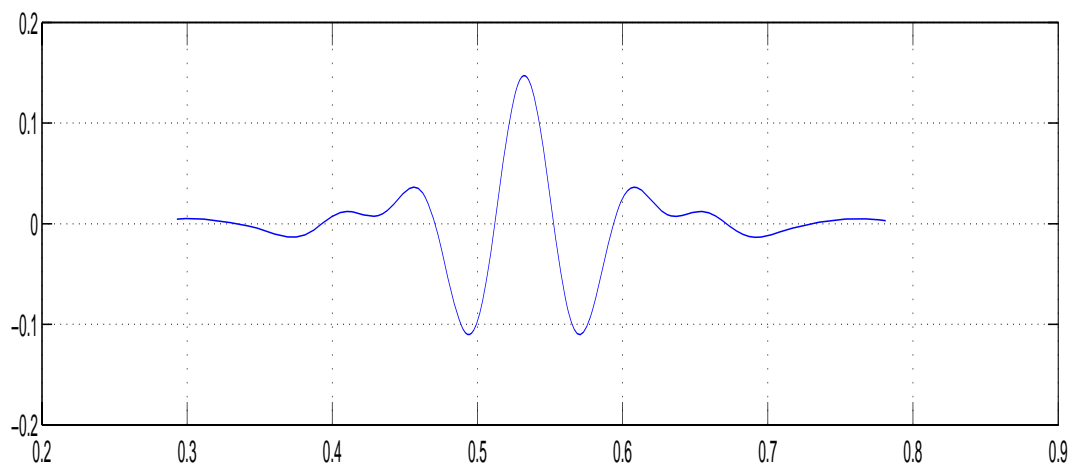
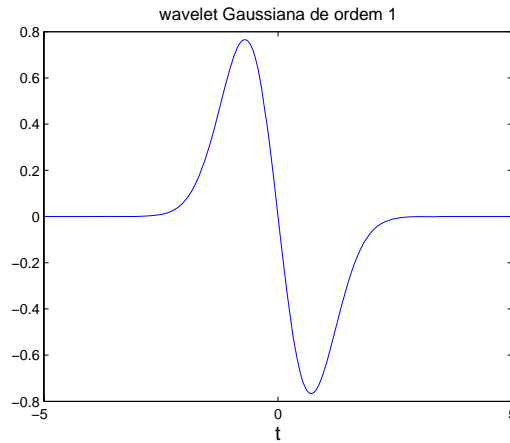
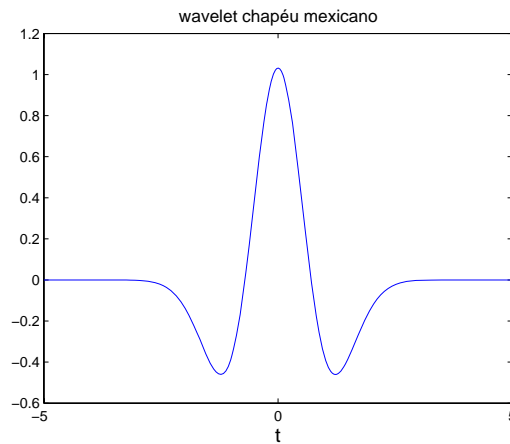


Figura 4.3: *Wavelet* de Meyer. Esta figura foi gerada pela função WaveLab850 `toon0114.m`.



(a)



(b)

Figura 4.4: (a): *wavelet* Gaussiana (relacionada à primeira derivada da PDF Gaussiana); (b): *wavelet* “chapéu mexicano” (relacionada à segunda derivada da PDF Gaussiana). Estas *wavelets* foram geradas pela função `gauswavf.m` do *toolbox wavelet* para MATLAB.

quadrado integrável⁵ $L^2(\mathbb{R})$. Atualmente, a transformada *wavelet* também tem sido utilizada como uma ferramenta de análise de sinais de tempo discreto.

A Fig. 4.6 ilustra uma taxonomia da transformada *wavelet*. Observe que há decomposições *wavelet* em tempo contínuo (*Continuous Wavelet Transform* (CWT)) e em tempo discreto. A CWT de um sinal $x(t)$ consiste num conjunto $C = \{W_\psi(s, \tau), s \in \mathbb{R}^+, \tau \in \mathbb{R}\}$, em que τ é o parâmetro de localização no tempo, s representa a escala e ψ denota uma função *wavelet*, de coeficientes *wavelet* no plano tempo-escala (também conhecido como plano tempo-frequência) contínuo dados por⁶

$$W_\psi(s, \tau) = \left\langle \psi_{0(s, \tau)}, x \right\rangle = \int_{-\infty}^{\infty} \frac{1}{\sqrt{s}} \psi_0^* \left(\frac{\lambda - \tau}{s} \right) x(\lambda) d\lambda, \quad (4.17)$$

em que $\psi_{0(s, \tau)}(t) = s^{-1/2} \psi_0 \left(\frac{t - \tau}{s} \right)$ denota uma versão dilatada e deslocada da *wavelet* “mãe” $\psi_0(t)$. O fator $1/\sqrt{s}$ em (4.17) é usado para que todas as funções da classe

$$\mathcal{W} = \left\{ \frac{1}{\sqrt{s}} \psi_0 \left(\frac{t - \tau}{s} \right) \in \mathbb{R} \right\} \quad (4.18)$$

tenham a mesma energia (norma).

Note que a idéia básica da CWT definida por (4.17) é correlacionar⁷ um sinal $x(t)$ com versões transladadas (por τ) e dilatadas (por s) de uma *wavelet* mãe (que tem um espectro do tipo passa-bandas). Como mencionado acima, a CWT é uma função de dois parâmetros. Ela é, portanto, uma transformada redundante, pois consiste no mapeamento de um sinal unidimensional sobre o plano tempo-escala.

Diferentemente da WFT, onde a reconstrução é feita a partir da mesma família de funções que foi usada na análise, na CWT a síntese é feita com funções $\tilde{\psi}_{s, \tau}$ que devem satisfazer

$$\tilde{\psi}_{s, \tau}(t) = \frac{1}{C_\psi} \frac{1}{s^2} \psi_{s, \tau}(t). \quad (4.19)$$

Sendo assim, $x(t)$ é recuperado completamente via transformada *wavelet* contínua

⁵O espaço $L^2(\mathbb{R})$ é um exemplo de **espaço de Hilbert**. Diz-se que um conjunto \mathcal{H} é um espaço de Hilbert se (i) \mathcal{H} é um espaço vetorial completo (no sentido de que toda sequência de Cauchy $\{x_n\}_{n \geq 1}$ converge em norma para algum elemento $x \in \mathcal{H}$) em \mathbb{C} (ou em \mathbb{R}) e se (ii) \mathcal{H} é equipado com uma operação de produto interno [40, pág.46], [118, pág.161]. Observe que $\{x_n\}_{n \geq 1}$ é convergente se e somente se $\{x_n\}_{n \geq 1}$ for uma sequência de Cauchy (condição necessária e suficiente) [119].

⁶Neste trabalho, o produto interno entre as funções $f(\cdot)$ e $g(\cdot)$ que pertencem a um espaço de funções definidas num domínio \mathcal{D} é dado por: $\langle f, g \rangle = \int_{\mathcal{D}} w(x) f^*(x) g(x) dx$, em que $w(x)$ denota uma função não negativa arbitrária.

⁷Medir a semelhança.

inversa (*Inverse Continuous Wavelet Transform* (ICWT)):

$$x(t) = \frac{1}{C_\psi} \int_0^\infty \left[\int_{-\infty}^\infty W_\psi(s, \tau) \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) d\tau \right] \frac{ds}{s^2}. \quad (4.20)$$

A diferença fundamental entre a CWT e a WFT reside no fato das funções $\psi_{s,\tau}$ sofrerem dilatações⁸ e compressões. A análise em escalas refinadas de tempo (pequenos valores de s) requer funções $\psi_{s,\tau}$ “rápidas”, isto é, de pequeno suporte, enquanto que a análise em escalas agregadas de tempo (valores elevados de s) requer funções $\psi_{s,\tau}$ mais “lentas”, isto é, de suporte mais largo. Conforme foi mencionado anteriormente, o produto interno definido por 4.17 é uma medida da semelhança entre a *wavelet* $\psi\left(\frac{t-\tau}{s}\right)$ e o sinal $x(t)$ num certo instante de tempo τ e numa determinada escala s . Para um τ fixo, grandes valores de s correspondem a uma análise em baixas frequências, ao passo que pequenos valores de s estão associados a uma análise em altas frequências. Portanto, a transformada *wavelet* possui uma **resolução temporal variável** (isto é, capacidade para analisar o sinal de perto - “*zoom in*” - ou de longe - “*zoom out*”), sendo adequada para o estudo de fenômenos que acontecem em várias escalas de tempo.

A Fig. 4.5 mostra a CWT de um sinal que é regular durante a primeira metade da sua duração e que contém singularidades em quase todos os pontos da sua segunda metade. Quando a escala decresce, a CWT decai rapidamente para zero nas regiões em que o sinal é regular. As singularidades isoladas na parte esquerda da figura produzem coeficientes de grandes valores em seus respectivos cones de influência, que convergem para as localizações das singularidades.

Segundo Kaiser [109], a CWT e a WFT são casos especiais de um método mais geral de análise e reconstrução de sinais, denominado teoria dos *frames* ou dos arcabouços em maior generalidade. O uso de *frames* na descrição de sinais é uma alternativa ao uso de *bases*. Enquanto as bases representam um número mínimo de vetores necessários para representar um vetor (sinal) qualquer, os *frames* são conjuntos com mais vetores que o mínimo necessário (ou seja, são redundantes). A família de funções $\psi_{s,\tau}$ é linearmente dependente porque cada vetor do *frame* (que é infinito-dimensional) pode ser decomposto como uma superposição linear contínua dos outros vetores⁹. A eq. 4.20 é válida porque o operador síntese \mathbf{S} da ICWT foi definida da forma

$$\mathbf{S} = (\mathbf{T}^* \mathbf{T})^{-1} \mathbf{T}^* \quad (4.21)$$

⁸ *Dilations*, em inglês. O termo “dilação” tem o significado de dilatação.

⁹ Observe-se que um *frame* contínuo é definido num *espaço linear de funções* ou *espaço funcional*, que tem dimensão infinita.

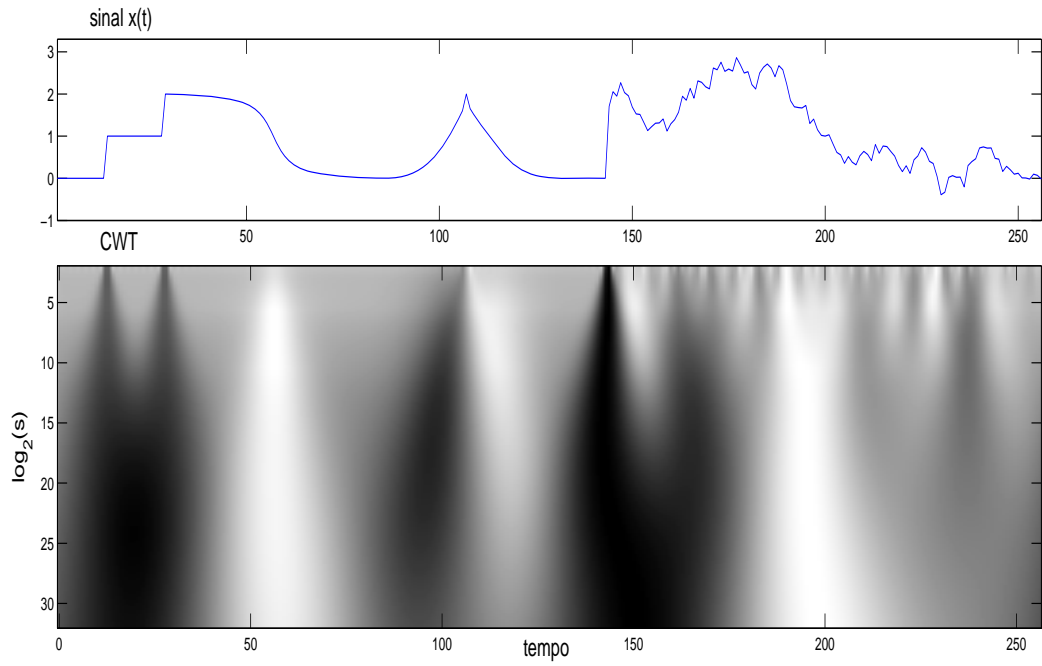


Figura 4.5: A imagem da parte inferior da figura é a CWT $W_\psi(s, \tau)$ do sinal da parte superior, calculada com uma *wavelet* que é a primeira derivada da PDF Gaussiana. Os parâmetros $\tau = t$ e s variam ao longo dos eixos horizontal e vertical, respectivamente. Grandes valores (positivos) de coeficientes *wavelet* são indicados pela cor preta. Observe que as singularidades isoladas na parte esquerda da figura produzem grandes coeficientes em seus respectivos cones de influência, que convergem para as localizações das singularidades. Esta figura foi criada com a função WaveLab `WTBrowser.m`.

em que $\mathbf{T} : X \rightarrow Y$ corresponde ao operador de análise definida pela CWT e $\mathbf{T}^* : Y \rightarrow X$ é o *operador adjunto* de \mathbf{T} , que deve satisfazer a relação

$$\langle y, \mathbf{T}x \rangle = \langle \mathbf{T}^*y, x \rangle . \quad (4.22)$$

Se $(\mathbf{T}^*\mathbf{T})^{-1}$ existe, como é o caso da CWT (e também da WFT), garante-se que vale a *resolução da identidade* [117, pág.24]

$$\mathbf{S}\mathbf{T} = \mathbf{I} . \quad (4.23)$$

em que o operador identidade \mathbf{I} é definido por $\mathbf{I}x \equiv x$, para todo x .

Kaiser [109] afirma que pode-se passar da descrição em tempo contínuo para uma descrição em tempo discreto em que $\tau = k\Delta\tau$, $k \in \mathbb{Z}$, e $s = \sigma^m$, $m \in \mathbb{R}$, desde que o tempo de amostragem $\Delta\tau$ seja suficientemente pequeno ($\Delta\tau \approx 0$) e que o fator de escala σ seja escolhido suficientemente próximo da unidade, isto é, $\sigma \approx 1$ (representação de $x(t)$ num plano tempo-escala finamente discretizado). Este resultado não surpreende, porque, se *frames* contínuos são infinitamente redundantes, então espera-se que *frames* discretos, obtidos com $\Delta\tau \approx 0$ e $\sigma \approx 1$, sejam altamente redundantes. Entretanto, Mallat [113] propôs, em meados da década de 1980, um método radicalmente diferente (e surpreendente) de implementação da transformada *wavelet* discreta, em que sinais são representados com $\Delta\tau$ finito e $\sigma \neq 1 + \epsilon$, com ϵ arbitrário, denominado análise de multirresolução (*Multiresolution Analysis* (MRA)). As propriedades de análise e reconstrução são mantidas, mesmo sem variações de escala e de tempo do tipo infinitesimal. A MRA é completamente recursiva, sendo portanto ideal para implementações computacionais.

Na MRA, $\Delta\tau = 1$ e $\sigma = 2$, o que resulta numa forma de decomposição de $x(t)$ em que as escalas de tempo são diádicas (potências de 2). A reconstrução de $x(t)$ é perfeita. As *wavelets* usadas na MRA formam conjuntos de *bases ortonormais* ao invés de *frames*. Portanto, essas novas *wavelets* não podem ser obtidas via discretização de um *frame* contínuo genérico, pois $\sigma = 2$, como dito acima. A teoria da MRA provê a “prescrição” para a construção dessas novas *wavelets*, as quais devem satisfazer outras restrições além da *condição de admissibilidade* (1).

4.2.2.2 Análise de Multirresolução e Transformada *Wavelet* Discreta

A Fig. 4.6 mostra que há dois tipos de DWT (veja que a CWT possui uma “filha” que também se chama DWT [120]): a DWT para sinais de tempo discreto e a DWT para sinais de tempo contínuo. A DWT pode ser formulada para

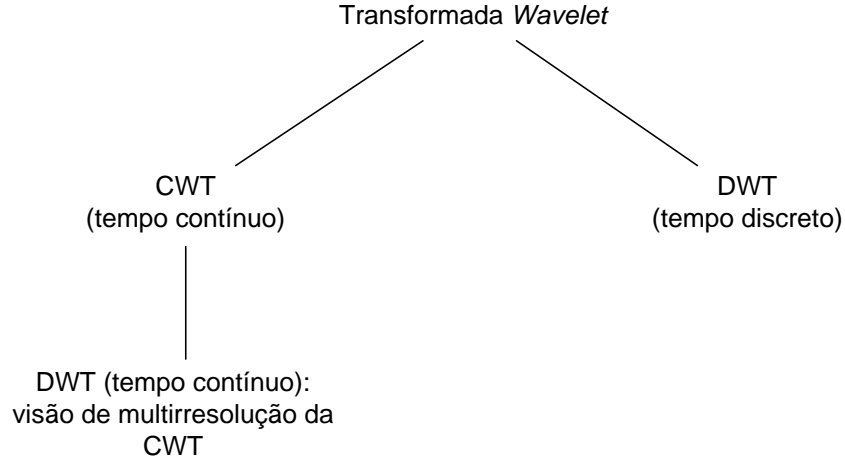


Figura 4.6: Taxonomia da transformada *wavelet*.

sinais de tempo discreto (como fazem, por exemplo, Percival e Walden em [67, Cap.4]), sem que haja o estabelecimento de uma conexão explícita com a CWT. Por outro lado, não se deve entender que o termo “discreto” da DWT para sinais de tempo contínuo queira dizer que esta transformada seja definida sobre um sinal de tempo discreto, mas tão somente que os coeficientes produzidos por esta transformada pertencem a um subconjunto $D = \{w_{j,k} = W_\psi(2^j, 2^j k), j \in \mathbb{Z}, k \in \mathbb{Z}\}$ do conjunto C [120], [110, pág.105]. De fato, os coeficientes da DWT para sinais de tempo contínuo também podem ser obtidos diretamente, por meio da integral (esta relação será demonstrada mais adiante, no estudo da análise de multirresolução)

$$w_{j,k} = \left\langle \psi_{0(2^j, 2^j k)}, x \right\rangle = \int_{-\infty}^{\infty} 2^{-j/2} \psi_0^*(2^{-j} \lambda - k) x(\lambda) d\lambda, \quad (4.24)$$

em que os índices j e k são chamados de escala e localização, respectivamente, que não envolve um sinal de tempo discreto, mas o sinal de tempo contínuo $x(t)$.

A Eq. (4.24) mostra que a DWT de tempo contínuo corresponde a uma versão criticamente amostrada da CWT definida por (4.17) nas escalas diádicas $s = 2^j$, $j = \dots, -1, 0, 1, 2, \dots$, em que os instantes de tempo na escala diádica $s = 2^j$ estão separados por múltiplos de 2^j (vide Fig. 4.7). A função ψ_0 de (4.24) deve ser definida a partir de uma análise de multirresolução (MRA) do sinal $x(t)$ [67], [117], [121], a qual é apresentada na sequência. Observe-se que a teoria da MRA de tempo contínuo é similar à de tempo discreto (veja [67], por exemplo). Apesar dos sinais de teletráfego serem de tempo discreto, o autor desta tese optou por apresentar a versão de tempo contínuo da MRA porque o estimador do parâmetro de Hurst baseado em *wavelets* proposto por Abry e Veitch [122] que será usado nos Caps. 5 e 6 é baseado na análise espectral de um processo “fictício” $\{\tilde{x}_t, t \in \mathbb{R}\}$

que é associado ao processo de tempo discreto $\{x_n, n \in \mathbb{Z}\}$ (vide [120] para obter maiores detalhes).

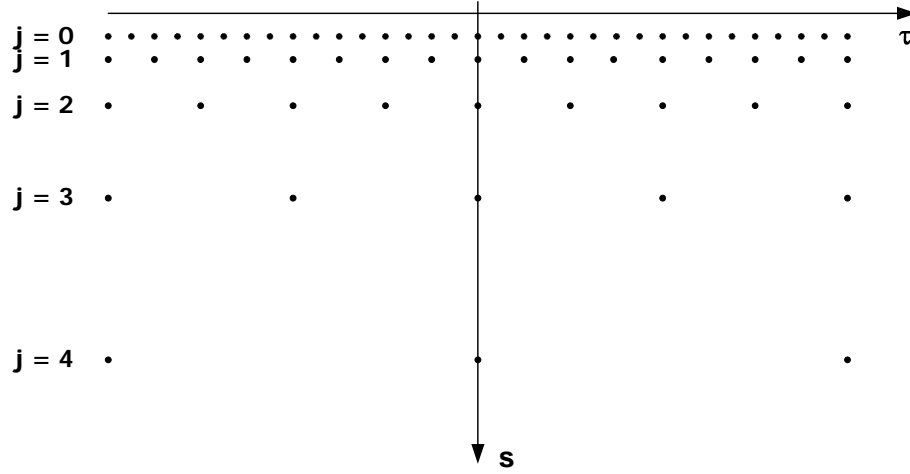


Figura 4.7: Amostragem crítica do plano tempo-escala por meio da discretização dos parâmetros da CWT ($s = 2^j$ e $\tau = 2^j k$). A CWT é definida em todos os pontos do plano (s, τ) e corresponde a uma representação redundante da informação presente no sinal. Note que o número de coeficientes dobra quando se vai de uma escala $s_1 = 2^{j+1}$ para uma escala mais rápida (ou mais refinada) $s_2 = 2^j$.

Uma MRA é, por definição, uma seqüência de subespaços fechados¹⁰ $\{V_j\}_{j \in \mathbb{Z}}$ de $L^2(\mathbb{R})$ tal que [67, pág.462], [117]:

1. $\dots V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \dots$;
2. $\bigcap_{j \in \mathbb{Z}} V_j = \{\}$;
3. $\bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R})$;
4. $x(t) \in V_j \Leftrightarrow x(2^j t) \in V_0, j > 0$ (em que t denota tempo e $x(t)$ é um sinal de energia);
5. Existe uma função $\phi_j(t) = 2^{-j/2} \phi_0(2^{-j} t)$ em V_j , denominada *função de escala*, tal que o conjunto $\{\phi_{j,k}, k \in \mathbb{Z}\}$ é uma base ortonormal de V_j , com $\phi_{j,k}(t) = 2^{-j/2} \phi_0(2^{-j} t - k) \forall j, k \in \mathbb{Z}$.

O subespaço V_j é conhecido como o **espaço de aproximação** associado à escala de tempo $s_j = 2^j$ (supondo-se que V_0 seja o espaço de aproximação com escala unitária).

¹⁰Um subespaço \mathcal{M} de um espaço de Hilbert \mathcal{H} é um subespaço **fechado** de \mathcal{H} se $\|x_n - x\| \rightarrow 0, \{x_n\}_{n \geq 1} \in \mathcal{M}$, implica que $x \in \mathcal{M}$ (ou seja, \mathcal{M} contém todos os seus pontos de acumulação).

Se a projeção sobre V_j de $x(t)$ é representada pelos coeficientes de escala

$$u_{j,k} = \langle \phi_{j,k}, x \rangle = \int_{-\infty}^{\infty} 2^{-j/2} \phi_0^*(2^{-j}t - k) x(t) dt, \quad (4.25)$$

então as propriedades 1 e 3 garantem que $\lim_{j \rightarrow -\infty} \sum_k \phi_{j,k}(t) u_{j,k} = x(t)$, $\forall x \in L^2(\mathbb{R})$. A propriedade 4 implica que o subespaço V_j é uma versão em escala do subespaço V_0 (multirresolução). A base ortonormal mencionada na propriedade 5 é obtida por translações no tempo da função passa-baixas ϕ_j .

Considere a seqüência de aproximações (também conhecidas na literatura como *wavelet smooths* [67] ou suavizações *wavelet*) sucessivas de $x(t)$

$$\mathcal{S}_j(t) = \sum_k \phi_{j,k}(t) u_{j,k} \quad j = \dots, -1, 0, 1, \dots \quad (4.26)$$

Como $V_{j+1} \subset V_j$, tem-se que $\mathcal{S}_{j+1}(t)$ é uma aproximação mais grosseira de $x(t)$ do que $\mathcal{S}_j(t)$. Este fato ilustra a idéia fundamental da MRA, que consiste em examinar a **perda de informação** quando se vai de $\mathcal{S}_j(t)$ para $\mathcal{S}_{j+1}(t)$:

$$\mathcal{S}_j(t) = \mathcal{S}_{j+1}(t) + \Delta x_{j+1}(t), \quad (4.27)$$

em que $\Delta x_{j+1}(t)$ (dito **detalhe** de $x_j(t)$) pertence ao subespaço W_{j+1} , denominado espaço do detalhe [67] (também às vezes chamado de subespaço *wavelet* [122]), o qual está associado às flutuações (ou variações) do sinal na escala de tempo mais refinada $s_j = 2^j$ (qualitativamente, os coeficientes *wavelet* $w_{j,k}$ da escala j são proporcionais às diferenças entre médias adjacentes do sinal x_t na escala de tempo $\tau_j = s_{j-1} = 2^{j-1}$ [67, pág.59]), e que corresponde ao complemento ortogonal de V_{j+1} em V_j ¹¹. A MRA mostra que os sinais de detalhe $\Delta x_{j+1}(t) = \mathcal{D}_{j+1}(t)$ podem ser obtidos diretamente a partir de projeções sucessivas do sinal original $x(t)$ sobre subespaços *wavelet* W_j . Além disso, a teoria da MRA demonstra que existe uma função $\psi_0(t)$, denominada “*wavelet* mãe”, que é obtida a partir de $\phi_0(t)$, tal que $\psi_{j,k}(t) = 2^{-j/2} \psi_0(2^{-j}t - k)$ $k \in \mathbb{Z}$ é uma base ortonormal de W_j .

O detalhe $\mathcal{D}_{j+1}(t)$ é obtido pela equação

$$\mathcal{D}_{j+1}(t) = \sum_k \psi_{j+1,k}(t) \langle \psi_{j+1,k}(t), x(t) \rangle, \quad (4.28)$$

em que o produto interno $\langle \psi_{j+1,k}(t), x(t) \rangle = w_{j+1,k}$ denota o coeficiente *wavelet* associado à escala $j+1$ e tempo discreto k e $\{\psi_{j+1,k}(t)\}$ é uma família de funções *wavelets* que gera o subespaço W_{j+1} , ortogonal ao subespaço V_{j+1} ($W_{j+1} \perp V_{j+1}$),

¹¹Além disso, W_{j+1} está contido no subespaço V_j .

isto é,

$$\langle \psi_{j+1,n}, \phi_{j+1,p} \rangle = 0, \forall n, p. \quad (4.29)$$

Portanto, o sinal de detalhe $\mathcal{D}_{j+1}(t)$ pertence ao subespaço complementar W_{j+1} de V_j , pois

$$V_j = V_{j+1} \oplus W_{j+1}, \quad (4.30)$$

ou seja, V_j é dado pela soma direta de V_{j+1} e W_{j+1} , e isto quer dizer que qualquer elemento em V_j pode ser determinado a partir da soma de dois elementos ortogonais pertencentes a V_{j+1} e W_{j+1} . Iterando-se (4.30), tem-se que

$$V_j = W_{j+1} \oplus W_{j+2} \oplus \dots \quad (4.31)$$

A Eq. (4.31) diz que a aproximação $\mathcal{S}_j(t)$ é dada por

$$\mathcal{S}_j(t) = \sum_{i=j+1}^{\infty} \sum_k w_{i,k} \psi_{i,k}(t). \quad (4.32)$$

A MRA de um sinal de tempo contínuo $x(t)$ é iniciada com a determinação dos coeficientes¹² $u_0(k) = \langle \phi_{0,k}(t), x(t) \rangle$, em que $k = 0, 1, \dots, N-1$, que estão associados à projeção de $x(t)$ no subespaço de aproximação V_0 . Em seguida, a sequência $\{u_0(k)\}$ é decomposta via filtragem e subamostragem por um fator de 2 (*downsampling*) em duas sequências: $\{u_1(k)\}$ e $\{w_1(k)\}$, cada uma contendo $N/2$ pontos. Este processo de filtragem e subamostragem é repetido várias vezes, obtendo-se as sequências

$$\left\{ \{u_0(k)\}_N, \{u_1(k)\}_{\frac{N}{2}}, \{u_2(k)\}_{\frac{N}{4}}, \dots, \{u_j(k)\}_{\frac{N}{2^j}}, \dots, \{u_J(k)\}_{\frac{N}{2^J}} \right\} \quad (4.33)$$

e

$$\left\{ \{w_1(k)\}_{\frac{N}{2}}, \{w_2(k)\}_{\frac{N}{4}}, \dots, \{w_j(k)\}_{\frac{N}{2^j}}, \{w_J(k)\}_{\frac{N}{2^J}} \right\}. \quad (4.34)$$

A literatura [120], [122] denomina o conjunto de coeficientes

$$\left\{ \{w_1(k)\}_{\frac{N}{2}}, \{w_2(k)\}_{\frac{N}{4}}, \dots, \{w_J(k)\}_{\frac{N}{2^J}}, \{u_J(k)\}_{\frac{N}{2^J}} \right\} \quad (4.35)$$

como a DWT do sinal $x(t)$.

A Fig. 4.8 ilustra a DWT de 3 níveis (decomposição nas escalas $j = 1, 2, 3$) associada a 1024 amostras do sinal de tempo discreto $x(k) = \sin(3k) + \sin(0,3k) + \sin(0,03k)$, que corresponde à superposição de 3 senóides nas frequências $f_1 \approx 0,004775$, $f_2 \approx 0,04775$ e $f_3 \approx 0,4775$. A Fig. 4.9 mostra a DEP deste sinal.

¹²A sequência $u_0(k)$ é obtida amostrando-se a saída de um filtro com resposta impulsiva $\phi^*(-t)$ (filtro casado com a função $\phi_0(t) = \phi(t)$) nos instantes $k = 0, 1, 2, \dots$, ou seja, $u_0(k) = x(t) \star \phi^*(-t)$ para $k = 0, 1, 2, \dots$, em que \star denota convolução.

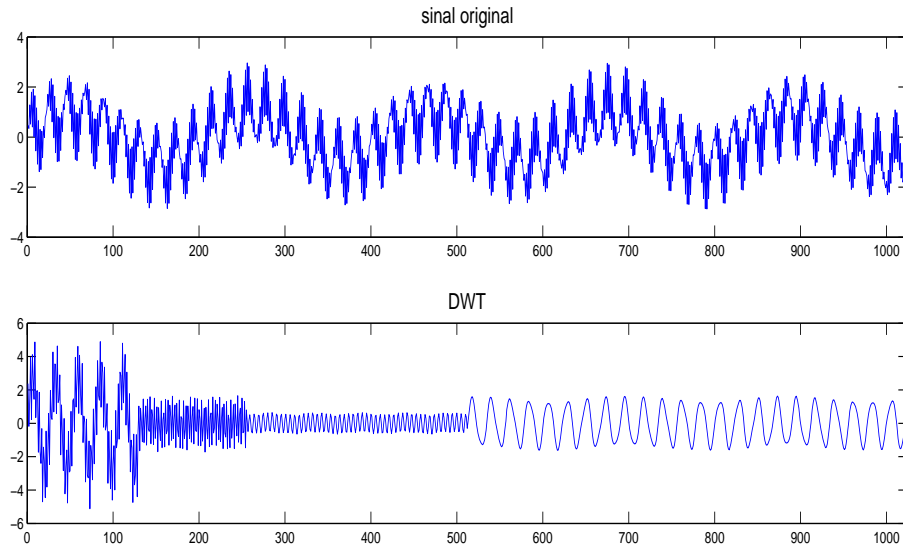


Figura 4.8: Uma ilustração da DWT de 3 níveis do sinal de tempo discreto $x(k) = \sin(3k) + \sin(0,3k) + \sin(0,03k)$. O gráfico concatena as seqüências dos coeficientes de escala $\{u_3(k)\}_{128}$ e dos coeficientes *wavelet* $\{w_3(k)\}_{128}$, $\{w_2(k)\}_{256}$ e $\{w_1(k)\}_{512}$ da esquerda para a direita, ou seja, os primeiros 128 pontos correspondem à seqüência $\{u_3(k)\}_{128}$; seguem-se os 128 pontos da seqüência $\{w_3(k)\}_{128}$, os 256 pontos da seqüência $\{w_2(k)\}_{256}$ e os 512 pontos da seqüência $\{w_1(k)\}_{512}$.

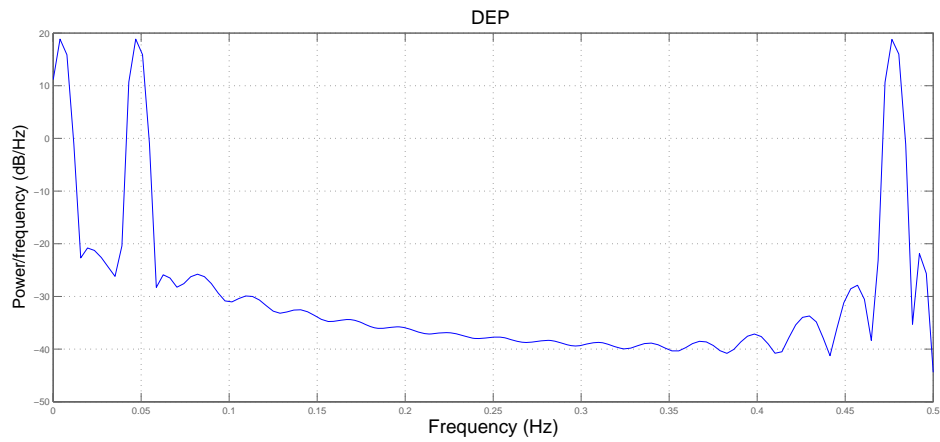


Figura 4.9: DEP do sinal $x(k) = \sin(3k) + \sin(0,3k) + \sin(0,03k)$.

A reconstrução de $x(t)$ é implementada via filtragem e sobreamostragem por um fator de 2 (*upsampling*) das seqüências (4.33) e (4.34), obtendo-se uma aproximação de $x(t)$ no subespaço V_0

$$\mathcal{S}_0(t) = \mathcal{S}_J(t) + \mathcal{D}_1(t) + \mathcal{D}_2(t) + \cdots + \mathcal{D}_J(t) \quad (4.36)$$

ou

$$x(t) \approx \sum_k u(J, k) \phi_{J,k}(t) + \sum_{j=1}^J \sum_k w_{j,k} \psi_{j,k}(t). \quad (4.37)$$

A Eq. (4.37) define a transformada discreta *wavelet* inversa - *Inverse Discrete Wavelet Transform* (IDWT). A Fig. 4.10 ilustra a síntese do sinal $x(k) = \sin(3k) + \sin(0,3k) + \sin(0,03k)$ conforme (4.36) (utilizou-se a *wavelet* de Haar).

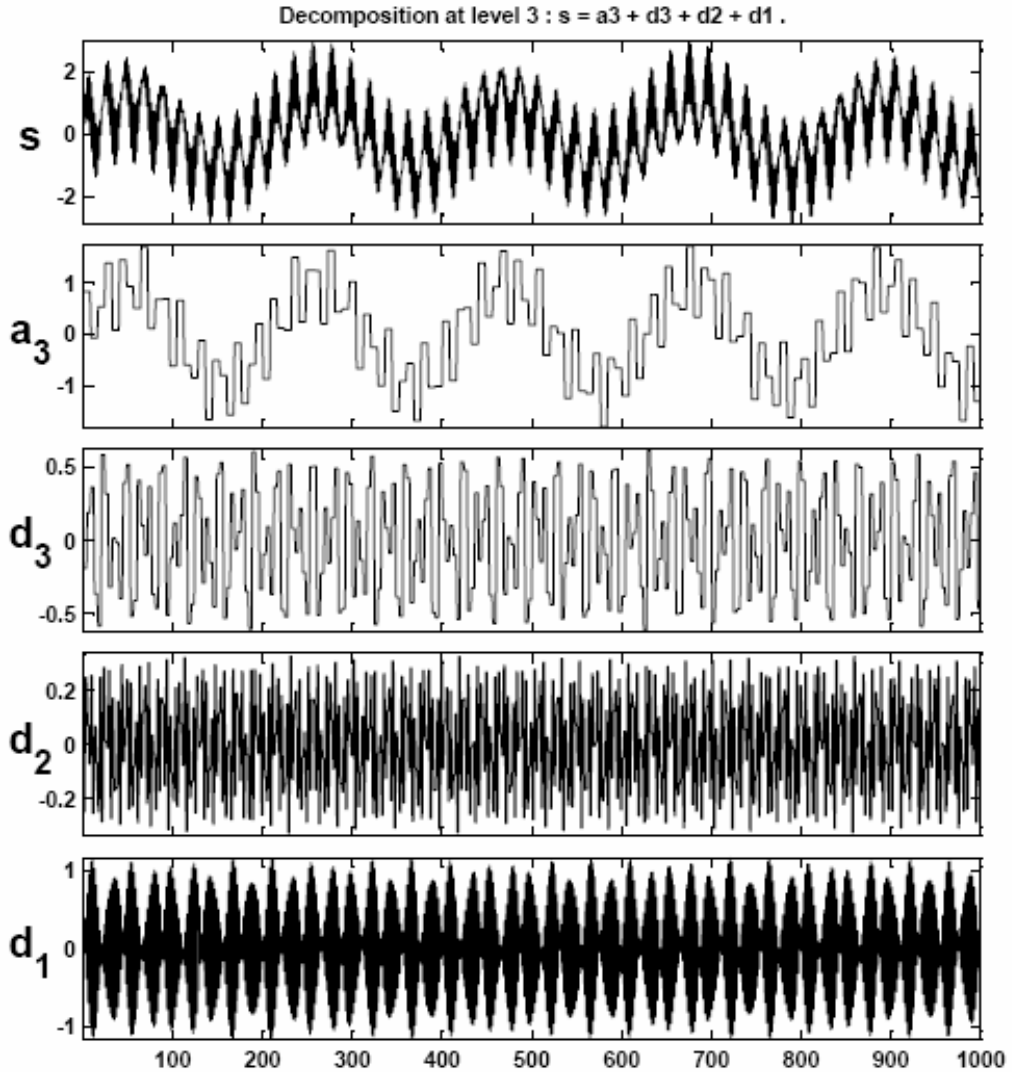


Figura 4.10: Síntese do sinal $x(k) = \sin(3k) + \sin(0,3k) + \sin(0,03k)$ em termos da soma $\mathcal{S}_3(t) + \mathcal{D}_1(t) + \mathcal{D}_2(t) + \mathcal{D}_3(t)$ (aproximação na escala $j = 3$ e detalhes nas escalas 1, 2 e 3). Na Fig., $a_3 = \mathcal{S}_3$ e $d_j = \mathcal{D}_j$, $j = 1, 2, 3$.

Diz-se que a função $\phi_0(t) = \phi(t)$ determina uma MRA de $x(t)$ de acordo com

(4.36), se a mesma obedece às seguintes condições:

1. ortonormalidade intra-escala (propriedade 5)

$$\langle \phi(t-m), \phi(t-n) \rangle = \delta_{m,n}, \quad (4.38)$$

em que $\delta_{m,n}$ é o delta de Kronecker ($\delta_{m,n} = 1$ se $m = n$, $\delta_{m,n} = 0$ para $m \neq n$). A Eq. (4.38) impõe uma condição de ortonormalidade na escala $j = 0$.

2. média unitária

$$\int_{-\infty}^{\infty} \phi(t) dt = 1. \quad (4.39)$$

- 3.

$$\frac{1}{\sqrt{2}} \phi\left(\frac{t}{2}\right) = \sum_n g_n \phi(t-n), \quad (4.40)$$

pois “cabem” várias $\phi(t-k)$ em $\phi(\frac{t}{2})$ (é uma consequência da propriedade (1) da MRA).

A Eq. 4.40 pode ser reescrita na forma

$$\phi(t) = \sum_n \sqrt{2} g_n \phi(2t-n), \quad (4.41)$$

conhecida como **Equação de Dilação**. As Eqs. 4.40 e 4.41 podem ser escritas, respectivamente, no domínio das frequências como

$$\sqrt{2} \Phi(2\nu) = G(\nu) \Phi(\nu), \quad (4.42)$$

e

$$\Phi(\nu) = \frac{1}{\sqrt{2}} G(\nu) \Phi\left(\frac{\nu}{2}\right), \quad (4.43)$$

em que $\Phi(\nu)$ é a transformada de Fourier de $\phi(t)$ e $G(\nu) = \sum_n g_n e^{-j2\pi\nu n}$, conhecido como **filtro de escala** (passa-baixas), representa um filtro periódico em ν .

Como o subespaço W_{j+1} é ortogonal a V_{j+1} e está contido em V_j , tem-se que

$$\frac{1}{\sqrt{2}} \psi\left(\frac{t}{2}\right) = \sum_n h_n \phi(t-n), \quad (4.44)$$

ou

$$\psi(t) = \sum_n \sqrt{2} h_n \phi(2t-n), \quad (4.45)$$

que é a **Equação da Wavelet**. Aplicando-se a transformada de Fourier em

(4.44) e (4.45) obtém-se, respectivamente,

$$\sqrt{2}\Psi(2\nu) = H(\nu)\Phi(\nu), \quad (4.46)$$

e

$$\Psi(\nu) = \frac{1}{\sqrt{2}}H(\nu)\Phi\left(\frac{\nu}{2}\right). \quad (4.47)$$

em que $H(\nu)$ é o **filtro *wavelet*** (passa-altas).

Reescrevendo-se (4.29) em termos do domínio das frequências e usando-se (4.42) e (4.46) resulta a **condição de ortogonalidade**

$$\int_{-\infty}^{\infty} G(\nu)H^*(\nu)|\Phi(\nu)|^2 d\nu = 0, \quad (4.48)$$

que o filtro H deve atender para que a família $\{\psi_{1,k}(t)\}$ seja ortogonal à família $\{\phi_{1,k}(t)\}$. Pode-se mostrar [109, pág.150], [67, pág.75] que a condição

$$h_n = (-1)^n g_{L-1-n}, \quad \leftrightarrow \quad H(z) = -z^{-L+1}G(-z^{-1}), \quad (4.49)$$

em que L denota o comprimento de um filtro FIR g_n , é suficiente para que valha (4.48). Diz-se que g_n e h_n são **filtros espelhados em quadratura** (ou ***Quadrature Mirror Filters (QMF)***) quando estão relacionados por (4.49). A Fig. 4.11 mostra os gráficos de resposta em frequência dos filtros QMF e também ilustra a resposta em frequência de filtros do tipo *brickwall*, que não são fisicamente realizáveis.

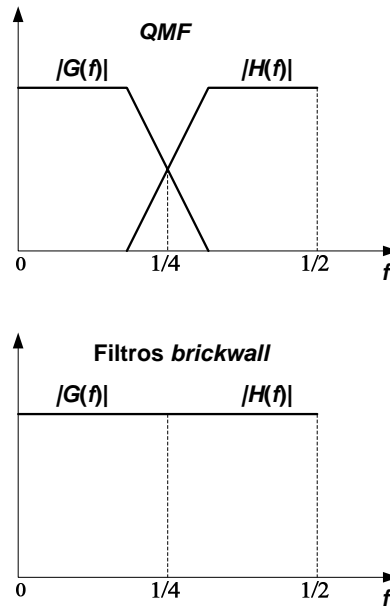


Figura 4.11: Resposta em frequência de filtros QMF (gráfico da parte superior) *vs* resposta em frequência de filtros do tipo *brickwall* (fisicamente não-realizáveis).

De acordo com (4.41), a MRA começa a partir de uma definição (dentre várias possíveis) da função de escala $\phi(t)$, que está relacionada ao filtro de escala g_n por (4.40). A Eq. (4.49) diz que a escolha de um filtro $\{g_n\}$ do tipo *Finite Impulse Response* (FIR) implica um $\{h_n\}$ que também seja FIR. Finalmente, a função *wavelet* é determinada por (4.44). As funções de escala $\phi(t)$ e *wavelet* $\psi(t)$ associadas aos filtros FIR $\{g_n\}$ e $\{h_n\}$ possuem suporte compacto, oferecendo, portanto, a funcionalidade de *resolução temporal*.

A função de escala mais simples que satisfaz (4.38) é a função característica¹³ do intervalo $I = [0, 1)$, que corresponde à função de escala de Haar:

$$\phi^{(H)}(t) = \chi_{[0,1)}(t) = \begin{cases} 1 & \text{se } 0 \leq t < 1 \\ 0 & \text{caso contrário.} \end{cases} \quad (4.50)$$

Neste caso (MRA de Haar), o filtro de escala de Haar associado é dado por

$$g_n = \{\dots, 0, g_0 = 1/\sqrt{2}, g_1 = 1/\sqrt{2}, 0, \dots\}, \quad (4.51)$$

o filtro *wavelet* de Haar por

$$h_n = \{\dots, 0, h_0 = g_1 = 1/\sqrt{2}, h_1 = -g_0 = -1/\sqrt{2}, 0, \dots\} \quad (4.52)$$

e a função *wavelet* de Haar por

$$\psi^{(H)}(t) = \chi_{[0,1/2)}(t) - \chi_{[1/2,1)}(t). \quad (4.53)$$

A Fig. 4.12 mostra as funções de escala e *wavelets* de Daubechies com $N = 2, 3, 4$ momentos ou cumulantes nulos (*vanishing moments*)

$$\int_{-\infty}^{\infty} t^m \psi(t) dt = 0, \quad m = 0, 1, \dots, N-1. \quad (4.54)$$

Ingrid Daubechies [116] foi a primeira a propor um método para construção de seqüências de funções de transferência $\{G^{(N)}(z)\}_{N=1,2,3,\dots}$ e $\{H^{(N)}(z)\}_{N=1,2,3,\dots}$, em que $G^{(N)}(z)$ está associada ao filtro FIR passa-baixas $g_n^{(N)}$ e $H^{(N)}(z)$ ao filtro passa-altas $h_n^{(N)}$. As funções de escala e *wavelet* correspondentes têm suporte em $[0, 2N-1]$. O primeiro membro da seqüência é o sistema de Haar $\phi^{(1)} = \phi^{(H)}$, $\psi^{(1)} = \psi^{(H)}$. Os filtros de Daubechies são generalizações do sistema de Haar para $N \geq 2$ (consulte [109] para obter maiores informações).

¹³Definida por

$$\chi_E(x) = \begin{cases} 1 & \text{se } x \in E \\ 0 & \text{se } x \notin E. \end{cases}$$

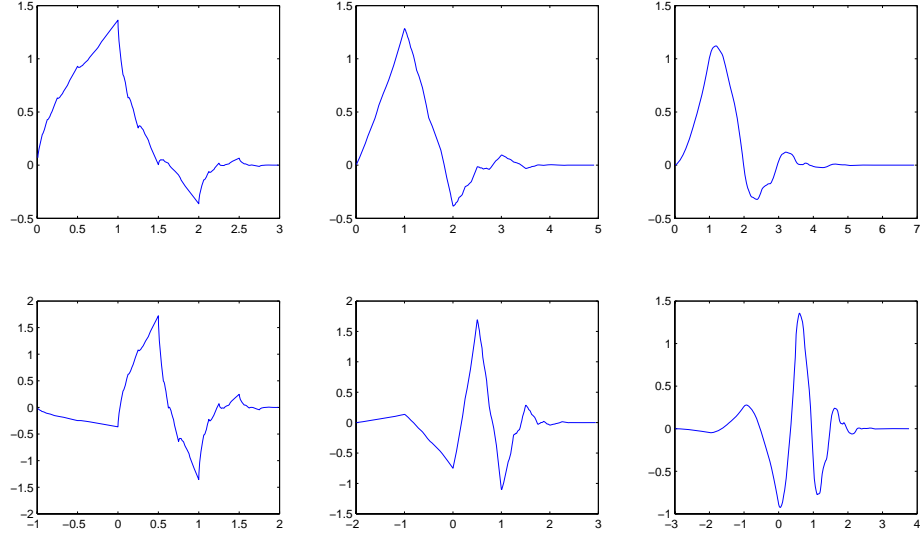


Figura 4.12: Os gráficos da parte inferior mostram as *wavelets* de Daubechies com $N = 2, 3, 4$ momentos nulos (*vanishing moments*), da esquerda para a direita, respectivamente. As funções de escala correspondentes estão na parte superior. Esta figura foi criada com a função WaveLab `WTBrowser.m`.

Demonstra-se que [113]:

$$u_j(n) = \sum_k g(k - 2n)u_{j-1}(k) \quad (4.55)$$

e que

$$w_j(n) = \sum_k h(k - 2n)u_{j-1}(k). \quad (4.56)$$

De acordo com (4.55) e (4.56), pode-se obter os coeficientes $u_j(n)$ e $w_j(n)$ a partir dos coeficientes de escala $u_{j-1}(m)$ por meio de uma operação de dizimação da seqüência $\{u_{j-1}(m)\}$ por um fator de 2. A dizimação (ou decimação) consiste no cascadeamento de um filtro passa-baixas $g(-m)$ (com função de transferência $\bar{G}(z) = G(1/z)$ e resposta em frequência $G^*(f)$) ou passa-altas $h(-m)$ (com função de transferência $\bar{H}(z) = H(1/z)$ e resposta em frequência $H^*(f)$) com um compressor (ou dizimador) por um fator de 2, conforme ilustrado pela Fig. 4.13a. Note que decimar um sinal por um fator D é o mesmo que reduzir sua taxa de amostragem em D vezes.

A Fig. 4.13a sugere que os coeficientes de uma DWT podem ser determinados a partir de um algoritmo piramidal baseado em convoluções com filtros espelhados em quadratura (esta foi a proposta original de Mallat em [113]). Sendo assim, a MRA é implementada via bancos de filtros de análise passa-baixas $G^*(f)$ e passa-altas $H^*(f)$ adequadamente posicionados para separação das seqüências

de coeficientes de escala das seqüências de coeficientes *wavelet*. Posteriormente é possível reconstruir o sinal original utilizando-se o banco de filtros duais de reconstrução QMF passa-baixas $G(f)$ e passa-altas $H(f)$ [113], conforme mostra a Fig. 4.13b¹⁴. De acordo com a Fig. 4.13b, $u_{j-1}(n)$ pode ser reconstruído por meio da inserção de zeros entre cada duas amostras de $u_j(m)$ e $w_j(m)$, gerando-se os sinais $u_j^{\text{up}}(n)$ e $w_j^{\text{up}}(n)$ nas saídas dos interpoladores (a inserção de zeros é conhecida como operação de interpolação, em que há um aumento da taxa de amostragem do sinal ou *upsampling*), aos quais se seguem convoluções com os filtros $G(f)$ e $H(f)$, respectivamente. É importante ressaltar que a complexidade do algoritmo da pirâmide é $O(N)$ (assumindo-se que se quer calcular a DWT de N amostras), ao passo que o cálculo “direto” da DWT (que envolve multiplicação de matrizes) é $O(N^2)$ [67].

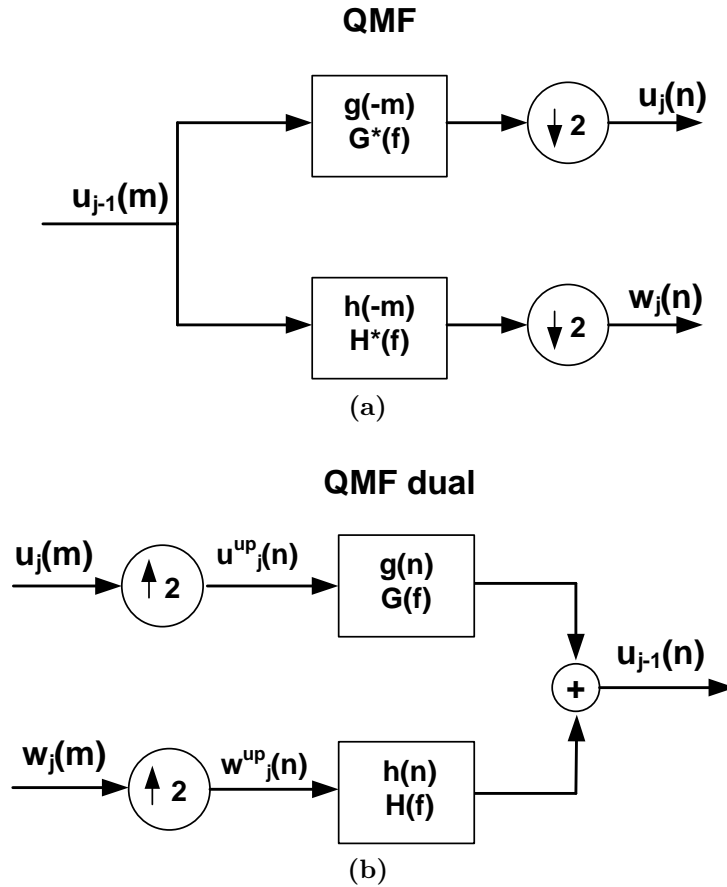


Figura 4.13: (a) banco de filtros QMF de análise $G^*(f)$ (passa-baixas) e $H^*(f)$ (passa-altas) com decimação (*downsampling*) por um fator de 2; (b) banco de filtros QMF de reconstrução com interpolação (*upsampling*) por um fator de 2. Note que são usados os filtros duais passa-baixas $G(f)$ e passa-altas $H(f)$.

A Fig. 4.14a é um diagrama de fluxo que ilustra a projeção inicial de um sinal $x(t)$ sobre V_0 seguida da MRA em W_1 , W_2 e V_2 em que, $G^*\left(\frac{k}{N/2^{j-1}}\right)$ e

¹⁴Pode-se adotar os filtros QMF de reconstrução passa-baixas $G(f)$ e passa-altas $H(f)$ desde que os filtros de análise sejam os filtros duais passa-baixas $G^*(f)$ e passa-altas $H^*(f)$.

$H^* \left(\frac{k}{N/2^j-1} \right)$ representam a Transformada Discreta de Fourier (TDF) dos filtros circulares de escala (passa-baixas) e *wavelet* (passa-altas), respectivamente [67]. O símbolo “ $\downarrow 2$ ” indica a operação de *downsampling* de um sinal. A Fig. 4.14b representa a reconstrução (aproximada) de $x(t)$ a partir de W_1 , W_2 e V_2 . Observe que os filtros circulares correspondem aos complexos conjugados dos filtros usados na análise. O símbolo “ $\uparrow 2$ ” indica a operação de *upsampling* de um sinal.

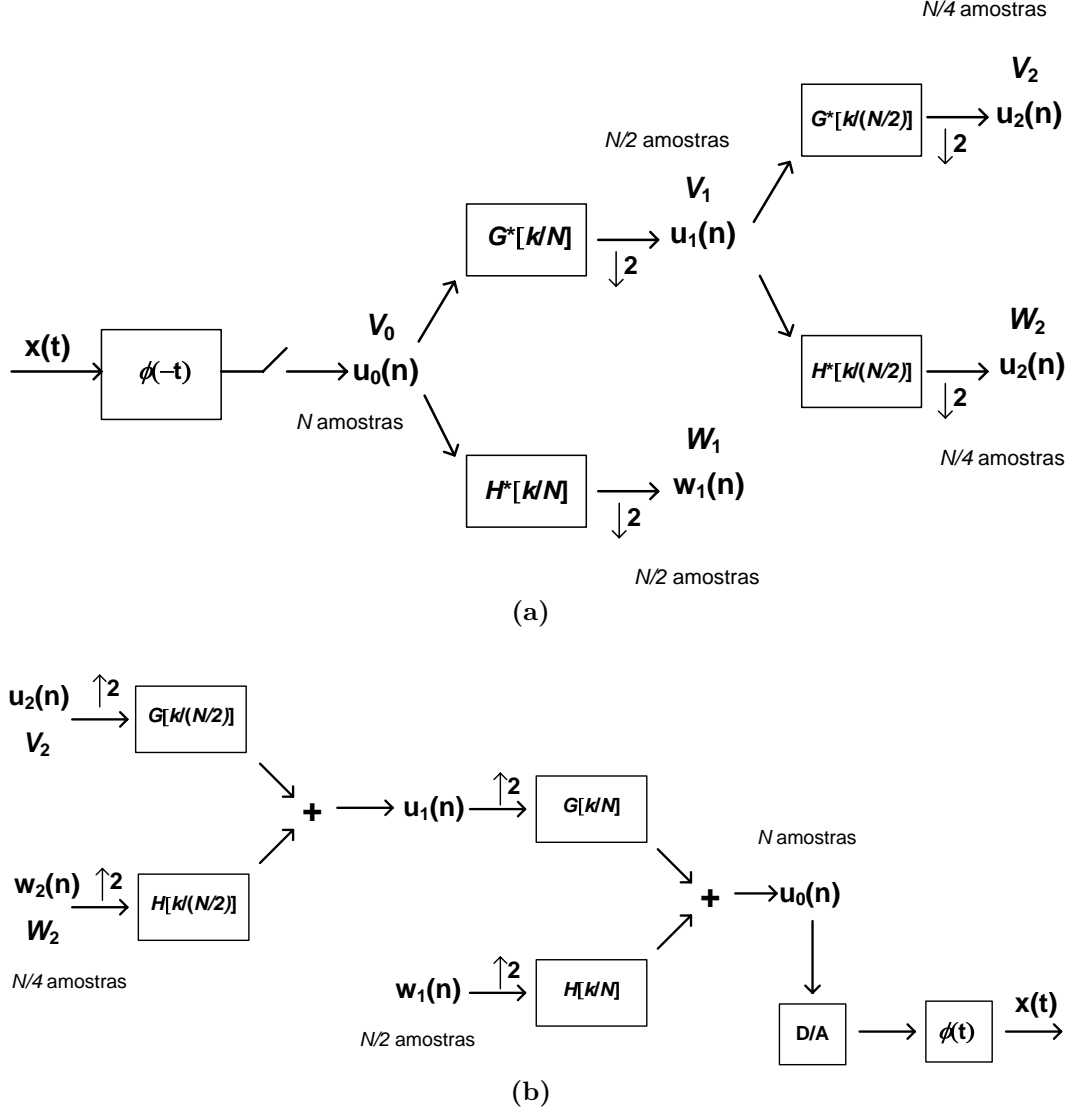


Figura 4.14: (a) Diagrama de fluxo que mostra a projeção inicial de um sinal $x(t)$ sobre V_0 seguida da decomposição em W_1 , W_2 e V_2 . (b) Diagrama de fluxo que ilustra a síntese aproximada de $x(t)$ a partir de W_1 , W_2 e V_2 .

A Fig. 4.15 mostra que o espectro $U_0(f)$ do sinal $u_0(n)$ da Fig. 4.14 é subdividido em três bandas de frequência (que cobrem duas oitavas): $0 \leq f < 1/8$, $1/8 \leq f < 1/4$ e $1/4 \leq f \leq 1$. É interessante observar que a Fig. 4.15 ilustra que a DWT utiliza o princípio da codificação de sub-bandas usado em sistemas práticos digitais multitaxa de codificação de sinais de voz (veja [123] e [52] para maiores detalhes).

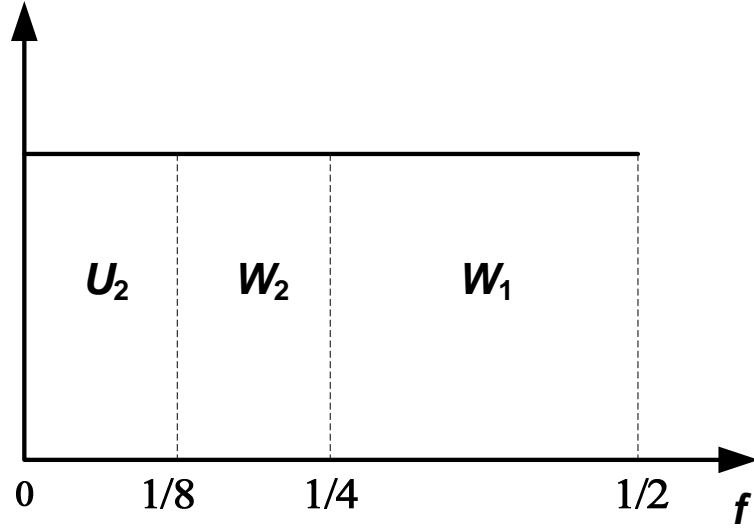


Figura 4.15: Diagrama de blocos que mostra que a DWT funciona de modo similar a um esquema de codificação de sub-bandas. O espectro $U_0(f)$ do sinal $u_0(n)$ da Fig. 4.14 é subdividido em três bandas de frequência (que cobrem duas oitavas): $0 \leq f < 1/8$, $1/8 \leq f < 1/4$ e $1/4 \leq f \leq 1$.

4.2.3 Modelo MWM

O MWM usa o sistema de MRA de Haar e está baseado numa cascata binomial multiplicativa no domínio *wavelet*, a qual garante que as séries simuladas são positivas (o que não acontece quando se utiliza modelos Gaussianos, tais como o FGN, na síntese de teletráfego) [11]. A cascata binomial é uma árvore binomial aleatória cuja raiz é o coeficiente $u_{J-1,0}$ (o MWM considera que $\frac{N}{2^{J-1}} = 1$, onde N denota o número de amostras)

$$x_t = x_{0,k} = u_{J-1,0}\phi_{J-1,0}(t) + \sum_{j=1}^{J-1} \sum_k w_{j,k}\psi_{j,k}(t), \quad (4.57)$$

em que $\phi_{J-1,0}(t)$ denota a função de escala de Haar na escala mais lenta (ordem $J-1$) e os $w_{j,k}$ são os coeficientes *wavelet*.

Os coeficientes de escala e *wavelet* de Haar podem ser calculados recursivamente por meio do seguinte conjunto de equações de síntese,

$$u_{j-1,2k} = 2^{-1/2}(u_{j,k} + w_{j,k}) \quad (4.58)$$

$$u_{j-1,2k+1} = 2^{-1/2}(u_{j,k} - w_{j,k}). \quad (4.59)$$

Portanto, sinais estritamente positivos podem ser modelados se $u_{j,k} \geq 0$ e

$$|w_{j,k}| \leq u_{j,k}. \quad (4.60)$$

É possível escolher um modelo estatístico para os $w_{j,k}$ que incorpore a condição (4.60). O MWM especifica um modelo multiplicativo,

$$w_{j,k} = M_{j,k} u_{j,k}, \quad (4.61)$$

em que o multiplicador $M_{j,k}$ pode ser modelado como uma variável aleatória com distribuição β simétrica com parâmetro de forma p_j , ou seja, $M_j \sim \beta(p_j, p_j)$. Neste caso, o MWM é conhecido como β -MWM e assume-se que os $M_{j,k}$'s são mutuamente independentes e independentes dos $u_{j,k}$ ¹⁵.

A variância de M_j é dada por [11]

$$\text{Var}[M_j] = \frac{1}{2p_j + 1}. \quad (4.62)$$

Desta forma, as equações (4.58) e (4.59) podem ser reescritas como

$$u_{j-1,2k} = \left(\frac{1 + M_{j,k}}{\sqrt{2}} \right) u_{j,k} \quad (4.63)$$

$$u_{j-1,2k+1} = \left(\frac{1 - M_{j,k}}{\sqrt{2}} \right) u_{j,k}. \quad (4.64)$$

Estas equações mostram que o MWM é de fato uma cascata binomial.

O MWM pode aproximar a DEP de uma sequência de treinamento por meio da modelagem do decaimento da variância dos coeficientes *wavelet*

$$\eta_j = \frac{\text{Var}[w_{j,k}]}{\text{Var}[w_{j-1,k}]} = \frac{2p_{(j-1)} + 1}{p_{(j)} + 1}, \quad (4.65)$$

que leva a

$$p_{(j-1)} = \frac{\eta_j}{2}(p_{(j)} + 1) - 1/2 \quad (4.66)$$

e

$$p_{(j)} = \frac{2p_{(j-1)} + 1}{\eta_j} - 1. \quad (4.67)$$

O MWM assume que $u_{J-1,0}$ (o coeficiente de escala “raiz”) seja aproximadamente Gaussiano. Pode-se mostrar que $p_{(j)}$ converge para

$$p_{-\infty} = \lim_{j \rightarrow -\infty} p_{(j)} = \frac{2^\alpha - 1}{2 - 2^\alpha}, \quad (4.68)$$

em que α e H estão relacionados por (3.4).

A Tabela 4.1 lista alguns valores assintóticos para o parâmetro de forma p .

O modelo MWM tem propriedades multifractais e a densidade de probabilidade

¹⁵Riedi et al[11] também investigaram o uso de outras distribuições para os multiplicadores.

Tabela 4.1: Valores assintóticos do parâmetro de forma p dos multiplicadores β em função de α (or H)[11].

α	0.1	0.2	0.5	0.8
p	0.077	0.175	0.707	2.86
H	0.55	0.6	0.75	0.9

marginal é lognormal [11].

4.3 Modelagem Paramétrica

4.3.1 Modelo ARFIMA

Conforme explicado no item 2.3, a modelagem de uma série temporal (linear) x_t consiste na estimação de uma função de transferência (ou modelo) $H(B)$ tal que

$$\mathbf{x}_t = H(B)\mathbf{w}_t, \quad (4.69)$$

em que \mathbf{w}_t é a inovação no instante t . Na prática, a modelagem é baseada na estimação da função inversa $G(B) = H(B)^{-1}$, pois espera-se que a filtragem de x_t por $G(B)$ produza uma série de resíduos w_t do tipo RB.

Granger e Joyeux [81] e Hosking [82] introduziram de forma independente a classe de modelos ARFIMA que possui as seguintes propriedades:

1. modelagem explícita da memória longa;
2. flexibilidade para modelar a estrutura de autocorrelação das séries nos pequenos e grandes *lags*;
3. possibilitar a simulação de séries LRD a partir do modelo.

Considere a equação

$$\Delta^d \mathbf{x}_t = \mathbf{w}_t, \quad (4.70)$$

em que d é um expoente fracionário¹⁶, $0 < d < 1/2$. Observe-se que

$$\Delta^d = (1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k, \quad (4.71)$$

¹⁶Uma característica de séries LRD é que as autocorrelações amostrais indicam não-estacionariedade [42, pág.460]. Portanto, faria sentido modelar uma série LRD, pelo menos numa primeira tentativa, como um processo integrado de primeira ordem ($\mathbf{x}_t \sim I(d=1)$). Entretanto, a DEP da série diferenciada tende a zero na frequência zero (não é um ruído branco) [43, pág.260], [81] ou seja, parece ser “super-diferenciada”. Este fato justifica, de maneira intuitiva, a modelagem de séries LRD por meio de processos de integração fracionária.

com coeficientes binomiais¹⁷

$$\binom{d}{k} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}, \quad (4.72)$$

resulta no **filtro de diferença fracionária**

$$\Delta^d = 1 - dB + \frac{1}{2!}d(d-1)B^2 - \frac{1}{3!}d(d-1)(d-2)B^3 + \dots, \quad (4.73)$$

que é definido para qualquer real $d > -1$. De acordo com (4.73), o modelo (4.70) é do tipo $\text{AR}(\infty)$ (vide (2.45)). A Eq. (4.70) define o **processo fracionário integrado** (também chamado de modelo $\text{FD}(d)$ [67] ou **RB fracionário** [42]), que é uma extensão do modelo integrado $\text{ARIMA}(0, d, 0)$ (2.66), $d \in \mathbb{Z}_+$. O processo FD consegue modelar a singularidade do tipo $1/f^\alpha$ na origem do espectro de uma série LRD. O FD é estacionário e LRD quando $0 < d < 1/2$; é estacionário e SRD quando $-1/2 < d < 0$; é não-estacionário¹⁸ quando $|d| > 1/2$.

Na prática, observa-se que o decaimento das SACF para **pequenos valores de lag** de algumas séries reais de teletráfego é bem modelado por processos SRD [50], [11], [77], [80], [106], [78], ou seja, têm autocorrelações significativas que decaem de modo exponencial para pequenos *lags*, a qual não é modelada pelo processo $\text{FD}(d)$. Isto não quer dizer que este tipo de série de tráfego não seja assintoticamente LRD (lembre-se que a SACF de séries LRD, para **valores suficientemente grandes de lag**, decresce segundo uma função potência, isto é, o decaimento para zero é extremamente lento e do tipo hiperbólico), mas tão somente que a característica de SRD pode se manifestar por meio da existência de “picos locais” de DEP (além da singularidade na origem do espectro que é devida à memória longa), conforme ilustrado pela Fig. 4.16 (imagine o espectro resultante da superposição dos espectros dos processos $\text{FD}(0, 4)$ e $\text{AR}(4)$). É daí que surge a necessidade de se introduzir a classe $\text{ARFIMA}(p, d, q)$ de modelos (mais flexível do que a classe FD)

$$\phi(B)\Delta^d \mathbf{x}_t = \theta(B)\mathbf{w}_t, \quad (4.74)$$

em que $-1/2 < d < 1/2$, $\phi(B)$ é o operador auto-regressivo de ordem p (2.33), $\theta(B)$ é o operador de média móvel de ordem q (2.34) e \mathbf{w}_t é um RB Gaussiano. O modelo (4.74) é **LRD, estacionário e invertível** quando $0 < d < 1/2$ e se os pólos e zeros de $\theta(z)/\phi(z)$ estiverem dentro do círculo de raio unitário (vide (2.36) e (2.37)).

O parâmetro d modela a estrutura da autocorrelação de ordens altas (em que

¹⁷A função Gamma estende a função fatorial para números reais e complexos: $d! = \Gamma(d+1)$.

¹⁸Neste caso, \mathbf{x}_t tem variância infinita [81].

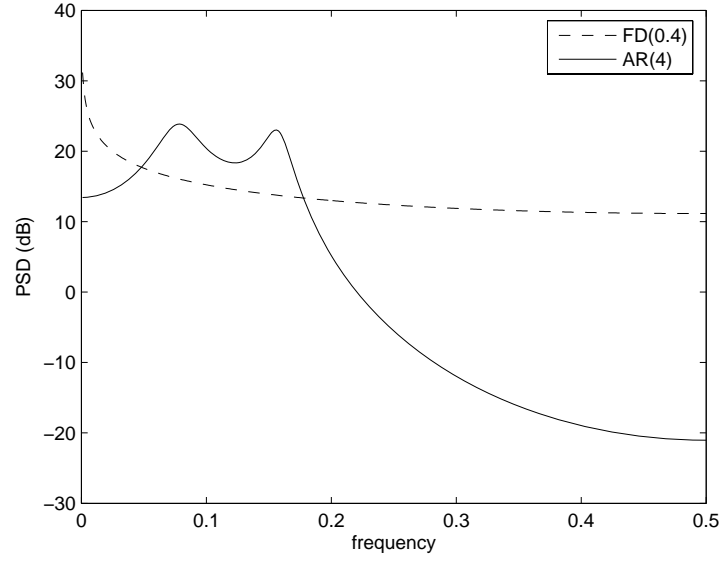


Figura 4.16: DEPs para modelos AR(4) e FD(0,4) de mesma potência.

o decaimento é lento, do tipo hiperbólico). Por outro lado, os parâmetros dos polinômios $\phi(B)$ e $\theta(B)$ são responsáveis pela modelagem da autocorrelação em *lags* de ordens baixas (decaimento rápido do tipo exponencial). A Eq. (4.74) pode ser reescrita na forma AR(∞)

$$\frac{\phi(B)}{\theta(B)} \Delta^d \mathbf{x}_t = \mathbf{w}_t. \quad (4.75)$$

A DEP de um modelo ARFIMA(p, d, q) é dada por [42], [78]

$$P_x(f) = \frac{\sigma_w^2 |1 - e^{-j2\pi f}|^{-2d} |1 - \theta_1 e^{-j2\pi f} - \dots - \theta_q e^{-jq2\pi f}|^2}{|1 - \phi_1 e^{-j2\pi f} - \dots - \phi_p e^{-jp2\pi f}|^2}, \quad (4.76)$$

em que σ_w^2 é a potência de \mathbf{w}_t e $w = 2\pi f$ é a frequência angular normalizada ($-\pi \leq w \leq \pi$). A Eq.(4.76) tem a seguinte forma simplificada para o caso ARFIMA(0, d , 0):

$$P_x(f) = |1 - e^{-j2\pi f}|^{-2d} \sigma_w^2 = [2(1 - \cos(2\pi f))]^{-d} \sigma_w^2 \quad (4.77)$$

ou

$$P_x(f) = [2 \sin(\pi f/2)]^{-2d} \sigma_w^2. \quad (4.78)$$

Como $\sin w \approx w$ para w próximo de zero, então (4.78) reduz-se a

$$P_x(f) = (\pi f)^{-2d} \sigma_w^2. \quad (4.79)$$

Note-se que (4.79) segue o espectro especificado pela Definição 3.1, Eq. (3.3).

4.3.2 Previsão de Modelos ARFIMA

4.3.2.1 Estimação Ótima

Sejam as três formas básicas do modelo ARIMA(p, d, q) (2.66) da seção 2.5.2 no instante $t + h$, em que t é o instante atual e h denota um **horizonte** de previsão:

(a) ARMA($p + d, q$) (similar à Eq. (2.31))

$$\mathbf{x}_{t+h} = \sum_{k=1}^{p+d} \varphi_k \mathbf{x}_{t+h-k} + \mathbf{w}_{t+h} - \sum_{k=1}^q \theta_k \mathbf{w}_{t+h-k}; \quad (4.80)$$

(b) AR(∞)

$$\mathbf{x}_{t+h} = \sum_{k=1}^{\infty} g_k \mathbf{x}_{t+h-k} + \mathbf{w}_{t+h}. \quad (4.81)$$

(c) MA(∞)¹⁹

$$\mathbf{x}_{t+h} = \sum_{k=0}^{\infty} \psi_k \mathbf{w}_{t+h-k} \quad (4.82)$$

A Eq. (4.82) sugere que a previsão do valor futuro de origem t e horizonte $h \geq 1$, denotada por $\hat{\mathbf{x}}_{t+h}$, seja uma combinação linear das inovações $\{\mathbf{w}_t, \mathbf{w}_{t-1}, \mathbf{w}_{t-2}, \dots\}$. Seja

$$\hat{\mathbf{x}}_{t+h} = \psi_h^* \mathbf{w}_t + \psi_{h+1}^* \mathbf{w}_{t-1} + \psi_{h+2}^* \mathbf{w}_{t-2} + \dots \quad (4.83)$$

a previsão de Erro Quadrático Médio Mínimo (EQMM) (*Minimum Mean Squared Error* (MMSE)). Então os coeficientes ψ_{h+k}^* , $k = 0, 1, 2, \dots$, podem ser determinados minimizando-se o Erro Quadrático Médio (EQM) de previsão

$$E[(\mathbf{e}_{t+h})^2] = E[(\mathbf{x}_{t+h} - \hat{\mathbf{x}}_{t+h})^2] = E \left[\left(\sum_{k=0}^{\infty} \psi_k \mathbf{w}_{t+h-k} - \sum_{k=0}^{\infty} \psi_{h+k}^* \mathbf{w}_{t-k} \right)^2 \right]. \quad (4.84)$$

Como

$$\sum_{k=0}^{\infty} \psi_k \mathbf{w}_{t+h-k} = \sum_{k=-h}^{\infty} \psi_{h+k} \mathbf{w}_{t-k},$$

tem-se que \mathbf{e}_{t+h} é dado por

$$\mathbf{e}_{t+h} = \psi_0 \mathbf{w}_{t+h} + \psi_1 \mathbf{w}_{t+h-1} + \dots + \psi_{h-1} \mathbf{w}_{t+1} - \sum_{k=0}^{\infty} (\psi_{h+k} - \psi_{h+k}^*) \mathbf{w}_{t-k}. \quad (4.85)$$

¹⁹A seqüência ψ_t em (4.82) denota a resposta impulsiva do modelo ARIMA. Adotou-se esta nova notação para que o leitor não confunda o horizonte h com um coeficiente $h_k = \psi_k$ da resposta impulsiva.

Deste modo,

$$E[(\mathbf{e}_{t+h})^2] = (1 + \psi_1^2 + \dots + \psi_{h-1}^2)\sigma_{\mathbf{w}}^2 + \sum_{k=0}^{\infty} (\psi_{h+k} - \psi_{h+k}^*)^2 \sigma_{\mathbf{w}}^2, \quad (4.86)$$

em que $\psi_0 = 1$, pois as inovações \mathbf{w}_t são não-correlacionadas. Segue-se que $\psi_{h+k}^* = \psi_{h+k}$ minimiza (4.86).

Portanto, a previsão ótima segundo o critério MMSE é dada por

$$\hat{x}_{t+h} = \psi_h w_t + \psi_{h+1} w_{t-1} + \psi_{h+2} w_{t-2} + \dots = \sum_{k=0}^{\infty} \psi_{h+k} w_{t-k} \quad (4.87)$$

e o erro mínimo de previsão por

$$\mathbf{e}_{t+h} = \mathbf{w}_{t+h} + \psi_1 \mathbf{w}_{t+h-1} + \dots + \psi_{h-1} \mathbf{w}_{t+1} \quad (4.88)$$

possui a variância

$$V_h = (1 + \psi_1^2 + \dots + \psi_{h-1}^2)\sigma_{\mathbf{w}}^2, \quad (4.89)$$

dado que

$$E[\mathbf{e}_{t+h} | \mathcal{F}_t^{(\infty)}] = 0, \quad (4.90)$$

em que $\mathcal{F}_t^{(\infty)} = \{x_t, x_{t-1}, \dots\}$ denota o conjunto de todas as observações passadas da série.

Note-se que

$$\mathbf{x}_{t+h} = \hat{x}_{t+h} + \mathbf{e}_{t+h}, \quad h \geq 1. \quad (4.91)$$

Observe-se que a esperança condicional de \mathbf{x}_{t+h} dadas as observações passadas da série

$$E[\mathbf{x}_{t+h} | \mathcal{F}_t^{(\infty)}] = \hat{x}_{t+h}, \quad (4.92)$$

é igual à previsão MMSE (vide (4.90)) (isto não acontece por acaso). De fato, pode-se demonstrar [47], [124] que o preditor ótimo segundo o critério MMSE: a) é a esperança condicional $E[\mathbf{x}_{t+h} | \mathcal{F}_t^{(\infty)}]$ e b) que esse preditor ótimo é linear quando as inovações são Gaussianas.

4.3.2.2 Formas de Previsão

Tomando-se a esperança condicional em (4.80), obtém-se a previsão via equação de diferenças

$$\begin{aligned} \hat{x}_{t+h} = & \varphi_1 E[\mathbf{x}_{t+h-1} | \mathcal{F}_t^{(\infty)}] + \dots + \varphi_{p+d} E[\mathbf{x}_{t+h-p-d} | \mathcal{F}_t^{(\infty)}] \\ & + E[\mathbf{w}_{t+h} | \mathcal{F}_t^{(\infty)}] - \theta_1 E[\mathbf{w}_{t+h-1} | \mathcal{F}_t^{(\infty)}] - \dots - \theta_q E[\mathbf{w}_{t+h-q} | \mathcal{F}_t^{(\infty)}], \end{aligned} \quad (4.93)$$

para $h \geq 1$. Observe-se que

$$\begin{aligned} E[\mathbf{x}_{t+k}|\mathcal{F}_t^{(\infty)}] &= \hat{x}_{t+k}, \quad k > 0, \\ E[\mathbf{x}_{t+k}|\mathcal{F}_t^{(\infty)}] &= x_{t+k}, \quad k \leq 0, \\ E[\mathbf{w}_{t+k}|\mathcal{F}_t^{(\infty)}] &= 0, \quad k > 0, \\ E[\mathbf{w}_{t+k}|\mathcal{F}_t^{(\infty)}] &= \hat{w}_{t+k}, \quad k \leq 0. \end{aligned} \tag{4.94}$$

Note-se que [42, pág.227]:

- (a) \hat{x}_{t+h} depende de $\hat{x}_{t+h-1}, \hat{x}_{t+h-2}, \dots$, que são calculados de forma recursiva;
- (b) na prática, só se conhece um número finito de observações passadas, ou seja, $\mathcal{F}_t = \{x_t, x_{t-1}, \dots, x_1\}$. Portanto,

$$E[\mathbf{x}_{t+k}|\mathcal{F}_t^{(\infty)}] \approx E[\mathbf{x}_{t+k}|\mathcal{F}_t];$$

- (c) as previsões para um $\text{AR}(p)$ são exatas, pois pode-se mostrar que

$$E[\mathbf{x}_{t+k}|x_t, x_{t-1}, \dots] = E[\mathbf{x}_{t+k}|x_t, \dots, x_{t+1-p}]$$

Tomando-se a esperança condicional em (4.81) obtém-se a previsão por meio da forma $\text{AR}(\infty)$

$$\hat{x}_{t+h} = \sum_{k=1}^{\infty} g_k E[\mathbf{x}_{t+h-k}|\mathcal{F}_t^{(\infty)}] + E[\mathbf{w}_{t+h}|\mathcal{F}_t^{(\infty)}], \tag{4.95}$$

como $E[\mathbf{w}_{t+h}|\mathcal{F}_t^{(\infty)}] = 0$, pode-se reescrever (4.95) na forma

$$\hat{x}_{t+h} = g_1 \hat{x}_{t+h-1} + g_2 \hat{x}_{t+h-2} + \dots + g_h x_t + g_{h+1} x_{t-1} + \dots \quad . \tag{4.96}$$

4.3.2.3 Intervalo de Confiança

O modelo ARFIMA (4.74) assume que a seqüência de inovações \mathbf{w}_t seja um RB Gaussiano de média nula, ou seja, $\mathbf{w}_t \sim \mathcal{N}(0, \sigma_w^2)$. Segue-se então que a distribuição condicional de \mathbf{x}_{t+h} dado $\mathcal{F}_t^{(\infty)}$ é do tipo $\mathcal{N}(\hat{x}_{t+h}, V_h)$ e que

$$Z = \frac{x_{t+h} - \hat{x}_{t+h}}{\sqrt{V_h}} \sim \mathcal{N}(0, 1). \tag{4.97}$$

A expressão do intervalo de confiança para \mathbf{x}_{t+h} , ao nível de confiança $(1-\beta)$, é dada por

$$\hat{x}_{t+h} - z_{\beta/2} \sqrt{V_h} \leq \mathbf{x}_{t+h} \leq \hat{x}_{t+h} + z_{\beta/2} \sqrt{V_h}. \tag{4.98}$$

Como na prática o valor de σ_w^2 é desconhecido, utiliza-se a estimativa

$$\hat{V}_h = (1 + \psi_1^2 + \dots + \psi_{h-1}^2) \hat{\sigma}_w^2, \quad (4.99)$$

obtida na fase de estimação do modelo. Finalmente, obtém-se a expressão final do intervalo de confiança para \mathbf{x}_{t+h}

$$\hat{x}_{t+h} - z_\beta \hat{\sigma}_w \left[1 + \sum_{k=1}^{h-1} \psi_k^2 \right]^{1/2} \leq \mathbf{x}_{t+h} \leq \hat{x}_{t+h} + z_\beta \hat{\sigma}_w \left[1 + \sum_{k=1}^{h-1} \psi_k^2 \right]^{1/2}. \quad (4.100)$$

4.3.2.4 Previsão do ARFIMA

Considere o modelo ARFIMA(p, d, q) estacionário e invertível, $-0,5 < d < 0,5$, dado por (4.75). Pode-se reescrever o processo na forma AR(∞)

$$\sum_{k=0}^{\infty} g_k \mathbf{x}_{t-k} = \mathbf{w}_t, \quad (4.101)$$

em que $g_0 = 1$ e

$$\sum_{k=0}^{\infty} g_k B^k = \phi(B) \theta^{-1}(B) (1 - B)^d = \pi(B). \quad (4.102)$$

Então, pode-se prever um valor futuro de \mathbf{x}_t utilizando-se (4.102) e (4.96) [42, pág.474]. A variância do erro de previsão é dada por (4.89). Note-se que o polinômio $\pi(B)$ é de ordem infinita (pois $|d| < 1/2$). Como na prática o que se tem é uma série com N observações, utilizam-se somente os L primeiros termos $\pi(B)$, com $L < N$. O próximo Capítulo abordará a questão da escolha do valor de L .

4.4 Testes Estatísticos de Memória Longa

4.4.1 Estatística R/S

Seja uma série temporal x_t , $t = 1, 2, \dots, N$. Hurst [71] propôs o teste de memória longa

$$Q_N = \frac{1}{\hat{s}_N} \left[\max_{1 \leq k \leq N} \sum_{j=1}^k (x_j - \bar{x}) - \min_{1 \leq k \leq N} \sum_{j=1}^k (x_j - \bar{x}) \right], \quad (4.103)$$

em que $\hat{s}_N = \sqrt{\hat{C}_0}$, conhecido como estatística **Range Over Standard Deviation** (R/S) ou **rescaled adjusted range** [43], [58].

Hurst observou que o gráfico log-log de R/S (para a série dos níveis anuais

mínimos do Rio Nilo) *versus* N espalhava-se ao longo de uma reta com inclinação superior a $1/2$, ou seja, que $\log(R/S)$ *versus* N apresentava um comportamento do tipo CN^H (efeito de Hurst), em que C é uma constante e $1/2 < H < 1$ denota o parâmetro de Hurst. Esta descoberta empírica contradizia o comportamento esperado para processos Markovianos (que são SRD), em que R/S deve ter um comportamento assintótico do tipo $CN^{1/2}$ [58].

De acordo com [43, pág.262] e [58, pág.82], a estatística $N^{-1/2}Q_N$ converge para uma variável aleatória bem definida (para $N \rightarrow \infty$) quando \mathbf{x}_t é um processo RBG. É por isto que o gráfico log-log de R/S *versus* N apresenta um comportamento assintótico do tipo $CN^{1/2}$. Por outro lado, é a estatística $N^{-H}Q_N$ que converge para uma variável aleatória bem definida quando \mathbf{x}_t é LRD ²⁰.

Posteriormente, Lo [125] mostrou que (4.103) não é robusta quanto à presença de SRD na série e desenvolveu uma versão estendida de (4.103)

$$Q_T = \frac{1}{\hat{\sigma}_{NW}} \left[\max_{1 \leq k \leq N} \sum_{j=1}^k (x_j - \bar{x}) - \min_{1 \leq k \leq N} \sum_{j=1}^k (x_j - \bar{x}) \right], \quad (4.104)$$

em que $\hat{\sigma}_{NW}$ denota a raiz quadrada da estimativa de Newey-West para a variância de longo termo (*long run variance*) de um processo \mathbf{x}_t (estacionário e ergódico) [43, pág.85] [126]. A variância de longo termo é definida como

$$lrv(\mathbf{x}_t) = \sum_{\tau=-\infty}^{\infty} C_{\tau}. \quad (4.105)$$

Como $C_{-\tau} = C_{\tau}$, (4.105) pode ser reescrita como

$$lrv(\mathbf{x}_t) = C_0 + 2 \sum_{\tau=1}^{\infty} C_{\tau}. \quad (4.106)$$

O estimador de Newey-West para (4.105) é dado por

$$\widehat{lrv}_{NW}(x_t) = \hat{C}_0 + 2 \sum_{\tau=1}^T w_{\tau,T} \hat{C}_{\tau}, \quad (4.107)$$

em que os $w_{\tau,T}$ denotam coeficientes (cuja somatória é igual a um) e um parâmetro de truncamento que satisfaz $T = O(N^{1/3})$.

²⁰Para maiores detalhes, vide os Teoremas 4.1 e 4.2 em [58].

4.4.2 Teste GPH

Geweke e Porter-Hudak [127] propuseram um teste de memória longa baseado na DEP do processo ARFIMA(0, d , 0) dada por²¹

$$P_x(f) = [4 \sin^2(\pi f)]^{-d} \sigma_w^2, \quad (4.108)$$

em que σ_w^2 denota a potência do RB w_t . Note-se que o parâmetro d pode ser estimado por meio da seguinte regressão

$$\ln P_x(f_j) = -d \ln[4 \sin^2(\pi f_j)] + 2 \ln \sigma_w, \quad (4.109)$$

para $j = 1, 2, \dots, z(N)$, em que $z(N) = N^\alpha$, $0 < \alpha < 1$ (N denota o número de amostras). Geweke e Porter-Hudak mostraram que se $P_x(f_j)$ for estimado pelo método do periodograma, então o estimador de mínimos quadrados \hat{d} utilizando a regressão (4.109) é normalmente distribuído em grandes amostras se $z(N) = N^\alpha$ com $0 < \alpha < 1$:

$$\hat{d} \sim \mathcal{N} \left(d, \frac{\pi^2}{6 \sum_{j=1}^{z(N)} (U_j - \bar{U})^2} \right), \quad (4.110)$$

em que $U_j = \ln[4 \sin^2(\pi f_j)]$ e \bar{U} é a média amostral dos U_j , $j = 1, 2, \dots, z(N)$. Sob a hipótese nula de que não haja LRD ($d = 0$), a estatística t

$$t_{d=0} = \hat{d} \left(\frac{\pi^2}{6 \sum_{j=1}^{z(N)} (U_j - \bar{U})^2} \right)^{-1/2} \quad (4.111)$$

tem distribuição normal no limite.

4.5 Alguns Métodos para Estimação de H e d

4.5.1 Abordagens Heurísticas

4.5.1.1 Estatística R/S

De acordo com a seção 4.4.1, o gráfico log-log da estatística R/S versus N , em que N denota o número de pontos da série, de uma série LRD se aproxima de uma reta com inclinação $1/2 < H < 1$. Este fato é utilizado pelo método de estimação do parâmetro H denominado análise R/S, que consiste no procedimento descrito a seguir. Primeiramente, calcula-se a estatística R/S usando-se N_1 observações consecutivas da série, em que N_1 deve ser um número suficientemente grande. Em seguida incrementa-se o número de observações por um fator f ; isto é, calcula-se

²¹A Eq. (4.108) é deduzida a partir de (4.76).

R/S sobre $N_i = fN_{i-1}$ amostras consecutivas para $i = 2, \dots, s$. Note-se que para se obter a estatística R/S com N_i observações consecutivas, pode-se dividir a série em $[N/N_i]$ blocos e obter-se $[N/N_i]$ valores, em que $[.]$ denota a parte inteira de um número real. A regressão do *plot* log-log de todas as estatísticas R/S versus N_i , $i = 1, \dots, s$, produz uma estimativa do parâmetro H [43], [58].

4.5.1.2 *Plot* da Variância

O gráfico da variância (*variance plot*) é um método heurístico de estimação do parâmetro de Hurst. Beran [58] mostra que a variância da média amostral de uma série LRD decresce com o seu tamanho N mais lentamente do que no caso tradicional (variáveis independentes ou não-correlacionadas) e da seguinte maneira

$$\text{Var}(\bar{x}) \approx cN^{2H-2}, \quad (4.112)$$

em que $c > 0$. Tem-se os seguintes passos [21]:

1. Seja k um número inteiro. Para diferentes k pertencentes à faixa $2 \leq k \leq N/2$, e para um número m_k suficiente de subséries de tamanho k , calcular as médias de m_k amostras de tamanho k , $\bar{x}_1(k), \bar{x}_2(k), \dots, \bar{x}_{m_k}(k)$ e a média global

$$\bar{x}(k) = m_k^{-1} \sum_{j=1}^{m_k} \bar{x}_j(k). \quad (4.113)$$

2. Para cada k , calcular a variância da amostra de m_k médias de amostras $\bar{x}_j(k)$, $j = 1, 2, \dots, m_k$:

$$s^2(k) = \frac{1}{m_k - 1} \sum_{j=1}^{m_k} (\bar{x}_j(k) - \bar{x}(k))^2. \quad (4.114)$$

3. Representar num gráfico $\log s^2(k)$ versus $\log k$.

Para o caso de dependência de curta duração ou independência, espera-se que o coeficiente angular $2H - 2$ do *plot* seja igual a $(2 \times 1/2) - 2 = -1$.

4.5.1.3 Método do Periodograma

De acordo com a definição (3.1), a DEP de um processo LRD é aproximada pela expressão $C_P|f|^{1-2H}$ quando $f \rightarrow 0$. Como a DEP pode ser estimada pelo

periodograma, um log-log *plot* do periodograma *versus* frequência deve acompanhar uma reta com inclinação $1 - 2H$ para frequências próximas de zero. Este procedimento de estimação é conhecido como método do periodograma.

O estimador $\hat{P}_x(f)$ da DEP é obtido pelo método não-paramétrico²² do periodograma [44], com janelamento de dados (*data tapering*, para redução de vazamento de potência) e suavização (*smoothing*, para redução da variabilidade de $\hat{P}_x(f)$). O periodograma é calculado via²³

$$\hat{P}_x(f) = \frac{1}{N} |X(f)|^2. \quad (4.115)$$

4.5.2 Método de Whittle

O estimador de Whittle também é baseado no periodograma e envolve a minimização da função [128]

$$Q(\theta) = \int_{-0.5}^{0.5} \frac{\hat{P}_x(f)}{P_x(\theta, f)} df \quad (4.116)$$

em que $\hat{P}_x(f)$ denota o periodograma da série x_t , $P_x(\theta, f)$ é a DEP teórica do modelo ARFIMA(p, d, q) x_t na frequência f e $\theta = [p, d, q]$ representa o vetor de parâmetros desconhecidos. Vide [58] para maiores detalhes sobre este método.

4.5.3 Estimador Aproximado de MV de Haslett e Raftery

Considere o modelo ARFIMA (4.74), reescrito na forma

$$\mathbf{x}_t = \Delta^{-d} \phi(B)^{-1} \theta(B) \mathbf{w}_t. \quad (4.117)$$

Sejam \hat{x}_t a previsão ótima de passo-1 de \mathbf{x}_t dadas as observações passadas $\mathcal{F}_{t-1} = \{x_{t-1}, x_{t-2}, \dots, x_1\}$, $e_t = x_t - \hat{x}_t$ o erro de previsão de passo-1 e

$$\zeta = [\sigma_w^2, \phi_1, \dots, \phi_p, d, \theta_1, \dots, \theta_q] \quad (4.118)$$

o vetor de parâmetros do modelo (4.117). Harvey [129, pág.91] mostra que a função de log-verossimilhança de (4.117) é dada por uma expressão conhecida

²²Os métodos paramétricos de análise espectral são baseados em modelos AR, MA e ARMA. Portanto não devem ser aplicados para estimação da DEP de um ruído $1/f^\alpha$.

²³A definição foi dada sem incluir o janelamento e a suavização, para melhor compreensão da natureza essencial do estimador.

como **decomposição do erro de previsão**

$$\log [L(\zeta)|\mathcal{F}_{t-1}] = -\frac{N}{2} \log 2\pi - \frac{N}{2} \log \sigma_w^2 - \frac{1}{2} \sum_{t=1}^N \log f_t - \frac{1}{2\sigma_w^2} \sum_{t=1}^N e_t^2 / f_t, \quad (4.119)$$

em que $f_t = \text{Var}[e_t]/\sigma_w^2$.

Haslett e Raftery [130] propuseram um procedimento rápido para a determinação de uma aproximação de (4.119), que é usada pelo programa **S-PLUS**[®]. O leitor interessado em obter maiores detalhes sobre o algoritmo de Haslett e Raftery deve consultar a seção 4.3 do artigo referenciado acima.

4.5.4 Estimador *Wavelet* de Abry e Veitch

Considere um sinal estacionário $x(t)$, $t \in \mathbb{R}$ e a sua IDWT (4.37)

$$x(t) \approx \sum_k u(J, k) \phi_{J,k}(t) + \sum_{j=1}^J \sum_k w_{j,k} \psi_{j,k}(t).$$

A estacionariedade de x_t implica a estacionariedade das seqüências de coeficientes *wavelet* $\{w_{j,k}\}$ em todas as escalas j . Seja $\mathbb{P}_j = E[w_{j,k}^2] = \text{Var}[w_{j,k}]$ a potência do sinal $\{w_{j,k}\}$ ou **variância *wavelet*** [67, pág.296] numa determinada escala j .

Demonstra-se que [122], [131]

$$\mathbb{P}_j = \int_{\mathbb{R}} |\Psi(\nu)|^2 P_x(\nu/2^j) d\nu, \quad (4.120)$$

em que $\Psi(\nu)$, $-\infty < \nu < \infty$, denota a transformada de Fourier da *wavelet* $\psi(t)$ e $P_x(\nu)$ denota a DEP de $x(t)$. Considere $j \rightarrow \infty$. Então $P_x(\nu/2^j)$ pode ser vista como uma versão magnificada (dilatada) da DEP de $x(t)$ na região de baixas freqüências ($\nu \rightarrow 0$). Sendo assim, depreende-se que (4.120) é uma aproximação da potência de $x(t)$ na região das baixas freqüências para $j \rightarrow \infty$ (pois a integral (4.120) é ponderada pela função $|\Psi(\nu)|^2$).

Suponha que $x(t)$ seja LRD (vide definição 3.1), isto é,

$$P_x(\nu) \sim C_P |\nu|^{-\alpha}, \quad \nu \rightarrow 0, \quad 0 < \alpha < 1, \quad (4.121)$$

em que \sim significa que a razão entre os lados esquerdo e direito de (4.121) converge para 1. As Eqs. (4.120) e (4.121) implicam que a seguinte relação é válida para $j \rightarrow \infty$:

$$\mathbb{P}_j \sim C_P \int_{\mathbb{R}} |\Psi(\nu)|^2 |\nu/2^j|^{-\alpha} d\nu = C_P C 2^{j\alpha}, \quad (4.122)$$

em que $C = C(\Psi, \alpha) = \int_{\mathbb{R}} |\Psi(\nu)|^2 |\nu|^{-\alpha} d\nu$.

A Eq. (4.122) sugere que $H = (1 + \alpha)/2$ pode ser estimado por meio de

$$\log_2(\mathbb{P}_j) \sim (2H - 1)j + \text{constante}, \quad j \rightarrow \infty. \quad (4.123)$$

Dado o sinal amostrado x_k , $k = 0, 1, \dots, N-1$, associado ao sinal original $x(t)$, pode-se estimar $\log_2(\mathbb{P}_j)$ utilizando-se os coeficientes $w_{j,k}$, $k = 0, 1, \dots, N_j - 1$, $j = 1, 2, \dots, J$, da DWT de x_k . O estimador de $\log_2(\mathbb{P}_j)$ é dado por

$$S_j = \log_2 \left(\frac{1}{N_j} \sum_{k=0}^{N_j-1} w_{j,k}^2 \right) \approx \log_2(\mathbb{P}_j). \quad (4.124)$$

O conjunto de estatísticas S_j , $j = 1, 2, \dots, J$, é denominado **espectro *wavelet*** do sinal x_k . A relação (4.123) diz que o espectro *wavelet* de x_k é **linear com inclinação $\alpha = 2H - 1$** nas escalas dilatadas de tempo. A aplicação de uma regressão linear entre as escalas j_1 e j_2 do espectro *wavelet* resulta na seguinte fórmula explícita (Eq. (2.10) do artigo [122]) para estimação de H :

$$\hat{H}_{[j_1, j_2]} = \frac{1}{2} \left[\frac{\sum_{j=j_1}^{j_2} \varepsilon_j j S_j - \sum_{j=j_1}^{j_2} \varepsilon_j j \sum_{j=j_1}^{j_2} \varepsilon_j S_j}{\sum_{j=j_1}^{j_2} \varepsilon_j \sum_{j=j_1}^{j_2} \varepsilon_j j^2 - \left(\sum_{j=j_1}^{j_2} \varepsilon_j j \right)^2} + 1 \right] \quad (4.125)$$

em que $\varepsilon_j = (N \ln^2 2)/2^{j+1}$. A Eq. (4.125) define o estimador *wavelet* de Abry e Veitch (AV).

O estimador *wavelet* $\hat{H}_{[j_1, j_2]}$ apresenta um bom desempenho quando as séries não estão muito distantes de um FGN. Estudos empíricos indicam que o mesmo é robusto quanto a tendências determinísticas suaves e mudanças na estrutura de dependência de curta duração das séries temporais [122],[132],[131].

A discussão desta seção mostrou que o estimador AV é baseado na variância *wavelet* de sinais de tempo contínuo. Abry, Veitch e Taqqu propuseram em [120] um método para que o estimador AV possa ser aplicado a sinais de tempo discreto, tais como sinais de tráfego em redes. O método consiste numa pré-filtragem do sinal de tempo discreto original x_k (“fase de inicialização” da DWT), a qual produz a seqüência inicial (a ser decomposta pela DWT)

$$u_{\tilde{x}}(k) = \int_{-\infty}^{\infty} \tilde{x}_t \phi_0(t - k) dt, \quad (4.126)$$

em que

$$\tilde{x}(t) = \sum_{n=-\infty}^{\infty} x(n) \text{sinc}(t - n). \quad (4.127)$$

Note que $\tilde{x}(t)$ é um sinal de tempo contínuo “fictício” que o procedimento de inicialização da DWT associa ao sinal original x_k e que (4.126) mostra que o estimador AV é calculado por meio da DWT de uma versão filtrada de x_k .

Aplicando-se (4.127) em (4.126) tem-se que

$$\begin{aligned}
 u_{\tilde{x}}(k) &= \int_{-\infty}^{\infty} \tilde{x}_t \phi_0(t - k) dt \\
 &= \sum_{n=-\infty}^{\infty} x(n) \int_{-\infty}^{\infty} \phi_0(t - k) \text{sinc}(t - n) dt \\
 &= \sum_{n=-\infty}^{\infty} x(n) I(k - n) \\
 &= x(k) \star I(k),
 \end{aligned} \tag{4.128}$$

em que

$$I(m) = \int_{-\infty}^{\infty} \text{sinc}(t + m) \phi_0(t) dt. \tag{4.129}$$

A função MATLAB[®] `LDestimate.m` de Veitch e Abry [133] é utilizada nesta tese para a estimação de d pelo método *wavelet* AV.

A Fig. 4.17 ilustra os espectros *wavelet* de um RBG, do modelo AR(4) $\mathbf{x}_t = 2,7607\mathbf{x}_{t-1} - 3,8106\mathbf{x}_{t-2} + 2,6535\mathbf{x}_{t-3} - 0,9238\mathbf{x}_{t-4} + \mathbf{w}_t$ e do *trace BellcoreAug89* (*bin* de 10 milissegundos). O espectro *wavelet* do RBG é plano, ao passo que o espectro do *trace BellcoreAug89* é aproximadamente linear entre as escalas $j = 3$ e $j = 10$. Os *spikes* na escalas 2 e 3 do espectro do modelo AR(4) sugerem a presença de dependência de curta duração. Os parâmetros de Hurst estimados por meio do ajuste da linha vermelha ao espectro *wavelet* do RBG entre as escalas $j = 10, 11, \dots, 20$, da linha vermelha ao espectro do processo AR(4) entre $j = 7, 8, \dots, 12$ e da linha vermelha ao espectro do *trace BellcoreAug89* entre $j = 3, 4, \dots, 12$, são iguais a $\hat{H} \approx 0,5$, $\hat{H} \approx 0,51$ e $\hat{H} \approx 0,814$, respectivamente. A Fig. 4.18 ilustra os periodogramas alisados pelo método *Welch's Overlapped Segment Averaging* (WOSA) da série AR(4) e do *trace BellcoreAug89* da Fig. 4.17.

4.6 Biespectro e Testes de Linearidade

Sejam $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T$ e $\omega = [\omega_1, \omega_2, \dots, \omega_n]^T$, em que T denota a operação de transposição, um vetor aleatório real n -dimensional e um vetor de parâmetros reais com n componentes, respectivamente²⁴. Os momentos conjuntos de ordem

²⁴Admite-se que todos os vetores sejam vetores-coluna, a menos que se diga o contrário.

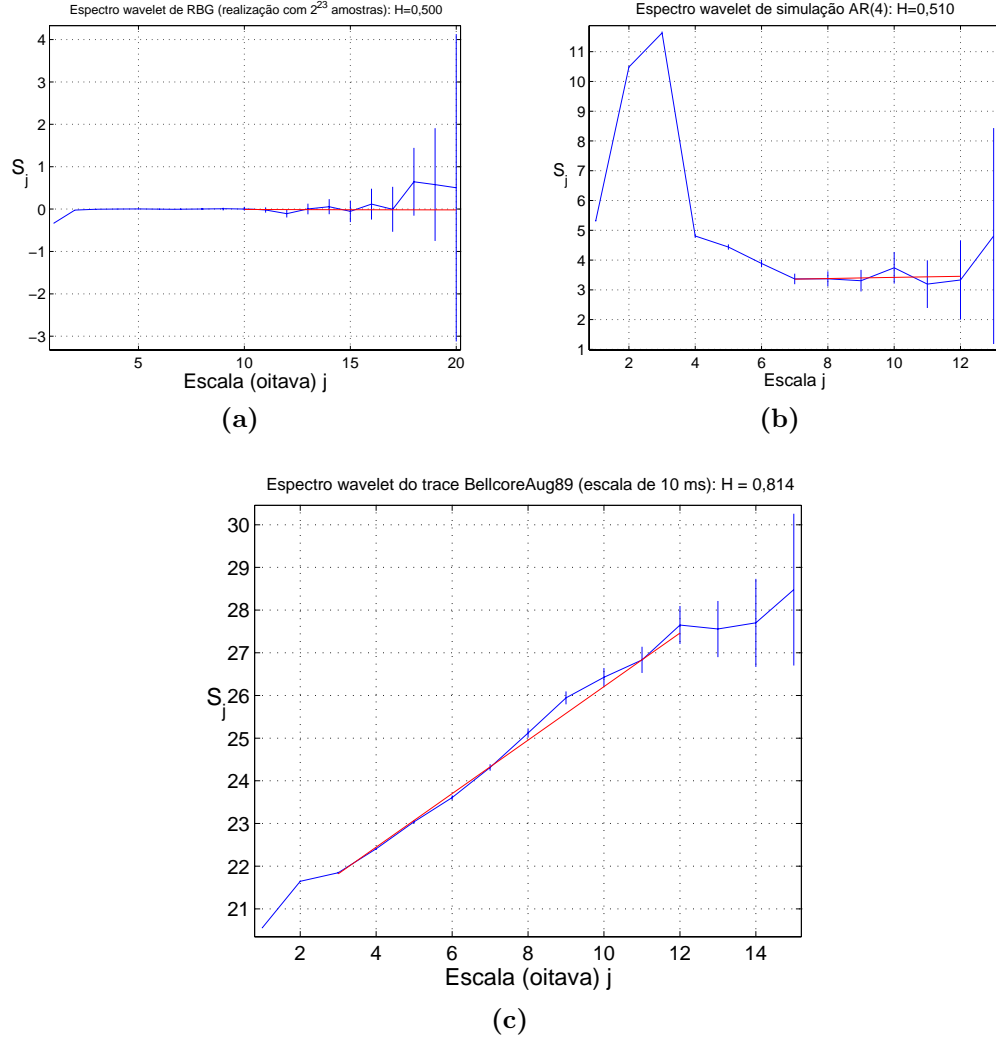


Figura 4.17: Espectros *wavelet* (a) de um RBG, (b) do modelo AR(4) $\mathbf{x}_t = 2,7607\mathbf{x}_{t-1} - 3,8106\mathbf{x}_{t-2} + 2,6535\mathbf{x}_{t-3} - 0,9238\mathbf{x}_{t-4} + \mathbf{w}_t$ e do (c) *trace* BellcoreAug89 (*bin* de 10 milissegundos). Note que o espectro *wavelet* (a) é plano, ao passo que o espectro (c) é aproximadamente linear entre as escalas $j = 3$ e $j = 10$. Os *spikes* na escalas 2 e 3 de (b) sugerem a presença de dependência de curta duração. Os parâmetros de Hurst estimados por meio do ajuste da linha vermelha ao espectro *wavelet* (a) entre as escalas $j = 10, 11, \dots, 20$, da linha vermelha ao espectro (b) entre $j = 7, 8, \dots, 12$ e da linha vermelha ao espectro (c) entre $j = 3, 4, \dots, 12$, são iguais a $\hat{H} \approx 0,5$, $\hat{H} \approx 0,51$ e $\hat{H} \approx 0,814$, respectivamente.

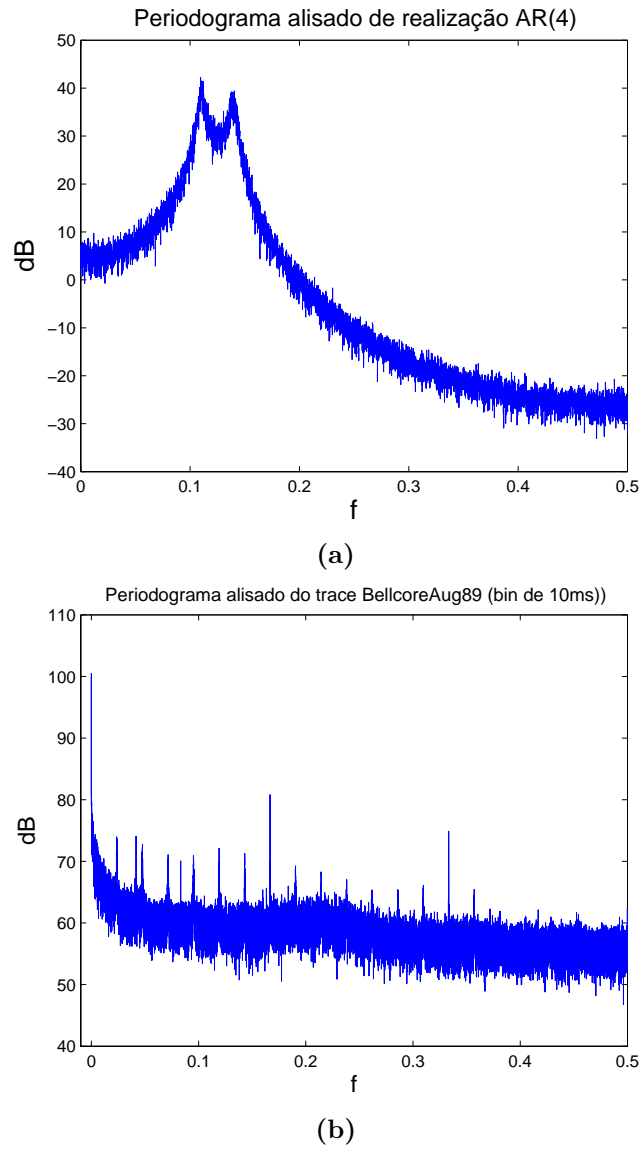


Figura 4.18: Periodogramas alisados pelo método WOSA: (a) série AR(4) da Fig. 4.17 e (b) *trace BellcoreAug89* (*bin* de 10 milissegundos).

$r = k_1 + k_2 + \dots + k_n$ de \mathbf{x} são dados por [46]

$$\text{Mom}\{\mathbf{x}_1^{k_1}, \mathbf{x}_2^{k_2}, \dots, \mathbf{x}_n^{k_n}\} \equiv E\{\mathbf{x}_1^{k_1} \mathbf{x}_2^{k_2} \dots \mathbf{x}_n^{k_n}\} = (-j)^r \frac{\partial^r \Phi_{\mathbf{x}}(\omega^T)}{\partial \omega_1^{k_1} \partial \omega_2^{k_2} \dots \partial \omega_n^{k_n}} \Big|_{\omega_1=\omega_2=\dots=\omega_n=0} \quad (4.130)$$

em que

$$\Phi_{\mathbf{x}}(\omega^T) \equiv E\{e^{j\omega^T \mathbf{x}}\} \quad (4.131)$$

é a função característica conjunta de \mathbf{x} ($E\{\cdot\}$ denota o operador esperança, como visto no Cap. 2).

Os cumulantes conjuntos de ordem r de \mathbf{x} são definidos como [134], [135]

$$\text{Cum}\{\mathbf{x}_1^{k_1}, \mathbf{x}_2^{k_2}, \dots, \mathbf{x}_n^{k_n}\} \equiv (-j)^r \frac{\partial^r \tilde{\Psi}_{\mathbf{x}}(\omega^T)}{\partial \omega_1^{k_1} \partial \omega_2^{k_2} \dots \partial \omega_n^{k_n}} \Big|_{\omega_1=\omega_2=\dots=\omega_n=0}, \quad (4.132)$$

em que

$$\tilde{\Psi}_{\mathbf{x}}(\omega^T) \equiv \ln \Phi_{\mathbf{x}}(\omega^T) \quad (4.133)$$

corresponde ao logaritmo natural da função característica conjunta.

Pode-se verificar que os momentos [134]

$$m_1 = E\{\mathbf{x}\} = \mu \quad m_2 = E\{\mathbf{x}^2\} \quad (4.134)$$

$$m_3 = E\{\mathbf{x}^3\} \quad m_4 = E\{\mathbf{x}^4\} \quad (4.135)$$

de uma variável aleatória \mathbf{x} estão relacionados aos seus cumulantes por

$$c_1 = \text{Cum}\{\mathbf{x}\} = m_1 \quad (4.136)$$

$$c_2 = \text{Cum}\{\mathbf{x}, \mathbf{x}\} = m_2 - m_1^2 \quad (4.137)$$

$$c_3 = \text{Cum}\{\mathbf{x}, \mathbf{x}, \mathbf{x}\} = m_3 - 3m_2m_1 + 2m_1^3 \quad (4.138)$$

$$c_4 = \text{Cum}\{\mathbf{x}, \mathbf{x}, \mathbf{x}, \mathbf{x}\} = m_4 - 4m_3m_1 - 3m_2^2 + 12m_2m_1^2 - 6m_1^4. \quad (4.139)$$

Assuma que os momentos de ordem menor ou igual a n de um processo estacionário $\{\mathbf{x}_k\}$ existam. Então

$$\text{Mom}\{\mathbf{x}(k), \mathbf{x}(k + \tau_1), \dots, \mathbf{x}(k + \tau_{n-1})\} = E\{\mathbf{x}(k) \mathbf{x}(k + \tau_1) \dots \mathbf{x}(k + \tau_{n-1})\} \equiv m_n^{\mathbf{x}}(\tau_1, \tau_2, \dots, \tau_{n-1}) \quad (4.140)$$

em que $\tau_1, \tau_2, \dots, \tau_{n-1}$, $\tau_i = 0, \pm 1, \pm 2, \dots$ para todo i , denotam *lags*. Similarmente, os cumulantes de ordem n de $\{\mathbf{x}_k\}$ podem ser escritos como

$$c_n^{\mathbf{x}}(\tau_1, \tau_2, \dots, \tau_{n-1}) \equiv \text{Cum}\{\mathbf{x}(k), \mathbf{x}(k + \tau_1), \dots, \mathbf{x}(k + \tau_{n-1})\}. \quad (4.141)$$

Nikias *et al* [136] demonstram que as seguintes relações são válidas:

$$c_1^x = m_1^x = \mu \text{ (média)} \quad (4.142)$$

$$c_2^x(\tau_1) = m_2^x(\tau_1) - (m_1^x)^2 \text{ (função de autocovariância)} \quad (4.143)$$

$$c_3^x(\tau_1, \tau_2) = m_3^x(\tau_1, \tau_2) - m_1^x[m_2^x(\tau_1) + m_2^x(\tau_2) + m_2^x(\tau_2 - \tau_1)] + 2(m_1^x)^3 \quad (4.144)$$

e

$$\begin{aligned} c_4^x(\tau_1, \tau_2, \tau_3) = & m_4^x(\tau_1, \tau_2, \tau_3) - m_2^x(\tau_1)m_2^x(\tau_3 - \tau_2) - \\ & - m_2^x(\tau_2)m_2^x(\tau_3 - \tau_1) - m_2^x(\tau_3)m_2^x(\tau_2 - \tau_1) - \\ & - m_1^x[m_3^x(\tau_2 - \tau_1, \tau_3 - \tau_1) + m_3^x(\tau_2, \tau_3) + \\ & + m_3^x(\tau_2, \tau_4) + m_3^x(\tau_1, \tau_2)] + \\ & + (m_1^x)^2[m_2^x(\tau_1) + m_2^x(\tau_2) + m_2^x(\tau_3) + \\ & + m_2^x(\tau_3 - \tau_1) + m_2^x(\tau_3 - \tau_2) + m_2^x(\tau_2 - \tau_1)] - \\ & - 6(m_1^x)^4. \end{aligned} \quad (4.145)$$

Seja $\{\mathbf{x}_k\}$ um processo de média zero, isto é, $m_1^x = 0$. Se $\tau_1 = \tau_2 = \tau_3 = 0$ em (4.143), (4.144), e (4.145) tem-se que

$$\begin{aligned} E\{\mathbf{x}(k)^2\} &= c_2^x(0) = \sigma^2 \text{ (variância)} \\ E\{\mathbf{x}(k)^3\}/\sigma^3 &= c_3^x(0, 0)/\sigma^3 = S(\mathbf{x}_k) \text{ (assimetria)} \\ E\{\mathbf{x}(k)^4\}/\sigma^4 &= c_4^x(0, 0, 0)/\sigma^4 = K(\mathbf{x}_k) \text{ (curtose)}. \end{aligned} \quad (4.146)$$

As fórmulas (4.146) dão a variância, assimetria e curtose em termos de cumulantes.

Sejam $f = [f_1, f_2, \dots, f_n]^T$ e $\tau = [\tau_1, \tau_2, \dots, \tau_n]^T$ vetores de frequências normalizadas e de *lags*, respectivamente. O poliespectro de ordem n do processo \mathbf{x}_k é definido como [136]:

$$P_n^x(f^T) = \sum_{\tau_1=-\infty}^{\infty} \dots \sum_{\tau_{n-1}=-\infty}^{\infty} c_n^x(\tau^T) \exp\{-j2\pi(f^T \tau)\}. \quad (4.147)$$

A DEP, biespectro e triespectro correspondem aos espectros de segunda, terceira e quarta ordens, respectivamente:

$$P_2^x(f) = P^x(f) = \sum_{\tau_1=-\infty}^{\infty} c_2^x(\tau) e^{-j2\pi f \tau}, \quad (4.148)$$

$$P_3^x(f_1, f_2) = \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} c_3^x(\tau_1, \tau_2) e^{-j2\pi(f_1 \tau_1 + f_2 \tau_2)}, \quad (4.149)$$

$$P_4^x(f_1, f_2, f_3) = \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} \sum_{\tau_3=-\infty}^{\infty} c_4^x(\tau_1, \tau_2, \tau_3) e^{-j2\pi(f_1\tau_1 + f_2\tau_2 + f_3\tau_3)}. \quad (4.150)$$

O bispectro e o trispectro são complexos. Outra estatística útil é a bicoerência, definida como

$$B_3^x(f_1, f_2) = \frac{P_3^x(f_1, f_2)}{\sqrt{P^x(f_1 + f_2)P^x(f_1)P^x(f_2)}}. \quad (4.151)$$

Hinich [137] desenvolveu testes estatísticos para Gaussianidade e linearidade. Estes testes são baseados nas seguintes propriedades: 1) se $\{\mathbf{x}_k\}$ é Gaussiano, seus cumulantes de terceira ordem são nulos; portanto, o seu bispectro é nulo e 2) se $\{\mathbf{x}_k\}$ é linear e não-Gaussiano, então a sua bicoerência é uma constante não nula [138, pág.1-22]. Os testes de normalidade e linearidade aplicados nos Caps. 5 e Cap. 6 utilizam o *toolbox* MATLAB “Higher-Order Spectral Analysis” de Swami, Mendel e Nikias [138].

4.7 Teste de Estacionariedade KPSS

Testes de estacionariedade assumem a hipótese nula²⁵ (ou de trabalho) de que a série sob investigação seja do tipo $\mathbf{x}_t \sim I(0)$ [43]²⁶. O teste KPSS, proposto por Kwiatkowski, Phillips, Schmidt e Shin [139], [43], é baseado no modelo

$$\mathbf{y}_t = \beta^T d_t + \mu_t + u_t \quad (4.152)$$

$$\mu_t = \mu_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(0, \sigma_w^2) \quad (4.153)$$

em que β é um vetor de parâmetros, d_t é um vetor de componentes determinísticos (constante ou constante mais tendência) e $u_t \in I(0)$. Observe que μ_t é um passeio aleatório. A hipótese nula de que \mathbf{y}_t seja $I(0)$ é formulada como $H_0 : \sigma_w^2 = 0$, o que implica que μ_t é uma constante. A estatística KPSS para o teste de $\sigma_w^2 = 0$ contra a hipótese alternativa $\sigma_w^2 > 0$ é dada por

$$KPSS = \left(N^{-2} \sum_{t=1}^N \hat{S}_t^2 \right) / \hat{\lambda}^2 \quad (4.154)$$

em que N é o tamanho da amostra, $\hat{S}_t = \sum_{j=1}^t \hat{u}_j$, \hat{u}_t é o resíduo de uma regressão de y_t sobre d_t e $\hat{\lambda}^2$ é uma estimativa da variância de longo termo de u_t usando \hat{u}_t .

²⁵É a hipótese que se deseja rejeitar.

²⁶Testes de raízes unitárias como o teste Dickey-Fuller trabalham com a hipótese nula de que a série seja $I(1)$.

Sob a hipótese nula de que y_t seja I_0 , demonstra-se que $KPSS$ converge para uma função do movimento Browniano que depende da forma dos termos de tendência determinística d_t e que, por outro lado, independe do vetor β .

Parte II

Resultados

5 Modelagem de Teletráfego no Espaço de Estados

Este Capítulo apresenta um novo modelo de teletráfego LRD no espaço de estados. O novo modelo é uma representação finito-dimensional, isto é, truncada, do processo ARFIMA. Uma das vantagens da modelagem no espaço de estados é a possibilidade de utilização do filtro de Kalman para previsão *on-line* de valores futuros de sinais de teletráfego. Além disso, compara-se, por meio de simulações realizadas nos programas MATLAB e S-PLUS[®], as previsões obtidas pelo novo modelo com as que são produzidas por modelos AR de ordem alta. Os códigos-fonte para MATLAB e S+FinMetrics¹ encontram-se listados nos Apêndices A e B.

5.1 Introdução

O tráfego agregado de dados com comportamento LRD é freqüentemente modelado pelo processo ARFIMA [140], [141], [142], [143], [144]. O processo ARFIMA consegue modelar as autocorrelações empíricas do tráfego das redes para pequenos e grandes *lags*, diferentemente de um modelo não-paramétrico como o FGN [85], que só consegue modelar as correlações para grandes valores de *lag*.

A utilização da técnica de modelagem no estado de espaços em conjunto com a filtragem de Kalman [145] tem atraído a atenção dos pesquisadores da área de séries financeiras nos últimos anos [42] ([39] oferece uma boa visão geral do assunto, enquanto que [146] contém uma exposição detalhada da matéria), porque a abordagem de espaço de estados é flexível e mais genérica do que o sistema clássico ARIMA proposto por Box e Jenkins [146].

A incorporação de um modelo de espaço de estados no filtro de Kalman permite que: a) o procedimento de estimação do modelo seja eficiente e b) que sejam feitas as operações de filtragem, previsão e suavização (*smoothing*) do vetor

¹Versão do programa S-PLUS[®] com o módulo FinMetrics de funções estatísticas para análise de séries temporais financeiras.

de estados dadas as observações passadas da série. Quando se considera a questão da previsão *on-line* de tráfego, a vantagem da abordagem no espaço de estados consiste em permitir que as previsões sejam implementadas de forma recursiva, por meio do preditor de Kalman (sendo portanto adequada para um esquema de previsão *on-line*). Não obstante, cabe chamar a atenção para o seguinte fato: a modelagem no espaço de estados de um processo ARFIMA não é trivial, uma vez que a representação de um modelo LRD no espaço de estados é infinito-dimensional (vide (4.70), (4.71), (4.72) e (4.73)). Sendo assim, justifica-se a introdução de uma aproximação finito-dimensional do modelo ARFIMA.

Bhansali and Kokoszka [147] apontam dois tipos de métodos de previsão de séries temporais:

- Tipo-I: em que as previsões são feitas sem que seja preciso estimar o parâmetro H ;
- Tipo-II: em que primeiramente deve-se estimar o parâmetro H do modelo LRD.

Este Capítulo investiga o uso dos métodos acima mencionados e adota o erro de previsão como métrica de avaliação de desempenho. Para o método Tipo-I, postula-se que uma série LRD seja gerada por um modelo AR de ordem alta e obtêm-se as previsões por meio do filtro de Kalman. Para o método Tipo-II, assume-se que uma série LRD seja modelada pelo processo ARFIMA e desenvolve-se uma representação finito-dimensional no espaço de estados denominada TARFIMA (*Truncated* ARFIMA). Também são obtidas previsões para múltiplos passos via filtro de Kalman. É interessante observar que, até onde o autor desta tese conseguiu pesquisar, a literatura registra somente alguns artigos sobre o tópico da previsão de Kalman de teletráfego agregado, tais como [148], [149] e [150]. Contudo, nenhum destes estudos considera a utilização de modelos LRD de tráfego, o que corrobora a originalidade deste trabalho. De fato, o TARFIMA é um novo modelo de tráfego LRD, que tem como objetivo viabilizar esquemas de previsão *on-line* de teletráfego heterogêneo [106]. O diferencial do TARFIMA é a sua possibilidade de implementação prática e de, em quase todos os casos, ser tão bom quanto os modelos de memória infinita.

5.2 Modelagem no Espaço de Estados

5.2.1 Modelo TARFIMA

Considere o modelo ARFIMA(p, d, q) (4.74), que pode ser reescrito como

$$\frac{(1-B)^d \phi(B)}{\theta(B)} \mathbf{x}_t = \mathbf{w}_t. \quad (5.1)$$

Como o polinômio $\frac{(1-B)^d \phi(B)}{\theta(B)}$ (lado esquerdo de (5.1)) é infinito, o modelo (5.1) pode ser escrito em termos de um AR(∞). Na prática, a previsão a partir da representação AR(∞) é feita truncando-se o modelo AR(∞) num modelo AR(L), em que L é um valor suficientemente grande (esse valor é obtido empiricamente, conforme será mostrado mais adiante), ou seja, L é um valor que produz uma boa aproximação TARFIMA:

$$\frac{(1-B)^d \phi(B)}{\theta(B)} \approx 1 - \psi_1 B - \psi_2 B^2 - \dots - \psi_L B^L. \quad (5.2)$$

Portanto, uma representação estacionária truncada AR(L) de um modelo ARFIMA \mathbf{x}_t satisfaz a equação de diferenças

$$\psi(B) \mathbf{x}_t \approx \mathbf{w}_t, \quad (5.3)$$

em que $\psi(B) = 1 - \psi_1 B - \psi_2 B^2 - \dots - \psi_L B^L$.

Todo modelo linear de séries temporais tem representação em espaço de estados [42, pág.367]. O modelo de espaço de estados linear Gaussiano geral (*linear Gaussian state-space model* (LGSSM)) pode ser escrito como o sistema de equações² [39, pág. 508]

$$\underset{m \times 1}{\mathbf{x}_{t+1}} = \underset{m \times 1}{d_t} + \underset{m \times m}{T_t} \cdot \underset{m \times 1}{\mathbf{x}_t} + \underset{m \times r}{H_t} \cdot \underset{r \times 1}{\boldsymbol{\eta}_t}, \quad (5.4)$$

$$\underset{l \times 1}{\mathbf{y}_t} = \underset{l \times 1}{c_t} + \underset{l \times m}{Z_t} \cdot \underset{m \times 1}{\mathbf{x}_t} + \underset{l \times 1}{\boldsymbol{\varepsilon}_t}, \quad (5.5)$$

²As seguintes regras de notação serão formalmente adotadas a partir deste ponto: a) \mathbf{X} denota uma matriz com quantidades aleatórias, X é uma matriz determinística, \mathbf{x} pode denotar um vetor aleatório ou uma variável aleatória (neste caso o contexto esclarecerá qual é a grandeza em questão) e x pode ser um vetor determinístico ou uma observação, e b) todos os vetores, a menos que se diga o contrário, são vetores-coluna, como $\mathbf{y}_t = [\mathbf{y}_t, \mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-N+1}]^T$, em que T denota a operação de transposição.

em que $t = 1, 2, \dots, N$ e

$$\mathbf{x}_1 \sim \mathcal{N}(\mu_{1|0}, \Sigma_{1|0}) \quad (\text{estado inicial}), \quad (5.6)$$

$$\boldsymbol{\eta}_t \sim \mathcal{N}(0, I_r) \quad (\text{RBG}), \quad (5.7)$$

$$\boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, I_l) \quad (\text{RBG}). \quad (5.8)$$

e assume-se que

$$E[\boldsymbol{\varepsilon}_t \boldsymbol{\eta}_t^T] = \mathbf{0}. \quad (5.9)$$

A **equação do estado** (ou de **transição**) (5.4) descreve a evolução do vetor de estados \mathbf{x}_t no tempo usando uma estrutura Markoviana de primeira ordem. A **equação da observação** descreve o vetor de observações \mathbf{y}_t em termos do vetor de estados \mathbf{x}_t e de um vetor $\boldsymbol{\varepsilon}_t$ (ruído de medição do tipo RBG) com matriz de covariâncias I_l . Assume-se que o RBG $\boldsymbol{\eta}_t$ (inovação) com matriz de covariâncias I_r em (5.4) seja não-correlacionado com $\boldsymbol{\varepsilon}_t$. As matrizes determinísticas T_t, H_t, Z_t são denominadas **matrizes do sistema** e os vetores d_t e c_t contêm componentes fixas e podem ser usados para incorporar efeitos conhecidos ou padrões no modelo (caso contrário eles são iguais a zero). Para modelos univariados, $l = 1$ e, conseqüentemente, Z_t é um vetor linha.

Há diversas maneiras de se mapear (ou transformar) o modelo AR truncado

$$\mathbf{y}_t = \phi_1 \mathbf{y}_{t-1} + \dots + \phi_L \mathbf{y}_{t-L} + \boldsymbol{\xi}_t, \quad (5.10)$$

em que $\boldsymbol{\xi}_t$ denota a inovação, numa forma de espaço de estados [39]. Nesta tese, adota-se a transformação proposta por Harvey em [129]. Desta forma, (5.10) pode ser posta numa forma de espaço de estados, que será denominada **modelo de espaço de estados TARFIMA**, com as equações do estado e da observação

$$\mathbf{x}_{t+1} = T \mathbf{x}_t + H \boldsymbol{\xi}_t, \quad \boldsymbol{\xi}_t \sim \mathcal{N}(0, \sigma_{\boldsymbol{\xi}}^2) \quad (5.11)$$

$$\mathbf{y}_t = Z \mathbf{x}_t, \quad (5.12)$$

em que

$$T = \begin{pmatrix} \phi_1 & 1 & 0 & \dots & 0 \\ \phi_2 & 0 & 1 & & 0 \\ \vdots & & & \ddots & \vdots \\ \phi_{L-1} & 0 & 0 & & 1 \\ \phi_L & 0 & 0 & \dots & \end{pmatrix}, \quad H = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad Z = (1 \quad 0 \quad \dots \quad 0 \quad 0). \quad (5.13)$$

O vetor de estados tem a forma

$$\mathbf{x}_t = \begin{pmatrix} \mathbf{y}_t \\ \phi_2 \mathbf{y}_{t-1} + \dots + \phi_L \mathbf{y}_{t-L+1} \\ \phi_3 \mathbf{y}_{t-1} + \dots + \phi_L \mathbf{y}_{t-L+2} \\ \vdots \\ \phi_L \mathbf{y}_{t-1} \end{pmatrix}. \quad (5.14)$$

O modelo descrito por (5.11) e (5.12) não possui erro de medição³. Note-se que os coeficientes do modelo AR(L) constituem a primeira coluna da matriz de sistema T (e esta é uma vantagem da transformação de Harvey).

5.2.2 Previsão Multipasso com o Filtro de Kalman

Conforme visto no Cap. 4, a teoria de estimação linear afirma que a estimativa de EQMM (ou MMSE) $\hat{\mathbf{x}}_t$ do estado \mathbf{x}_t em (5.11), dadas as observações passadas

$$\mathcal{F}_t = [y_t, y_{t-1}, \dots, y_1],$$

corresponde à média condicional $E\{\mathbf{x}_t | \mathcal{F}_t\}$, denotada por $x_{t|t}$ [47]. O objetivo do algoritmo do filtro de Kalman para o modelo de espaço de estados definido por (5.11) e (5.12) é obter estimativas recursivas da média condicional $x_{t+1|t}$ e da matriz de covariâncias $\Sigma_{t+1|t}$ do vetor de estados \mathbf{x}_{t+1} dados \mathcal{F}_t e o modelo, ou seja

$$x_{t+1|t} = E\{T\mathbf{x}_t + H\xi_t | \mathcal{F}_t\} = Tx_{t|t}, \quad (5.15)$$

$$\Sigma_{t+1|t} = \text{Var}\{T\mathbf{x}_t + H\xi_t | \mathcal{F}_t\} = T\Sigma_{t|t}T^T + \sigma_\xi^2 HH^T. \quad (5.16)$$

Sejam

- $y_{t|t-1}$ a média condicional de \mathbf{y}_t dado \mathcal{F}_{t-1} ,
- $\tilde{\mathbf{y}}_t = (\mathbf{y}_t - y_{t|t-1})$ o erro de previsão de passo-1 de \mathbf{y}_t dado \mathcal{F}_{t-1} ,
- $\sigma_t^2 = \text{Var}\{\tilde{\mathbf{y}}_t | \mathcal{F}_{t-1}\}$ a variância do erro de previsão de passo-1,
- $C_t = \text{Cov}(\mathbf{x}_t, \tilde{\mathbf{y}}_t | \mathcal{F}_{t-1})$, e
- $K_t = TC_t/\sigma_t^2 = T\Sigma_{t|t-1}Z^T/\sigma_t^2$ o ganho de Kalman.

³O erro de medição é desprezado nesta tese. Entretanto, a medição de sinais de tráfego é “ruidosa”. Vide, por exemplo, a coluna “% Loss” da Tabela 2 do estudo [70] (taxa de perda de pacotes associada à coleta de *traces* variando de 0,01% a 0,14%).

As equações do algoritmo do filtro de Kalman são dadas por (veja [39], [146] e [129] para maiores detalhes):

$$\left\{ \begin{array}{lcl} \tilde{\mathbf{y}}_t & = & \mathbf{y}_t - Z\mathbf{x}_{t|t-1}, \\ \sigma_t^2 & = & Z\Sigma_{t|t-1}Z^T, \\ K_t & = & T\Sigma_{t|t-1}Z^T/\sigma_t^2, \\ L_t & = & T - K_tZ, \\ \mathbf{x}_{t+1|t} & = & T\mathbf{x}_{t|t-1} + K_t\sigma_t^2, \\ \Sigma_{t+1|t} & = & T\Sigma_{t|t-1}L_t^T + \sigma_{\xi}^2HH^T, \quad t = 1, \dots, N. \end{array} \right. \quad (5.17)$$

Como o modelo de espaço de estados dado por (5.11) e (5.12) é estacionário, pode-se provar que as matrizes $\Sigma_{t|t-1}$ convergem para uma matriz constante Σ_* . As grandezas σ_t^2 , K_t , and $\Sigma_{t+1|t}$ tornam-se constantes quando o regime permanente é alcançado.

Seja t a origem da previsão e que se queira prever \mathbf{y}_{t+h} para $h = 1, 2, \dots, H$. As equações de previsão do filtro de Kalman (as duas últimas equações de (5.17)) produzem previsões de passo-1. Em geral, as previsões $\hat{\mathbf{y}}_h$, $h = 1, \dots, H$, podem ser obtidas por meio das equações de previsão do filtro de Kalman estendendo-se o conjunto de dados \mathcal{F}_t com um conjunto de valores faltantes (ou seja, a previsão de passo-1 é usada para se obter a de passo-2, as previsões de passo-1 e de passo-2 são utilizadas para se obter a de passo-3 e assim sucessivamente) [146].

5.3 O Poder de Previsão do Modelo TARFIMA

Conforme visto na seção 3.3.1, autocorrelações que decaem lentamente para zero dificultam a estimação de constantes tais como a média de um processo estacionário. Por outro lado, o grau de previsibilidade de uma série aumenta com o aumento da dependência temporal entre as observações [58]. Portanto, boas previsões de curto e de longo prazo podem ser obtidas para processos LRD quando se tem acesso a um grande número de amostras passadas e se os estimadores dos parâmetros p , d e q do modelo forem consistentes e de baixo viés.

A idéia básica do modelo TARFIMA é fazer um bom uso da estrutura de autocorrelação das séries LRD. O restante desta seção compara o poder de previsão dos modelos TARFIMA(L) e ARFIMA($0, d, 0$) (cujo esquema de previsão demandaria um memória infinita, sendo portanto irrealizável). Para tal, adota-se

a métrica

$$P_h^2 = \frac{\sigma_x^2 - EQM_h}{\sigma_x^2} = 1 - \frac{EQM_h}{\sigma_x^2}, \quad (5.18)$$

em que EQM_h denota o EQM de previsão de passo- h (ou variância do erro de previsão de passo- h , uma vez que $EQM_h = E[(\mathbf{e}_{t+h})^2] = E[(\mathbf{x}_{t+h} - \hat{\mathbf{x}}_{t+h})^2]$), a qual é uma medida do poder de previsibilidade de um determinado modelo. Para uma previsão perfeita, $EQM_h = 0$ e $P_h^2 = 1$. Por outro lado, se \mathbf{x}_{t+h} não depende das amostras passadas, então a melhor previsão é a média não-condicional da série. Neste caso, a previsão não melhora quando se utilizam observações passadas. Portanto, $EQM_h = \sigma_x^2$ e $P_h^2 = 0$. Note-se também que $P_h^2 \rightarrow 0$ quando $h \rightarrow \infty$. Isto reflete o fato de que previsões de prazos muito longos se tornam cada vez mais difíceis [58].

Considere o modelo ARFIMA(0, d , 0)

$$\mathbf{x}_t = (1 - B)^{-d} \mathbf{w}_t$$

em que a inovação \mathbf{w}_t tem potência unitária ($\sigma_w^2 = 1$). Neste caso, demonstra-se que a variância de \mathbf{x}_t é dada por (vide Teorema 1 de [82])

$$\sigma_x^2 = \frac{\Gamma(1 - 2d)}{\Gamma^2(1 - d)}. \quad (5.19)$$

Também pode-se demonstrar que a variância de \mathbf{x}_t pode ser calculada pela fórmula alternativa [40, pág.91]

$$\sigma_x^2 = 1 + \sum_{j=1}^{\infty} \psi_j^2, \quad (5.20)$$

em que os ψ_j^2 denotam os coeficientes da representação MA(∞) de \mathbf{x}_t . Note-se que a fórmula (5.20) pode ser aplicada para o cálculo aproximado da variância do modelo TARFIMA(L). Para tal, deve-se obter uma representação truncada MA(L') do TARFIMA(L), em que L' é uma ordem suficientemente alta (igual a 100, por exemplo).

Tabela 5.1: Valores de P_h^2 para modelos AR(1) e TARFIMA(L), $L = \{1, 10\}$. Os valores das correlações de $lag-1$ do AR(1) foram escolhidos de tal forma que $\rho_1 = 1/9$ corresponde à correlação de $lag-1$ do ARFIMA(0; 0, 1; 0) e $\rho_1 = 2/3$ corresponde à correlação de $lag-1$ do ARFIMA(0; 0, 4; 0).

h	AR(1)		$L = 1$		$L = 10$	
	$\rho_1 = 1/9$	$\rho_1 = 2/3$	$d = 0, 1$	$d = 0, 4$	$d = 0, 1$	$d = 0, 4$
1	0.0123	0.4444	0,01	0,16	0,0172	0,3249
10	0	0.0003	0	0	0,0002	0,0380
20	0	0	0	0	0	0,0038
100	0	0	0	0	0	0

Tabela 5.2: Valores de P_h^2 para modelos TARFIMA(L), $L = \{100, 150\}$ e ARFIMA($0, d, 0$) (coluna em que $L = \infty$).

h	$L = 100$		$L = 150$		$L = \infty$	
	$d = 0, 1$	$d = 0, 4$	$d = 0, 1$	$d = 0, 4$	$d = 0, 1$	$d = 0, 4$
1	0,0188	0,4095	0,0189	0,4161	0,0191	0,5169
10	0,0019	0,1585	0,0020	0,1680	0,0022	0,3116
20	0,0009	0,1082	0,0010	0,1182	0,0013	0,2704
100	0	0,0167	0,0001	0,0277	0,0003	0,1955

Valores numéricos de P_h^2 , $h = 1, 10, 20, 100$, são dados pelas Tabelas 5.1 e 5.2. A Tabela 5.1 mostra os valores de P_h^2 para modelos AR(1) e TARFIMA(L), $L = \{1, 10\}$. Os valores das correlações de $lag-1$ do modelo AR(1)

$$\mathbf{x}_t - \phi_1 \mathbf{x}_{t-1} = \mathbf{w}_t$$

foram escolhidos de tal forma que $\rho_1 = 1/9$ e $\rho_1 = 2/3$ correspondam às correlações de $lag-1$ dos processos ARFIMA($0; 0, 1; 0$) e ARFIMA($0; 0, 4; 0$), respectivamente. A previsão ótima de \mathbf{x}_{t+h} dado $\mathcal{F}_t = \{\mathbf{x}_t, \mathbf{x}_{t-1}\}$, ou seja, a previsão que se obtém utilizando-se o modelo AR(1) é dada por⁴

$$\hat{x}_{t+h} = (\rho_x(1))^h x_t = (\phi_1)^h x_t \quad (5.21)$$

e a variância do erro de previsão de passo- h por [58, pág.168]

$$EQM_h^{AR(1)} = \sigma_x^2 [1 - (\rho_x(1))^{2h}]. \quad (5.22)$$

Os resultados das Tabelas 5.1 e 5.2 ilustram que a precisão das previsões com o modelo TARFIMA(L) melhoram consideravelmente com o aumento de L (ordem do truncamento) e de d . Este efeito é bastante pronunciado para previsões de longo prazo e para valores grandes de d (como 0,4, por exemplo). Os resultados tabelados também indicam que a previsão é mais precisa quando se faz um uso adequado da estrutura de correlação das séries e que os intervalos de confiança das previsões podem ser reduzidos quando se aumenta a ordem do truncamento do modelo TARFIMA. Note-se que o procedimento (simplista) de previsão baseado no modelo AR(1) revela-se bastante impreciso. Para $h = 10$, as Tabelas 5.1 e 5.2 mostram que a variância do erro de previsão do modelo TARFIMA(150) para $d = 0, 4$ é 16,8% menor do que a variância do erro da previsão feita com o modelo AR(1).

As Tabelas 5.1 e 5.2 também mostram que a previsão de longo prazo é mais confiável na presença de LRD. Beran [58] demonstra que P_h^2 decai exponencial-

⁴De acordo com (2.48), $\rho_x(1) = \phi_1$.

mente para zero no caso do modelo AR(1) e que P_h^2 decai hiperbolicamente (ou seja, lentamente) para zero para o ARFIMA(0, d , 0).

5.4 Análise Exploratória de Séries Simuladas e da Série do Rio Nilo

Nesta seção, é feita a análise exploratória de duas realizações aproximadas do processo ARFIMA(0, d , 0), $d = \{0, 3; 0, 4\}$ (Fig. 5.1), e de duas realizações MWM(H), $H = \{0, 8; 0, 9\}$ (Fig. 5.2), produzidas pelo gerador de sinais LRD baseado em *wavelets*⁵ desenvolvido por Mello, Lima, Lipas e Amazonas [77], [78] e da série dos níveis mínimos anuais do rio Nilo entre os anos de 1007 a 1206 (Fig. 5.3). As séries simuladas e a série do rio Nilo⁶ serão utilizadas como “séries de treinamento” do modelo TARFIMA, o qual será aplicado a um sinal de teletráfego no próximo Capítulo.

O parâmetro d das séries foi estimado pelos seguintes métodos:

- análise R/S (vide seção 4.5.1.1);
- regressão de mínimos quadrados no domínio da frequência baseada no periodograma (vide seção 4.5.1.3);
- Whittle (vide seção 4.5.2);
- estimador aproximado de MV Haslett-Raftery (vide seção 4.5.3);
- estimador *wavelet* AV (vide seção 4.5.4).

Também foram utilizadas as seguintes estatísticas ou testes:

- SACF (2.25);
- periodogramas alisados pelos métodos WOSA $\hat{P}_x^{\text{WOSA}}(f)$ e da janela de autocorrelação de Daniell $\hat{P}_x^{lw}(f)$ (vide seção 4.5.1.3);
- testes estatísticos de memória longa R/S e GPH (vide seção 4.4);
- testes de normalidade e de linearidade de Hinich baseados no conceito de biespectro (vide seção 4.6).

⁵Vide funções `Generate.m`, `Model.m` e `Recon.m` que estão listadas no Apêndice A.

⁶Esta série empírica já foi bastante investigada na prática, sendo bem modelada pelo processo ARFIMA [58].

A normalidade das distribuições marginais das séries e dos resíduos dos modelos estimados também foi investigada por meio de histogramas e de gráficos *Quantile-Quantile plot* (QQ-plot). Um QQ-plot é um gráfico dos quantis⁷ da normal padrão no eixo horizontal *vs* os valores ordenados da série no eixo vertical [83].

Foram estimados modelos $AR(p)$ (para a implementação de métodos de previsão Tipo-I) e $ARFIMA(p, d, q)$ (para previsão Tipo-II por meio do modelo $TARFIMA(L)$). Os diagnósticos dos modelos ajustados foram feitos por meio da análise da SACF e do QQ-plot dos resíduos.

5.4.1 Série ARFIMA(0; 0, 4; 0)

Estimou-se um modelo $ARFIMA(0; 0, 3684; 0)$ por meio da função **S+FinMetrics FARIMA** [43, pág.271]. A função **FARIMA** implementa um procedimento rápido de estimação do parâmetro d baseado no método aproximado de estimação por MV proposto por Haslett e Raftery (vide seção 4.5.3) e ajusta os parâmetros p e q com base na seleção de valores $p \leq p_{\max}$ e $q \leq q_{\max}$ que minimizam o BIC (utilizou-se $p_{\max} = q_{\max} = 2$)⁸. A Tabela 5.3 resume as estimativas do parâmetro d para a série $ARFIMA(0; 0, 4; 0)$ obtidas pelos métodos R/S, periodograma (com janela de Daniell), Whittle, Haslett-Raftery e Abry-Veitch. Note que o resultado da análise R/S é discordante dos demais, os quais estão situados na faixa $0,36 \leq \hat{d} \leq 0,38$.

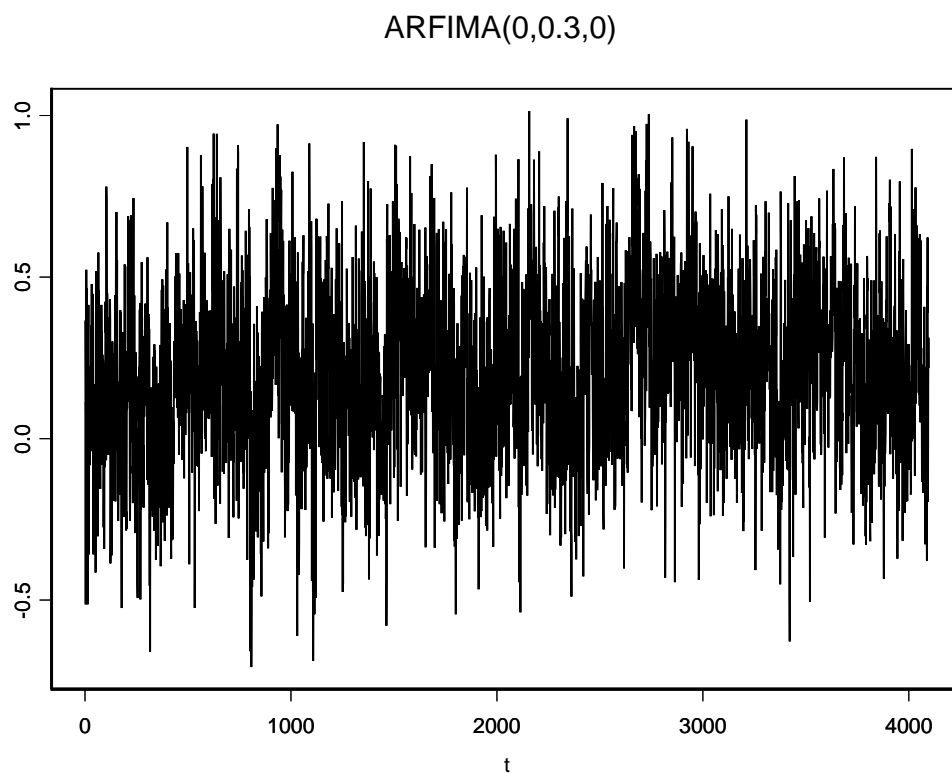
Tabela 5.3: Estimativas do parâmetro d para a realização $ARFIMA(0; 0, 4; 0)$.

análise	periodograma	estimador	estimador de MV	estimador
R/S	com janela	de Whittle	de Haslett-Raftery	<i>wavelet</i>
	de Daniell			Abry/Veitch
0,2735	0,3696	0,3773	0,3684	0,380

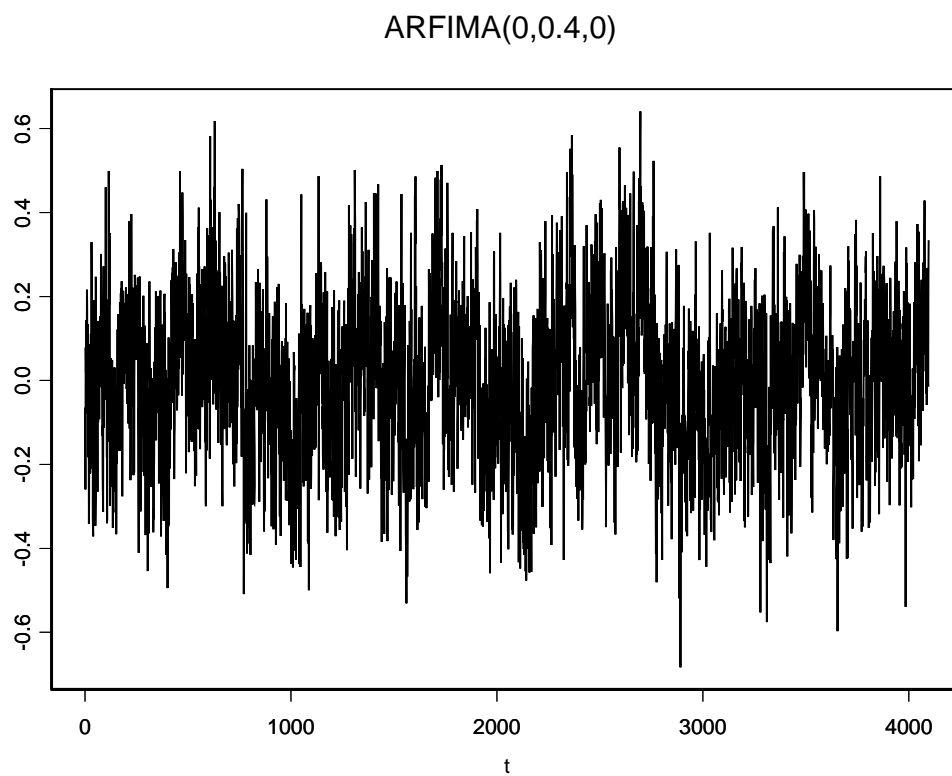
A Tabela 5.4 mostra que o teste KPSS não consegue rejeitar a hipótese nula (H_0 : a série é estacionária), porque $p_{af} > 0,05$ (probabilidade de “alarme falso” maior do que o nível de significância $\alpha = 0,05$), conforme indicado pelo “SIM” na coluna KPSS. O teste R/S não rejeita a hipótese nula de série SRD (H_0 : a série não possui memória longa), ao passo que o teste GPH sim. Os testes de Hinich não conseguem rejeitar as hipóteses nulas de normalidade (H_0 : série com biespectro nulo) e linearidade (H_0 : série com bicoerência constante). O número

⁷O valor x_α em que a função de distribuição de probabilidades de uma variável aleatória \mathbf{x} tem o valor α é denominado α -quantil ou 100α percentil: $F_{\mathbf{x}}(x_\alpha) = \alpha$. A mediana, por exemplo, é o 0,5-quantil de uma distribuição de probabilidades.

⁸Este tipo de ajuste é adotado para todas as séries deste Capítulo.

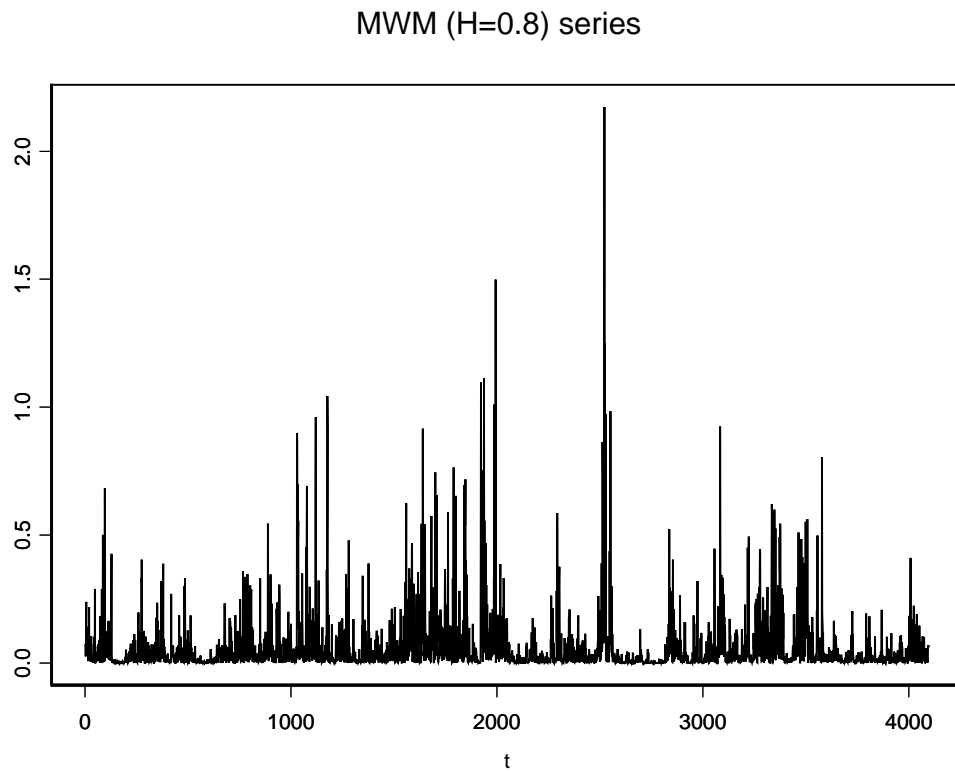


(a)

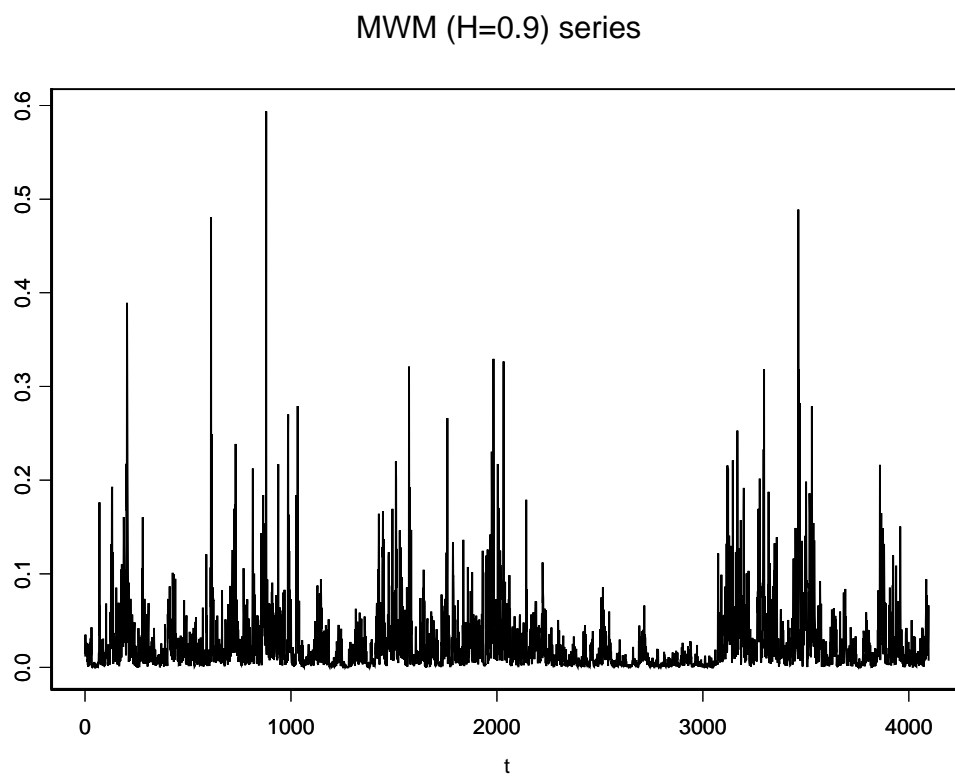


(b)

Figura 5.1: Simulações: (a) ARFIMA(0; 0, 3; 0) e (b) ARFIMA(0; 0, 4; 0).



(a)



(b)

Figura 5.2: Realizações: (a) $\text{MWM}(H = 0,8)$ e (b) $\text{MWM}(H = 0,9)$.

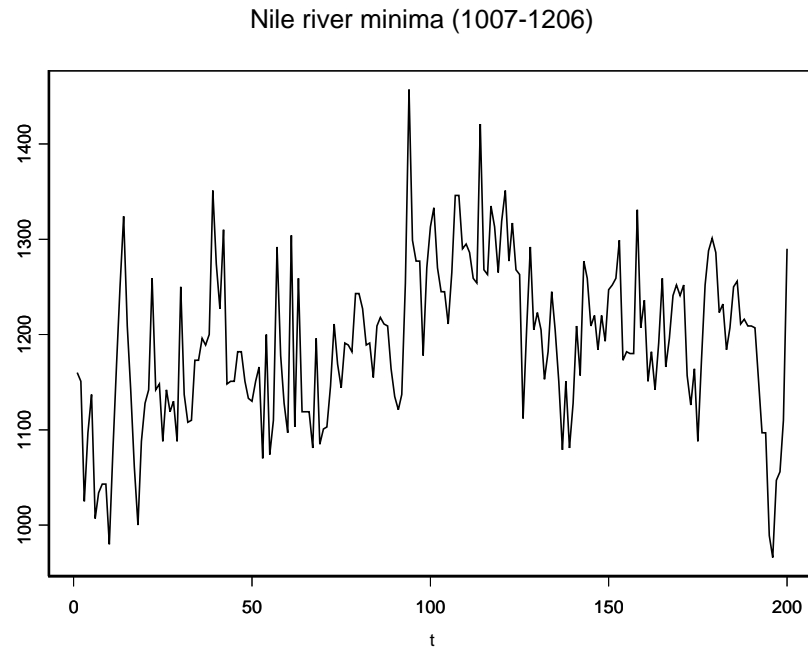


Figura 5.3: Série do rio Nilo entre os anos de 1007 e 1206.

estimado de raízes unitárias⁹ (coluna \hat{m}) pela função **S-PLUS**[®] **FARIMA** é igual a zero.

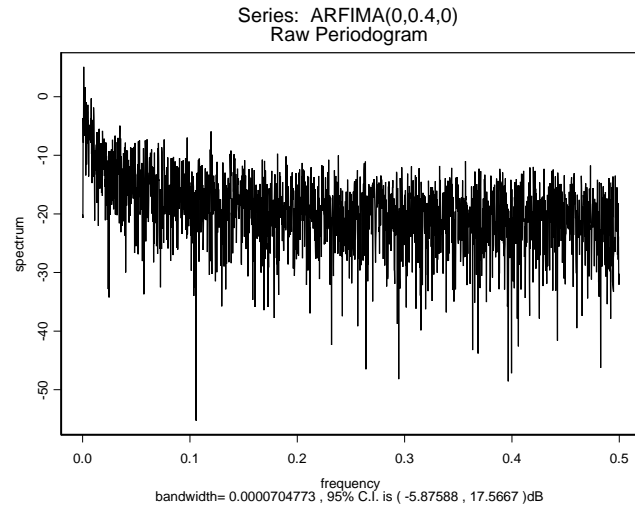
Tabela 5.4: Série ARFIMA(0; 0, 4; 0): testes de estacionariedade, memória longa, normalidade e linearidade, e número de raízes unitárias (coluna \hat{m}).

x_t é $I(0)$?	x_t é LRD?		Hinich		\hat{m}
KPSS	R/S	GPH	x_t é normal?	x_t é linear?	
SIM	NÃO	SIM	SIM	SIM	0
$p_{af} > 0,05$	$p_{af} > 0,05$	$p_{af} \leq 0,01$	$p_{af} = 0,5213$		

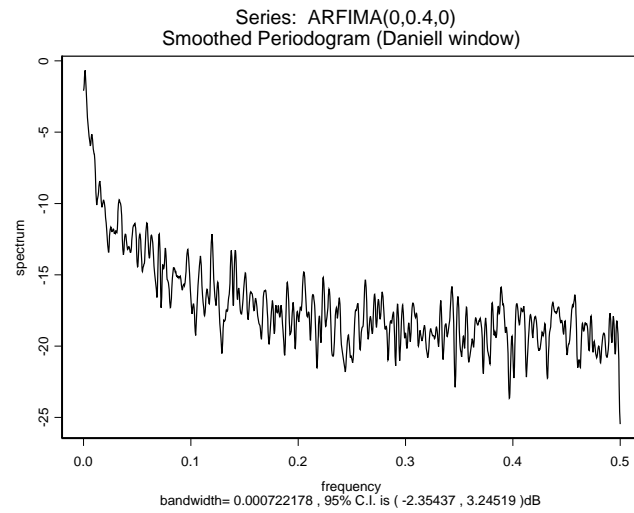
A Fig. 5.4 ilustra o periodograma (Fig. 5.4a), periodograma alisado com janela de autocorrelação Daniell [44, pág.264]) (Fig. 5.4b) e o periodograma alisado pelo método WOSA (Fig. 5.4c) [44, pág.289]. Observe que os periodogramas alisados sugerem que o espectro da série ARFIMA(0; 0, 4; 0) possui uma singularidade do tipo $1/f^{2H-1}$ na origem. O comportamento LRD da série é confirmado pelo espectro *wavelet* da Fig. 5.5, que é aproximadamente linear.

A Fig. 5.6 mostra o QQ-plot da série ARFIMA(0; 0, 4; 0) (Fig. 5.6a) e um gráfico com a SACF e ACFs teóricas dos modelos ARFIMA(0; 0, 3684; 0) e AR(15) estimados (Fig. 5.6b). O processo AR(15) é o melhor modelo AR de acordo com o critério AIC e foi estimado pela função **S-PLUS**[®] **AR** (método de Yule-Walker [151, pág.94]). O QQ-plot indica que a série é Gaussiana. O gráfico da Fig. 5.6b

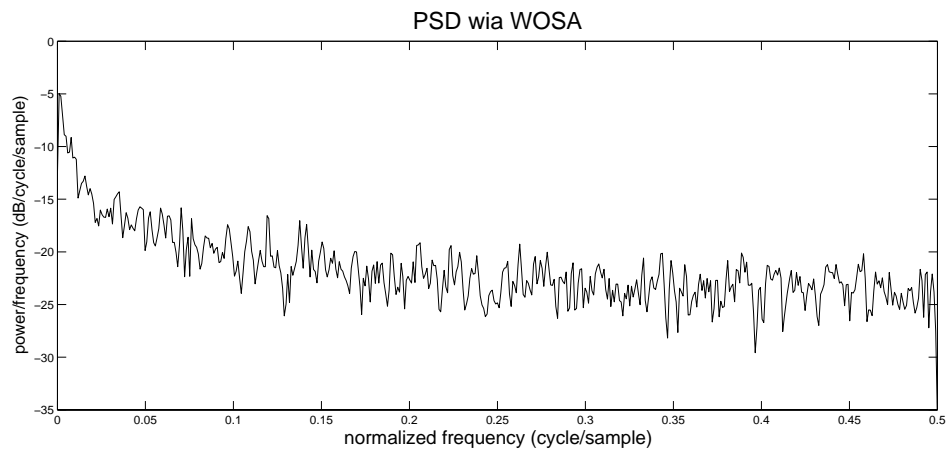
⁹Número de pólos no círculo de raio unitário.



(a)



(b)



(c)

Figura 5.4: Simulação ARFIMA(0; 0, 4; 0): (a) periodograma, (b) periodograma alisado com janela de Daniell e (c) periodograma estimado pelo método WOSA.

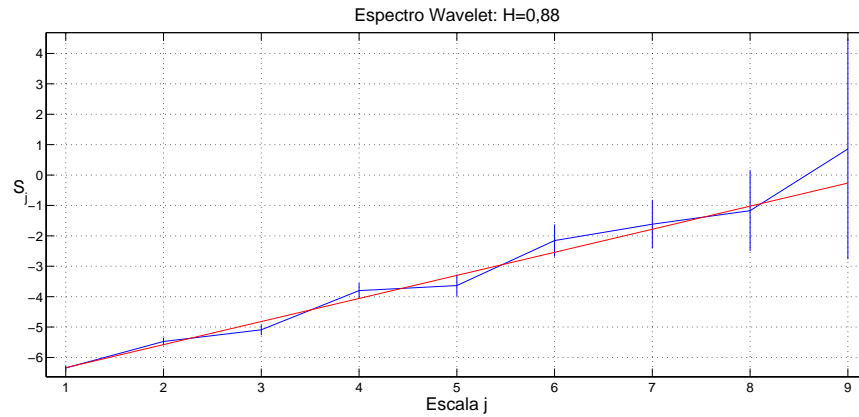


Figura 5.5: Espectro *wavelet* do sinal ARFIMA(0; 0, 4; 0).

mostra que o processo AR(15) não modela o decaimento lento da SACF da série. A Fig. 5.7 contém os gráficos do QQ-plot (Fig. 5.7a) e SACF (Fig. 5.6b) dos resíduos do modelo ARFIMA(0; 0, 3684; 0) estimado. Essas Figs. sugerem que a série dos resíduos é um RBG.

Finalmente, é curioso observar na Fig. 5.8, que ilustra o diagrama de pólos e zeros do modelo AR(15) estimado, que um dos pólos está bastante próximo do círculo unitário (pólo real positivo). Portanto, o processo AR de ordem alta ajustado consegue modelar a memória longa da série ARFIMA(0; 0, 4; 0) às custas do posicionamento de um pólo nas proximidades do círculo unitário.

5.4.2 Série MWM com $H = 0,9$

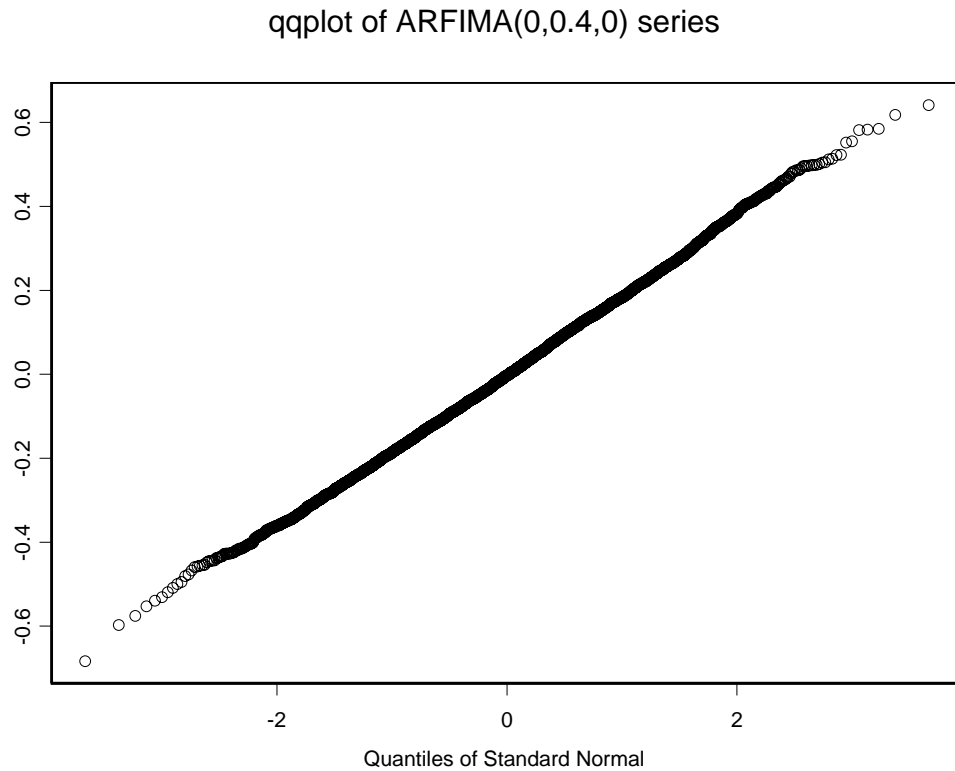
A função `S+FinMetrics FARIMA` ajustou¹⁰ um modelo ARFIMA(0; 0, 3572; 0) à série MWM($H = 0,9$). A Tabela 5.5 resume as estimativas do parâmetro d para a série MWM($H = 0,9$) obtidas pelos métodos R/S, periodograma (com janela de Daniell), Whittle, Haslett-Raftery e Abry-Veitch.

Tabela 5.5: Estimativas do parâmetro H para a realização MWM($H = 0,9$).

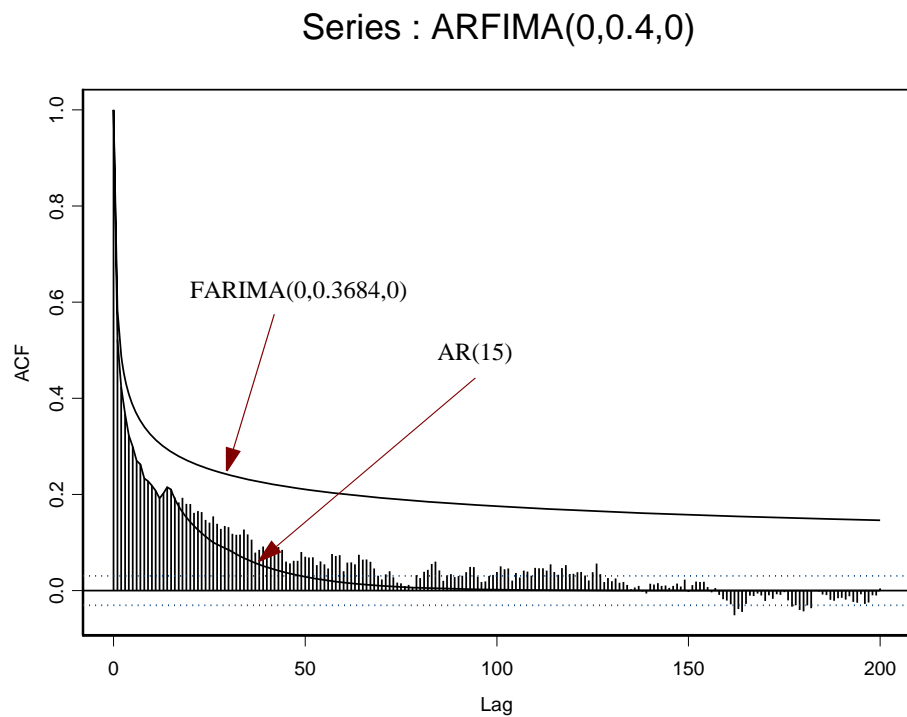
análise R/S	periodograma com janela de Daniell	estimador de Whittle	estimador de MV de Haslett-Raftery	estimador <i>wavelet</i> Abry/Veitch
0,3492	0,3267	0,3428	0,3572	0,362

A Tabela 5.6 mostra os resultados dos testes estatísticos e o número estimado de raízes unitárias da série MWM($H = 0,9$). O teste KPSS rejeitou a

¹⁰A função `FARIMA` utilizou os argumentos $p_{\max} = 2$ e $q = 0$ porque a estimação com $p_{\max} = q_{\max} = 2$ resultou num modelo ARFIMA(2, 0, 3719, 2), cuja ACF teórica não modelava o decaimento lento da SACF da série MWM com $H = 0,9$.

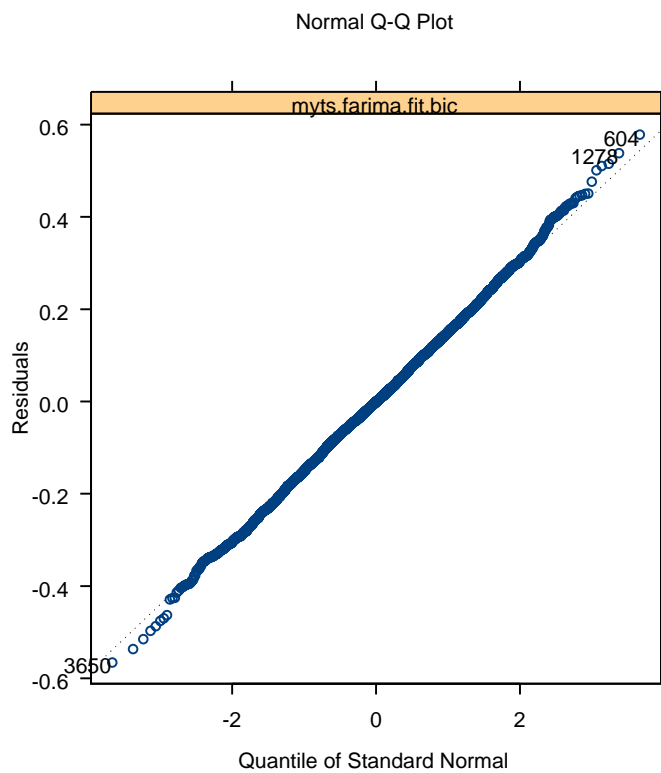


(a)

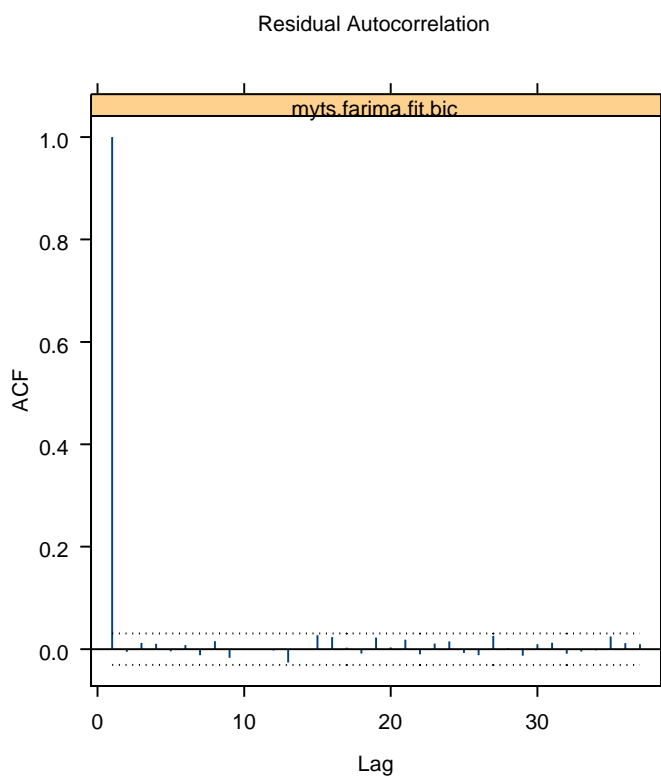


(b)

Figura 5.6: Simulação ARFIMA(0; 0, 4; 0) : (a) QQ-plot (gráfico dos quantis da normal padrão no eixo horizontal *vs* valores ordenados da simulação no eixo vertical) e (b) Gráfico com a SACF e ACFs teóricas dos modelos ARFIMA(0; 0, 3684; 0) e AR(15) estimados.



(a)



(b)

Figura 5.7: Resíduos do modelo ARFIMA(0; 0, 3684; 0) estimado para a série ARFIMA(0; 0, 4; 0): (a) QQ-plot e (b) SACF.

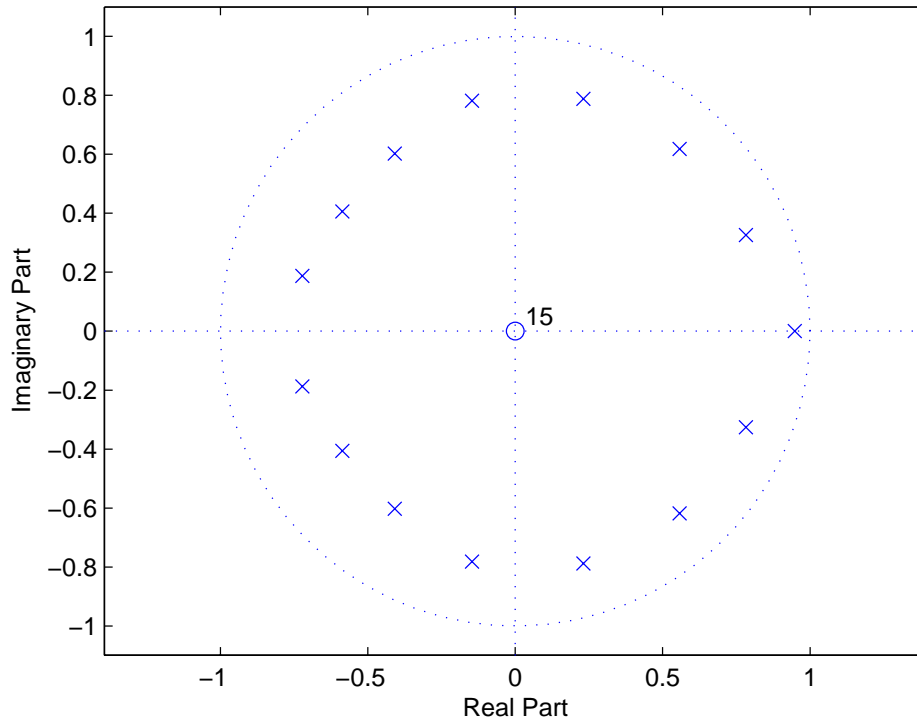


Figura 5.8: Diagrama de pólos e zeros do modelo AR(15).

estacionariedade dessa realização (o MWM é um processo estacionário de primeira ordem, mas não é estacionário de segunda ordem, conforme [11]). A série é LRD de acordo com os testes R/S e GPH. Os testes de Hinich afirmam que a série é não-linear, conforme é esperado para o modelo MWM, que é baseado numa cascata multiplicativa no domínio *wavelet*. O número estimado de raízes unitárias é igual a zero.

Tabela 5.6: Série MWM($H = 0,9$): resultados dos testes estatísticos e número estimado de raízes unitárias.

x_t é $I(0)$? KPSS	x_t é LRD?		Hinich		\hat{m}
	R/S	GPH	x_t é normal?	x_t é linear?	
NAO	SIM	SIM	NAO	NAO	0
$p_{af} \leq 0,05$	$p_{af} \leq 0,01$	$p_{af} \leq 0,01$	$p_{af} = 0$		

A Fig. 5.9 ilustra o periodograma (Fig. 5.9a), periodograma alisado com janela de Daniell (Fig. 5.9b) e o periodograma alisado pelo método WOSA (Fig. 5.9c). Os periodogramas suavizados mostram que o espectro da série MWM($H = 0,9$) possui uma singularidade do tipo $1/f^{2H-1}$ na origem. O comportamento LRD da série é confirmado pelo espectro *wavelet* aproximadamente linear da Fig. 5.10.

A Fig. 5.11 mostra o QQ-plot da série MWM($H = 0,9$) (Fig. 5.11a) e

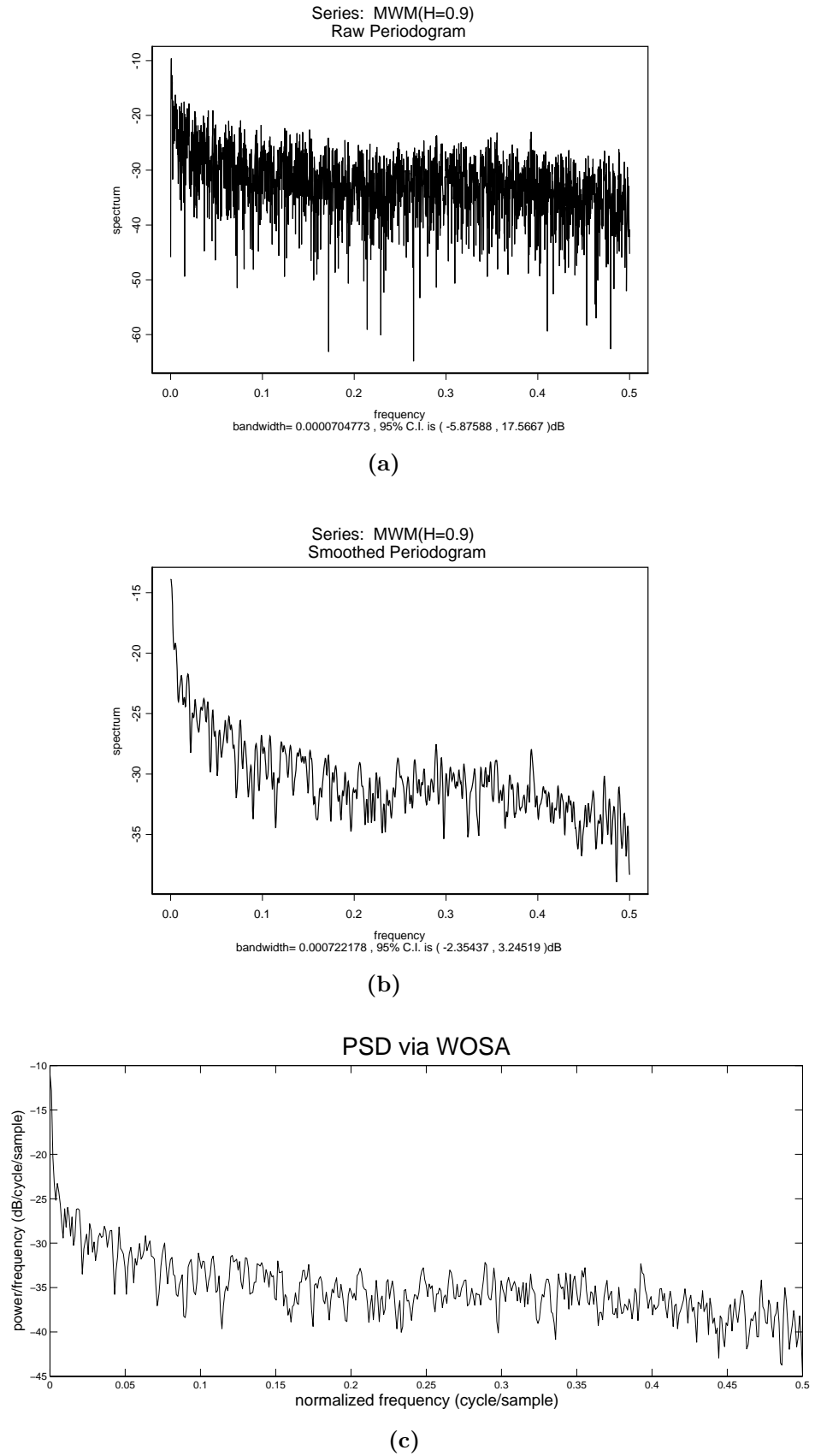


Figura 5.9: Simulação MWM($H = 0,9$): (a) periodograma, (b) periodograma alisado com janela de autocorrelação de Daniell e (c) periodograma estimado pelo método WOSA.

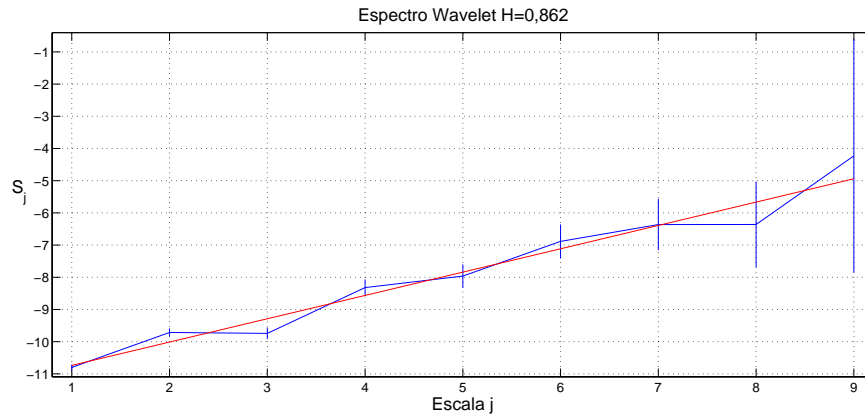


Figura 5.10: Espectro *wavelet* do sinal MWM($H = 0,9$).

um gráfico com a SACF e ACFs teóricas dos modelos ARFIMA(0; 0, 3572; 0) e AR(28) estimados (Fig. 5.11b). O processo AR(28) é o melhor modelo AR de acordo com o critério AIC. O QQ-plot indica que a série não é Gaussiana. A Fig. 5.11b mostra que o decaimento lento da SACF da série é modelada pela ACF teórica do processo ARFIMA(0; 0, 3572; 0) estimado.

A Fig. 5.12 contém os gráficos do QQ-plot (Fig. 5.12a) e SACF (Fig. 5.12b) dos resíduos do modelo ARFIMA(0; 0, 3572; 0) estimado. Essas Figs. sugerem que a série dos resíduos é um RB não-Gaussiano. A Fig. 5.13 mostra o diagrama de pólos e zeros do modelo AR(28) ajustado. Observe que o pólo real positivo está muito próximo do círculo unitário.

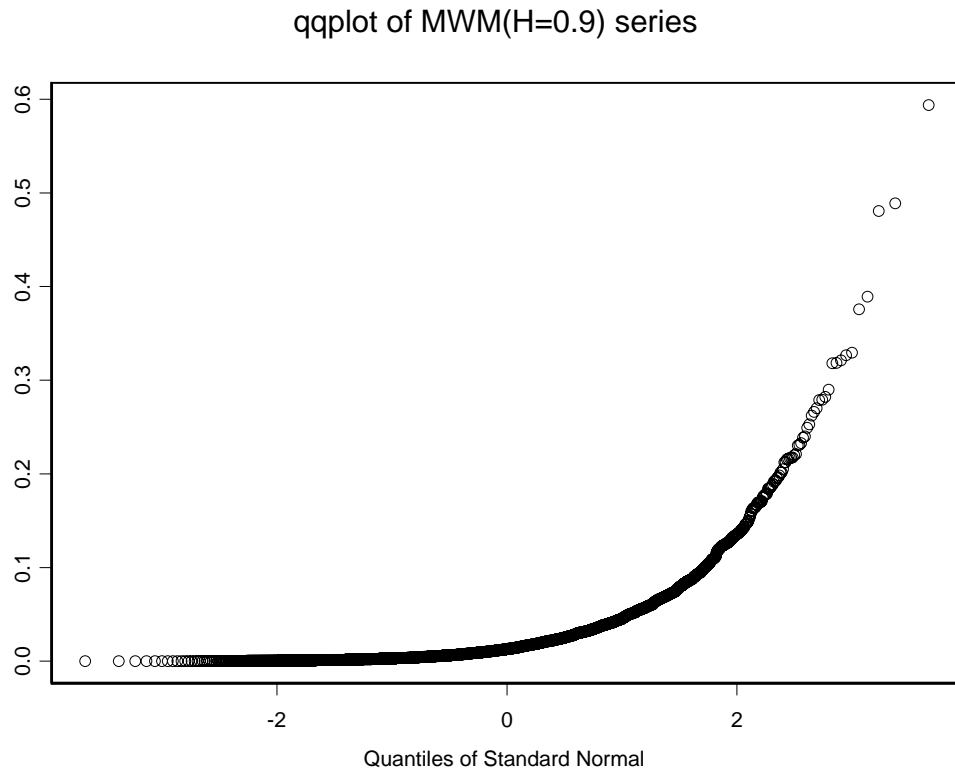
5.4.3 Série ARFIMA(0; 0, 3; 0)

A função `S+FinMetrics FARIMA` ajustou um modelo ARFIMA(0; 0, 2895; 0) à série ARFIMA(0; 0, 3; 0). A Tabela 5.7 resume as estimativas do parâmetro d obtidas pelos métodos R/S, periodograma com janela de Daniell, Whittle, Haslett-Raftery e Abry-Veitch. A Tabela 5.7 mostra que a estimativa pela análise R/S ($\hat{d} = 0,2422$) é discordante das demais, que estão situadas na faixa $0,28 \leq \hat{d} \leq 0,32$.

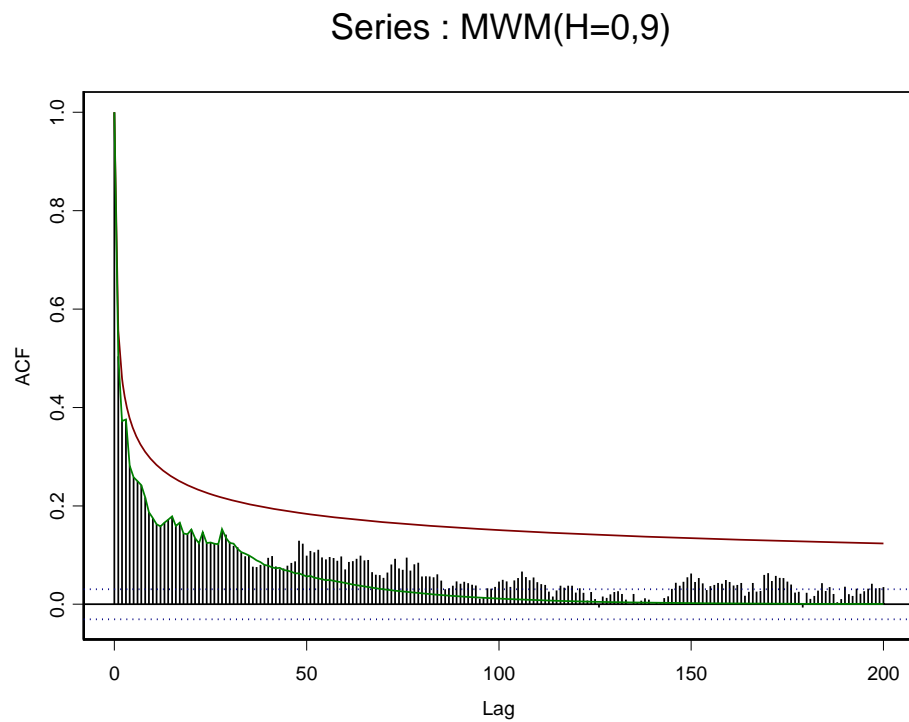
Tabela 5.7: Estimativas do parâmetro d para a realização ARFIMA(0; 0, 3; 0).

análise R/S	periodograma com janela de Daniell	estimador de Whittle	estimador de MV de Haslett-Raftery	estimador <i>wavelet</i> Abry/Veitch
0,2422	0,2917	0,2889	0,2895	0,313

A Tabela 5.8 mostra os resultados dos testes estatísticos e o número estimado de raízes unitárias da série ARFIMA(0; 0, 3; 0). O teste KPSS rejeitou a hipótese de estacionariedade da série. A série é LRD de acordo com os testes R/S e GPH.

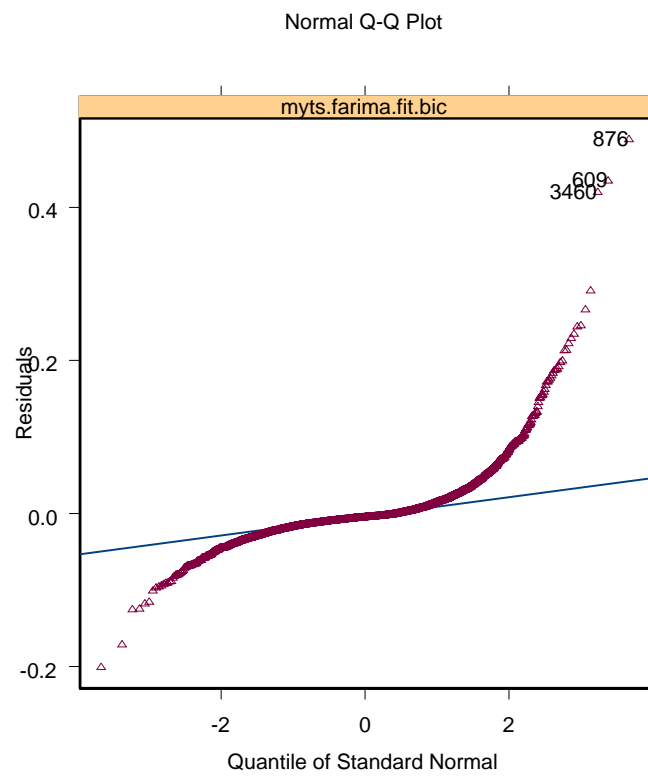


(a)

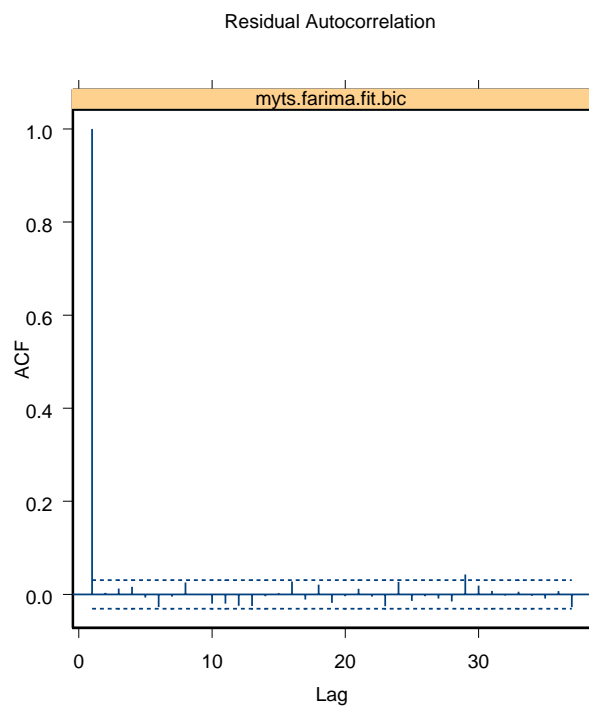


(b)

Figura 5.11: Simulação MWM($H = 0,9$) : (a) QQ-plot e (b) Gráfico com a SACF e ACFs teóricas dos modelos ARFIMA(0; 0, 3572; 0) (linha de cor vermelha) e AR(28) (linha de cor verde) estimados.



(a)



(b)

Figura 5.12: Resíduos do modelo ARFIMA(0;0,3572;0) estimado para a série MWM($H = 0,9$): (a) QQ-plot e (b) SACF.

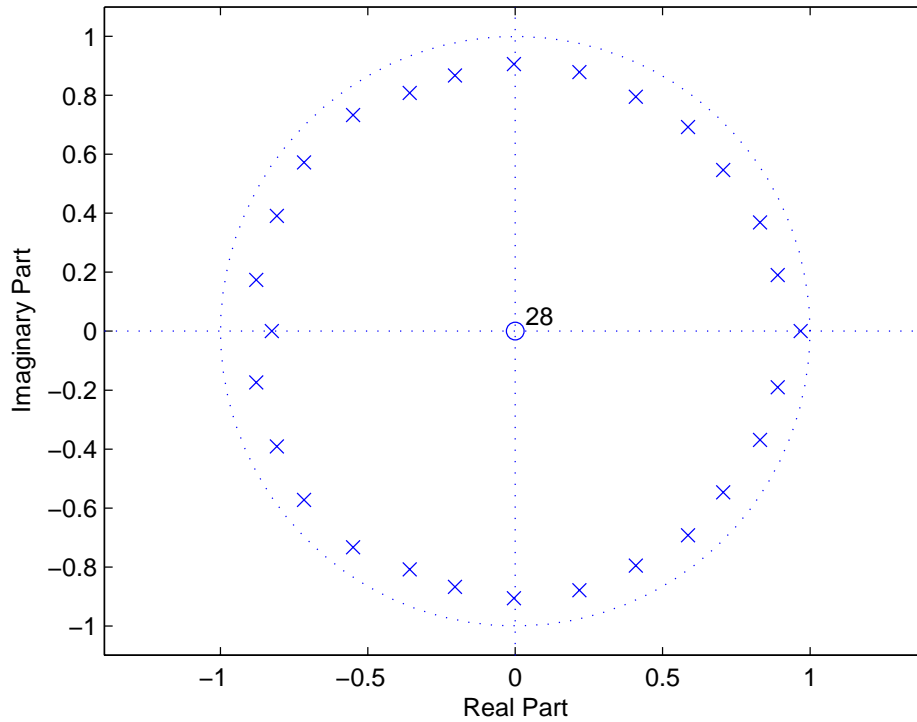


Figura 5.13: Diagrama de pólos e zeros do modelo AR(28).

A série é Gaussiana, de acordo com os testes de Hinich. O número estimado de raízes unitárias é igual a zero.

Tabela 5.8: Série ARFIMA(0;0,3;0): resultados dos testes estatísticos e número estimado de raízes unitárias.

x_t é $I(0)$? KPSS	x_t é LRD?		Hinich		\hat{m}
	R/S	GPH	x_t é normal?	x_t é linear?	
NAO	SIM	SIM	SIM	SIM	0
$p_{af} \leq 0,01$	$p_{af} \leq 0,01$	$p_{af} \leq 0,01$	$p_{af} = 0,4505$		

A Fig. 5.14 ilustra o periodograma (Fig. 5.14a), periodograma alisado com janela de Daniell (Fig. 5.14b) e o periodograma alisado pelo método WOSA (Fig. 5.14c). Os periodogramas suavizados mostram que o espectro da série possui uma singularidade do tipo $1/f^{2H-1}$ na origem. O comportamento LRD da série é confirmado pelo espectro *wavelet* aproximadamente linear da Fig. 5.15.

A Fig. 5.16 mostra o QQ-plot da série ARFIMA(0;0,3;0) (Fig. 5.16a) e um gráfico com a SACF e ACFs teóricas dos modelos ARFIMA(0;0,2895;0) e AR(16) estimados (Fig. 5.16b). O processo AR(16) é o melhor modelo AR de acordo com o critério AIC. O QQ-plot indica que a série é Gaussiana. O processo LRD ajustado consegue modelar o decaimento lento da SACF da série.

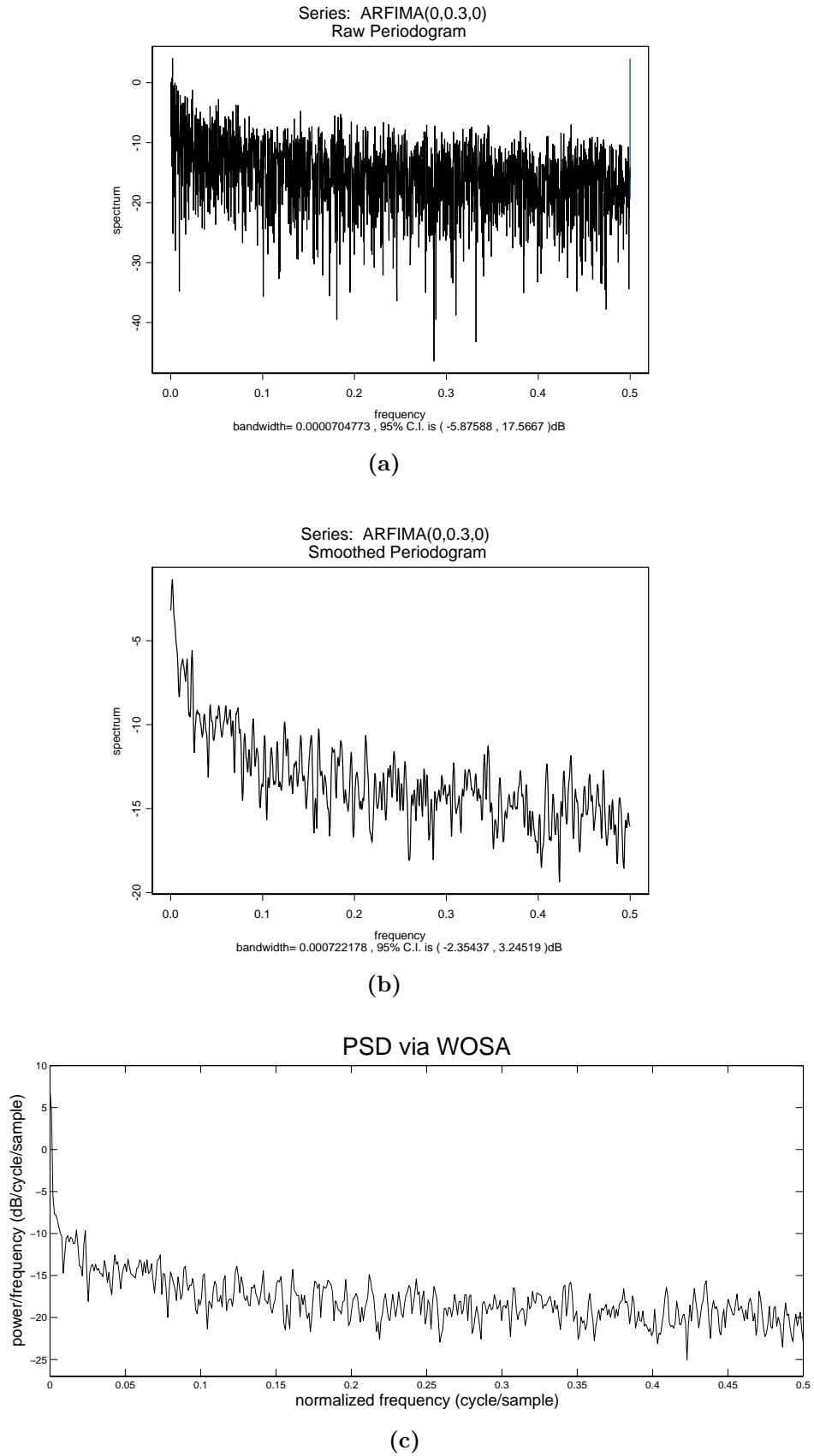


Figura 5.14: Série ARFIMA(0; 0, 3; 0): (a) periodograma, (b) periodograma alisado com janela de autocorrelação de Daniell e (c) periodograma estimado pelo método WOSA.

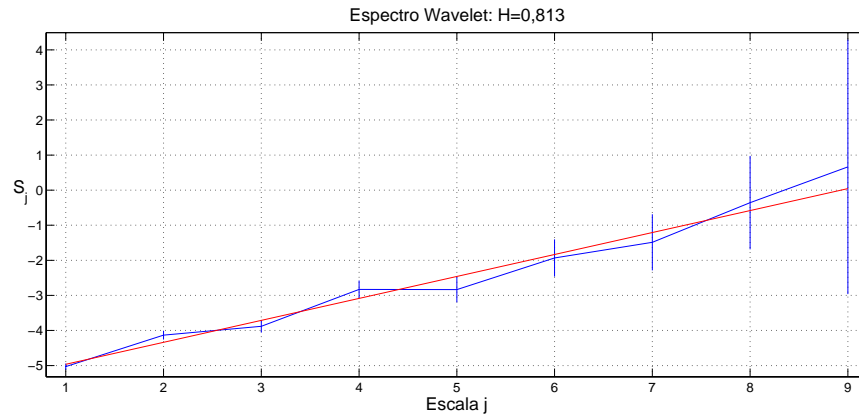


Figura 5.15: Espectro *wavelet* do sinal ARFIMA(0; 0, 3; 0).

A Fig. 5.17 ilustra os gráficos do QQ-plot (Fig. 5.17a) e SACF (Fig. 5.17b) dos resíduos do modelo de memória longa estimado. Essas Figs. sugerem que a série dos resíduos é um RBG. A Fig. 5.18 mostra o diagrama de pólos e zeros do modelo AR(16) ajustado. Observe que o pólo real positivo está situado nas proximidades do círculo unitário.

5.4.4 Série MWM com $H = 0,8$

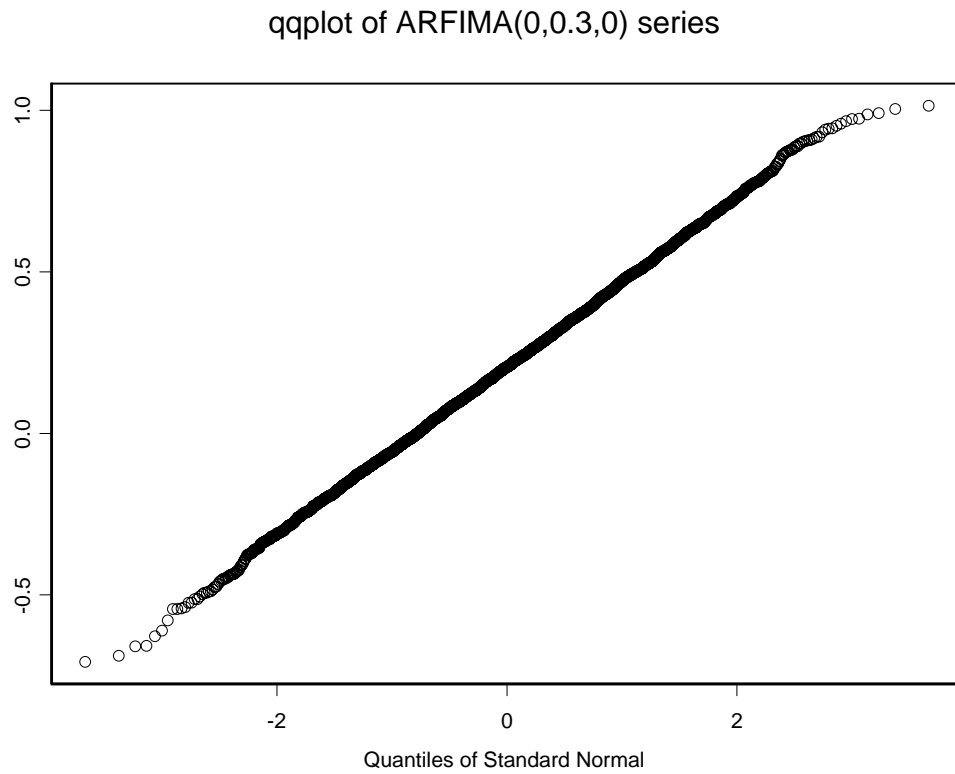
A função `S+FinMetrics FARIMA` ajustou um modelo ARFIMA(0; 0, 3164; 0) à série MWM($H = 0,8$). A Tabela 5.9 resume as estimativas do parâmetro d obtidas pelos métodos R/S, periodograma com janela de Daniell, Whittle, Haslett-Raftery e Abry-Veitch. A Tabela mostra que a estimativa pela análise R/S ($\hat{d} = 0,2197$) é discordante das demais, que estão situadas na faixa $0,27 \leq \hat{d} \leq 0,35$.

Tabela 5.9: Estimativas do parâmetro H para a realização MWM($H = 0,8$).

análise	periodograma	estimador	estimador de MV	estimador
R/S	com janela	de Whittle	de Haslett-Raftery	<i>wavelet</i>
	de Daniell			Abry/Veitch
0,2197	0,2742	0,3211	0,3164	0,341

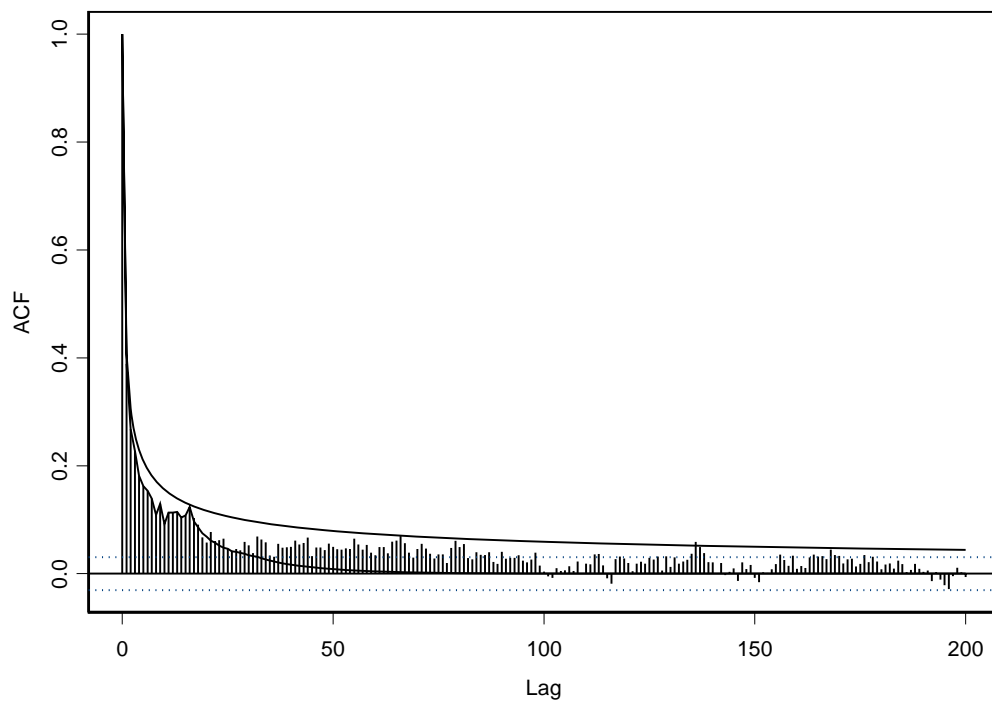
A Tabela 5.10 mostra os resultados dos testes estatísticos e o número estimado de raízes unitárias da série MWM($H = 0,8$). O teste KPSS não rejeitou a hipótese de estacionariedade da série. A série é LRD de acordo com os testes R/S e GPH. A série é não-linear, de acordo com os testes de Hinich. Não há raízes unitárias.

A Fig. 5.19 ilustra o periodograma (Fig. 5.19a), periodograma alisado com janela de Daniell (Fig. 5.19b) e o periodograma alisado pelo método WOSA (Fig. 5.19c). Os periodogramas suavizados mostram que o espectro da série



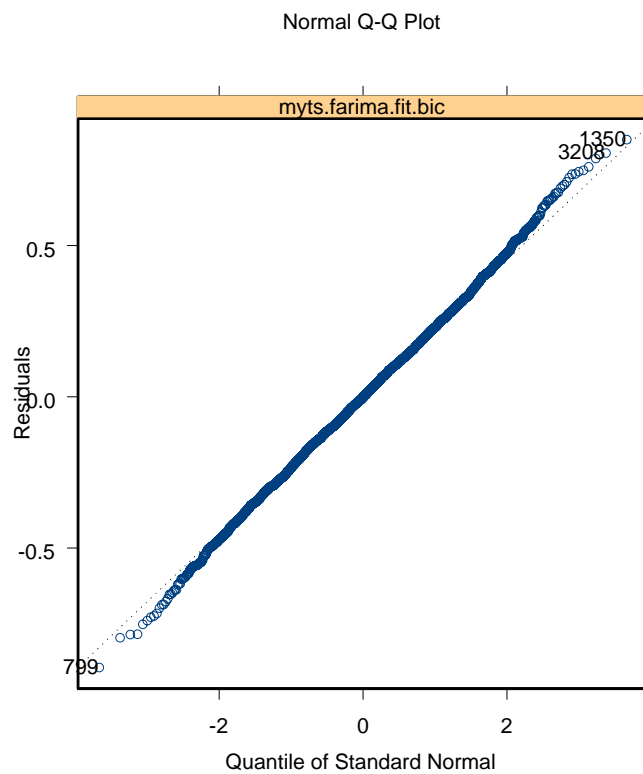
(a)

Series : ARFIMA(0,0.3,0)

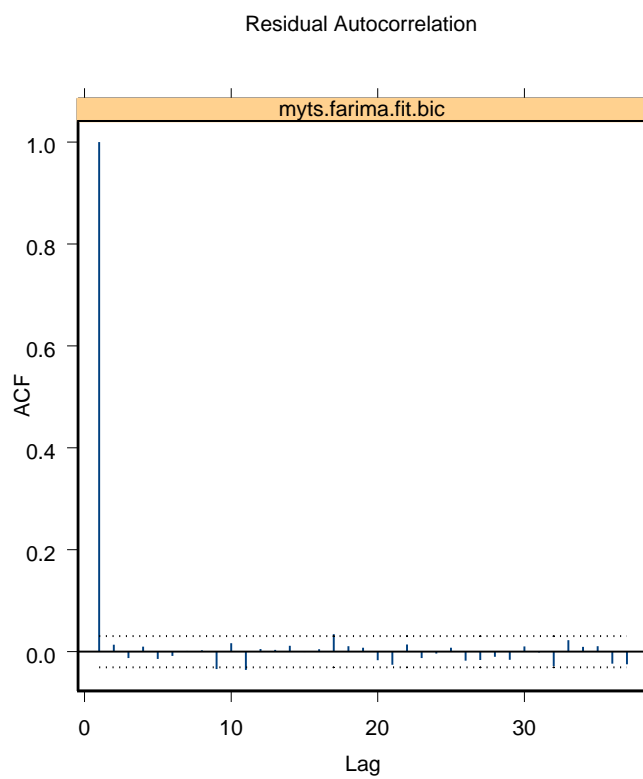


(b)

Figura 5.16: Série ARFIMA(0; 0, 3; 0): (a) QQ-plot e (b) gráfico com a SACF e ACFs teóricas dos modelos ARFIMA(0; 0, 2895; 0) e AR(16) estimados.



(a)



(b)

Figura 5.17: Resíduos do modelo ARFIMA(0; 0, 29; 0) estimado para a série ARFIMA(0; 0, 3; 0): (a) QQ-plot e (b) ACF.

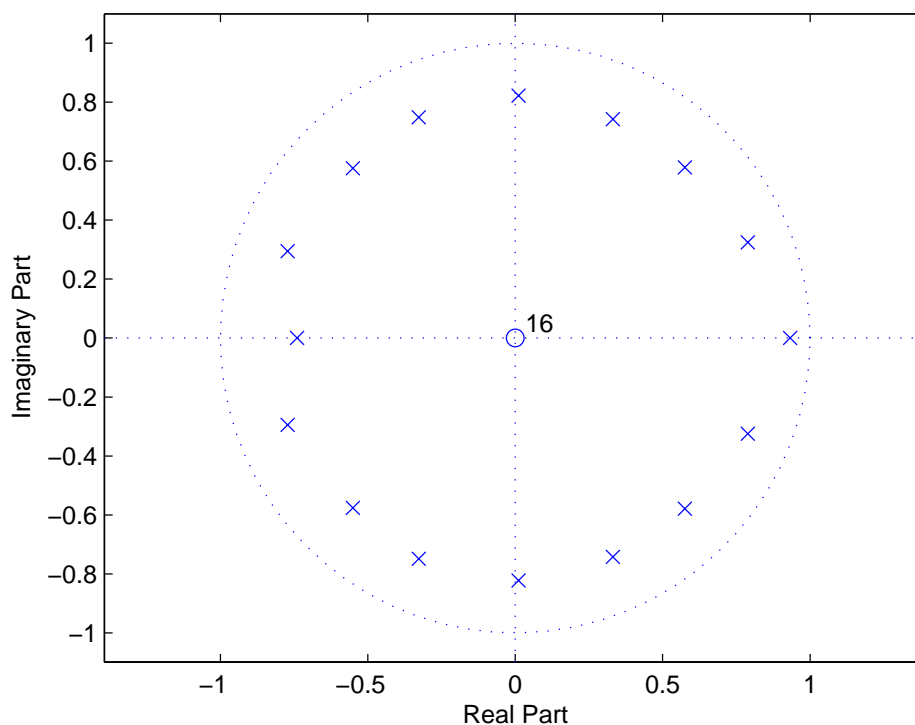


Figura 5.18: Diagrama de pólos e zeros do modelo AR(16).

Tabela 5.10: Série MWM($H = 0,8$): resultados dos testes estatísticos e número estimado de raízes unitárias.

x_t é $I(0)$?	x_t é LRD?		Hinich		\hat{m}
KPSS	R/S	GPH	x_t é normal?	x_t é linear?	
SIM	SIM	SIM	NÃO	NÃO	0
$p_{af} > 0,05$	$p_{af} \leq 0,01$	$p_{af} \leq 0,01$	$p_{af} = 0$		

possui uma singularidade do tipo $1/f^{2H-1}$ na origem. O comportamento LRD da série é confirmado pelo espectro *wavelet* aproximadamente linear da Fig. 5.20.

A Fig. 5.21 mostra o QQ-plot da série MWM($H = 0,8$) (Fig. 5.21a) e um gráfico com a SACF e ACFs teóricas dos modelos ARFIMA(0;0,3164;0) e AR(29) estimados (Fig. 5.21b). O processo AR(29) é o melhor modelo AR de acordo com o critério AIC. O QQ-plot indica que o sinal é não-Gaussiano. O processo LRD ajustado consegue modelar o decaimento lento da SACF da série.

A Fig. 5.22 ilustra os gráficos do QQ-plot (Fig. 5.22a) e SACF (Fig. 5.22b) dos resíduos do modelo LRD estimado. Essas Figs. indicam que a série dos resíduos é não-Gaussiana. A Fig. 5.23 mostra o diagrama de pólos e zeros do modelo AR(29) ajustado. Observe que há alguns pólos situados nas proximidades do círculo unitário.

5.4.5 Série do Rio Nilo

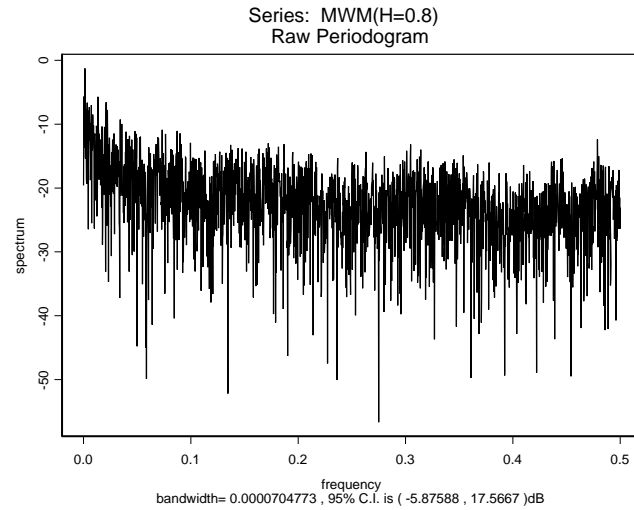
De acordo com Beran [58] e Percival e Walden [67, pág.386], a série do rio Nilo entre os anos de 622 e 1284 não é globalmente estacionária. Percival e Walden [67] demonstram, utilizando a DWT, que o comportamento da série muda por volta do ano 715 e que a mesma pode ser subdividida em duas séries ARFIMA localmente estacionárias. Percival e Walden [67] estimam $\hat{H} = 0,9532$ por meio do método aproximado de MV baseado na DWT.

A função **S+FinMetrics FARIMA** ajustou um modelo ARFIMA(0;0,505;0) (não-estacionário, pois $d > 0,5$). A Tabela 5.11 resume as estimativas do parâmetro d obtidas pelos métodos R/S, periodograma com janela de Daniell, Whittle, Haslett-Raftery e Abry-Veitch. A Tabela mostra que a estimativa pela análise R/S ($\hat{d} = 0,2853$) é discordante das demais, que estão situadas na faixa $0,40 \leq \hat{d} \leq 0,51$. Como o modelo ajustado pela função **FARIMA** é não-estacionário, esta seção adotará o modelo ARFIMA(0;0,4125;0), em que o valor do parâmetro d corresponde ao valor estimado pelo método de Whittle.

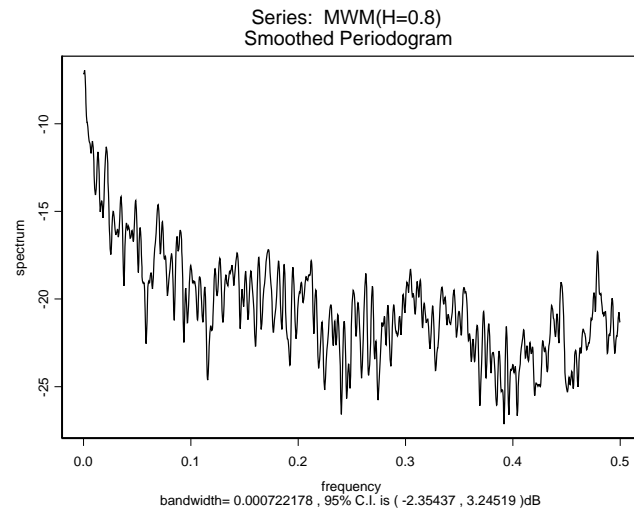
Tabela 5.11: Estimativas do parâmetro d para a série do rio Nilo.

análise	periodograma	estimador	estimador de MV	estimador
R/S	com janela	de Whittle	de Haslett-Raftery	<i>wavelet</i>
	de Daniell			Abry/Veitch
0,2853	0,4062	0,4125	0,5050	0,387

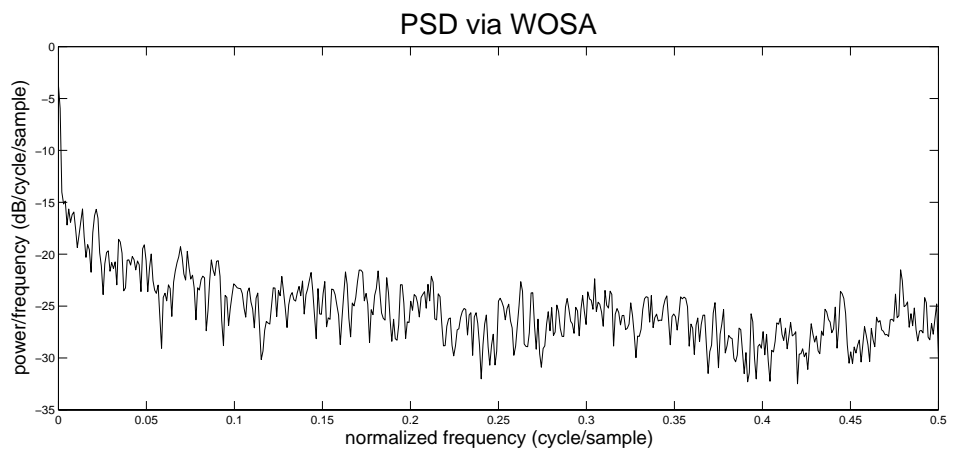
A Tabela 5.12 mostra os resultados dos testes estatísticos e o número estimado de raízes unitárias da série do rio Nilo. O teste KPSS rejeitou a hipótese



(a)



(b)



(c)

Figura 5.19: Simulação MWM($H = 0,8$): (a) periodograma, (b) periodograma alisado com janela de autocorrelação de Daniell e (c) periodograma estimado pelo método WOSA.

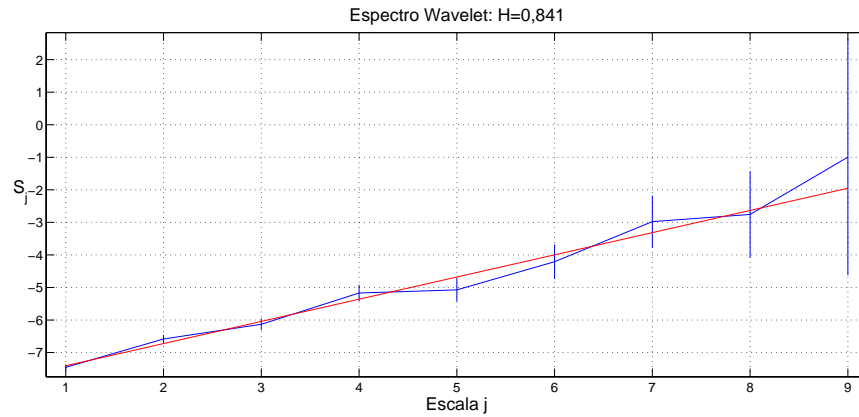


Figura 5.20: Espectro *wavelet* do sinal MWM($H = 0,8$).

de estacionariedade da série. A série é LRD de acordo com os testes R/S e GPH. A série é linear, porém não-Gaussiana, de acordo com os testes de Hinich. Os resultados dos testes de Hinich (série linear não-Gaussiana) sugerem que a distribuição marginal de probabilidades da série do rio Nilo possa ser modelada por uma distribuição estável (vide o teorema generalizado do limite central (3.2) apresentado no Cap. 3). Não há raízes unitárias.

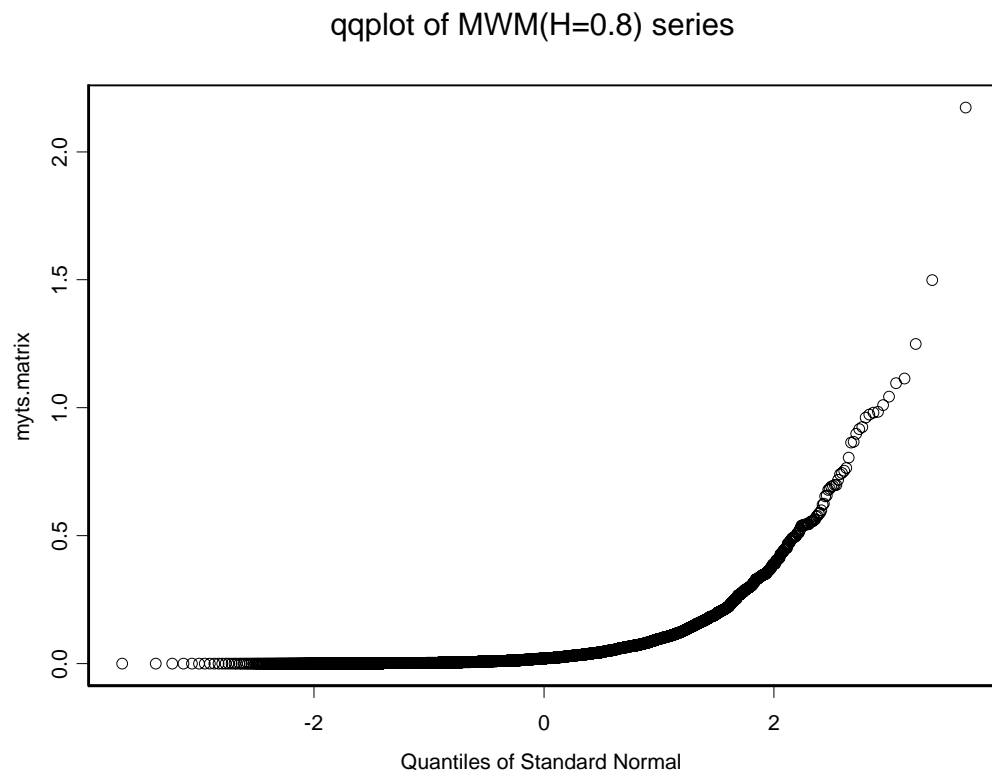
Tabela 5.12: Série do rio Nilo: resultados dos testes estatísticos e número estimado de raízes unitárias.

x_t é $I(0)$?	x_t é LRD?		Hinich		\hat{n}
KPSS	R/S	GPH	x_t é normal?	x_t é linear?	
NAO	SIM	SIM	NAO	SIM	0
$p_{af} < 0,01$	$p_{af} < 0,01$	$p_{af} < 0,05$	$p_{af} = 0$		

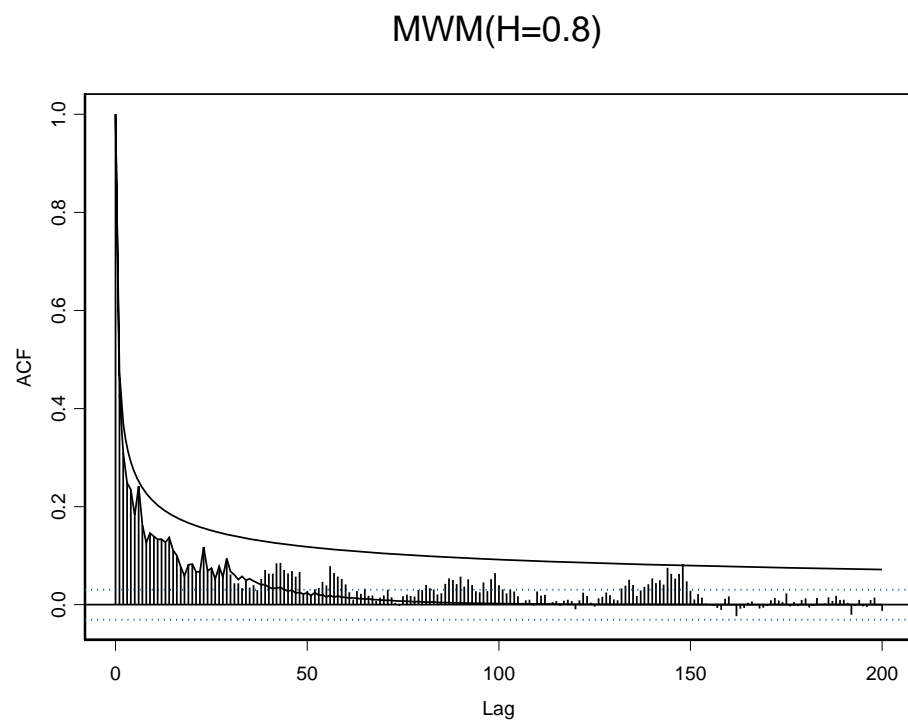
A Fig. 5.24 ilustra o periodograma (Fig. 5.24a), periodograma alisado com janela de Daniell (Fig. 5.24b) e o periodograma alisado pelo método WOSA (Fig. 5.24c). Os periodogramas alisados sugerem que o espectro da série possui uma singularidade do tipo $1/f^{2H-1}$ na origem. O comportamento LRD da série é confirmado pelo espectro *wavelet* aproximadamente linear da Fig. 5.25.

A Fig. 5.26 mostra o QQ-plot da série do rio Nilo entre os anos de 1007 e 1206 (Fig. 5.26a) e um gráfico com a SACF e ACFs teóricas dos modelos ARFIMA(0; 0,4125; 0) e AR(8) estimados (Fig. 5.26b). O processo AR(8) é o melhor modelo AR de acordo com o critério AIC.

A Fig. 5.27 ilustra os gráficos do QQ-plot (Fig. 5.27a) e SACF (Fig. 5.27b) dos resíduos do modelo ARFIMA estimado. Essas Figs. indicam que a série dos resíduos é um RB não-Gaussiano. A Fig. 5.28 mostra o diagrama de pólos e zeros do modelo AR(8) ajustado. Note que o pólo real positivo está posicionado

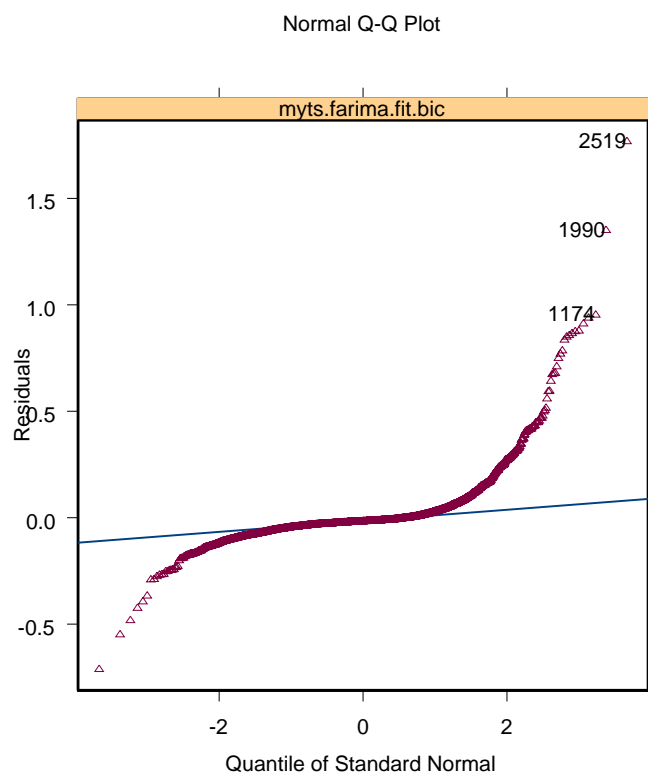


(a)

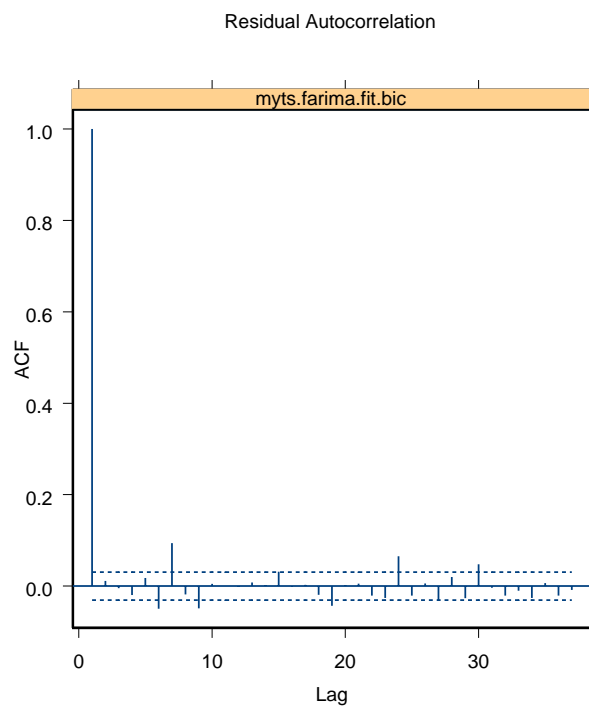


(b)

Figura 5.21: Simulação MWM($H = 0,8$): (a) QQ-plot e (b) Gráfico com a SACF e ACFs teóricas dos modelos ARFIMA(0;0,3164;0) e AR(29) estimados.



(a)



(b)

Figura 5.22: Resíduos do modelo ARFIMA(0;0,3164;0) estimado para a série MWM($H = 0,8$): (a) QQ-plot e (b) SACF.

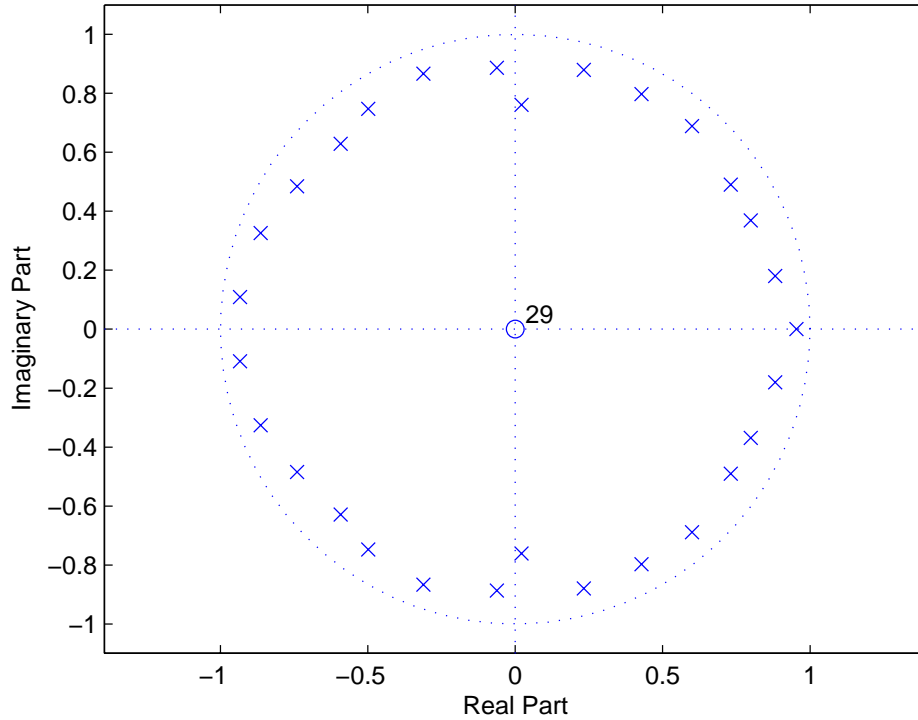


Figura 5.23: Diagrama de pólos e zeros do modelo AR(29).

nas proximidades do círculo unitário.

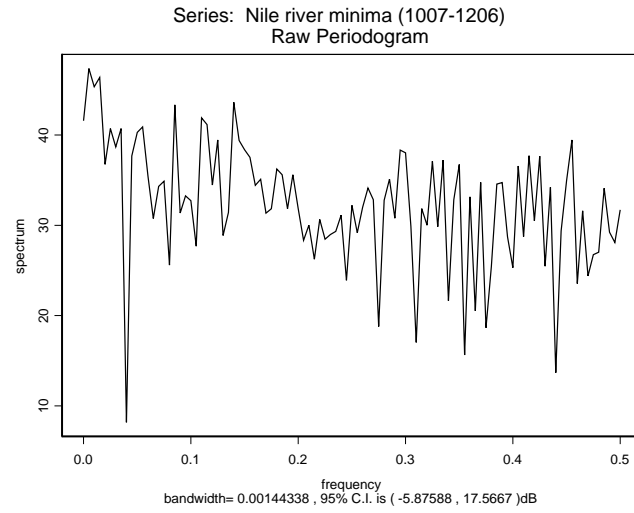
5.5 Estudo Empírico da Previsão com o Modelo TARFIMA

Os modelos TARFIMA(L) desta seção correspondem a versões aproximadas dos processos ARFIMA(p, d, q) estimados no item 5.4. Portanto, foram utilizadas as estimativas do parâmetro d de Haslett-Raftery.

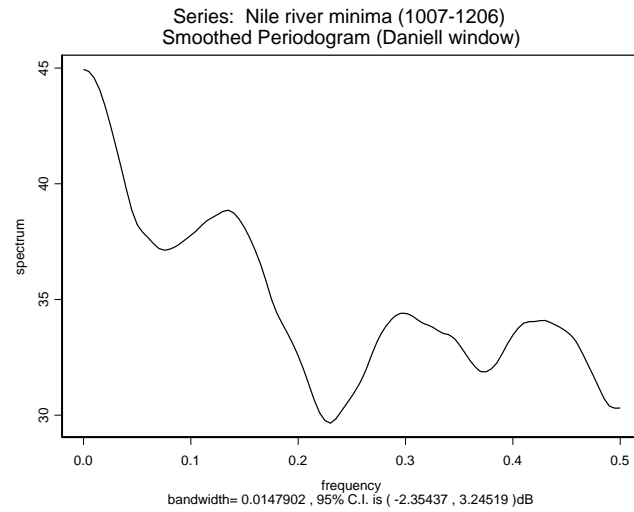
5.5.1 Séries ARFIMA(0, d , 0)

A Fig. 5.29 ilustra as previsões (Tipo-I e Tipo-II) de valores futuros da realização ARFIMA(0; 0, 4; 0) com origem em $t = 3996$, horizontes $h = 1, 2, \dots, 100$ e intervalos de 95% de confiança que foram obtidas por meio de modelos TARFIMA(100) e AR(15) (Fig. 5.29a) e TARFIMA(100) e AR(2) (Fig. 5.29b). O modelo TARFIMA(100) utilizou a estimativa (conservadora) $\hat{d} = 0,3684$ (vide Tabela 5.3).

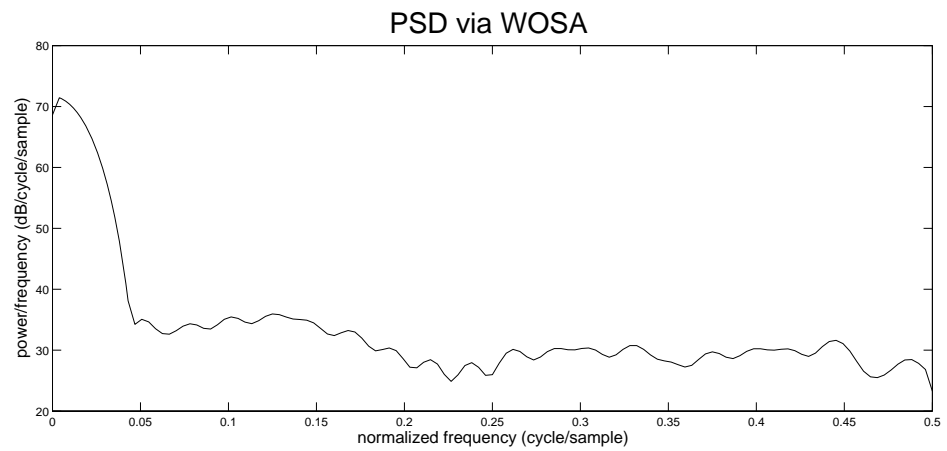
A Fig. 5.30a mostra a diferença entre os erros absolutos de previsão dos modelos AR(15) e TARFIMA(100) (gráfico $(\Delta_h(1) - \Delta_h(2))$ vs. t , em que $\Delta_h(1)$ é



(a)



(b)



(c)

Figura 5.24: Série dos níveis anuais mínimos do rio Nilo entre os anos de 1007 e 1206: (a) periodograma, (b) periodograma estimado com janela de Daniell e (c) periodograma estimado pelo método WOSA.

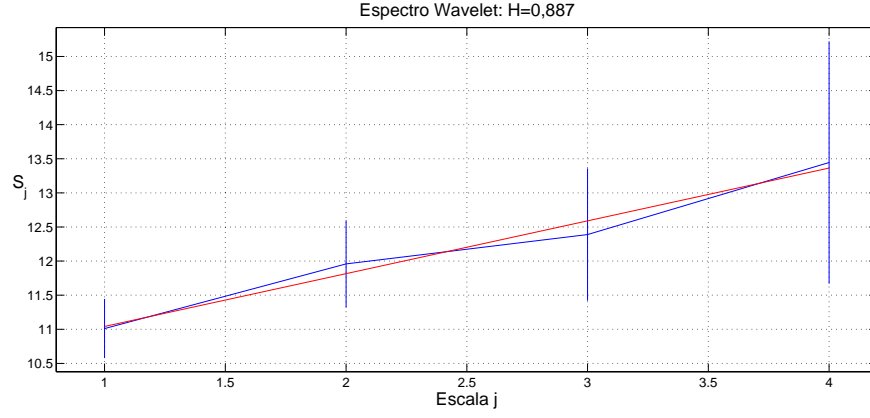


Figura 5.25: Espectro *wavelet* da série do rio Nilo entre os anos de 1007 e 1206.

o erro absoluto de previsão do modelo AR(15) e $\Delta_h(2)$ é o erro absoluto de previsão do modelo TARFIMA(100)) para a realização ARFIMA(0; 0, 4; 0). A Fig. 5.30b ilustra a diferença entre os erros absolutos de previsão dos modelos AR(2) e TARFIMA(100) (gráfico $(\Delta_h(1) - \Delta_h(2))$ vs. t , em que $\Delta_h(1)$ é o erro absoluto de previsão do modelo AR(2) e $\Delta_h(2)$ é o erro absoluto de previsão do modelo TARFIMA(100)) para a mesma realização. A Fig. 5.30 sugere que o desempenho do modelo TARFIMA(100) é superior, pois os erros absolutos de previsão dos modelos AR tendem a ser maiores do que o do processo TARFIMA(100) (note que o número de ocorrências de valores positivos de $(\Delta_h(1) - \Delta_h(2))$ nos gráficos das Figs. 5.30a e 5.30b é maior do que o número de ocorrências de valores negativos).

Também é interessante comparar o desempenho dos modelos em termos do critério da **variação proporcional do erro quadrático médio empírico acumulado de previsão** (*Proportionate Change in the Cumulative Empirical Mean Squared Error* (PCEMSE)) [147], que será definido a seguir. Primeiramente, o **erro quadrático médio empírico acumulado de previsão** (*Cumulative Empirical Mean Squared Error* (CEMSE)) com origem em t e horizonte máximo K ($h = 1, 2, \dots, h_{\max} = K$) de um modelo \mathbf{x}_t é definido como [42]

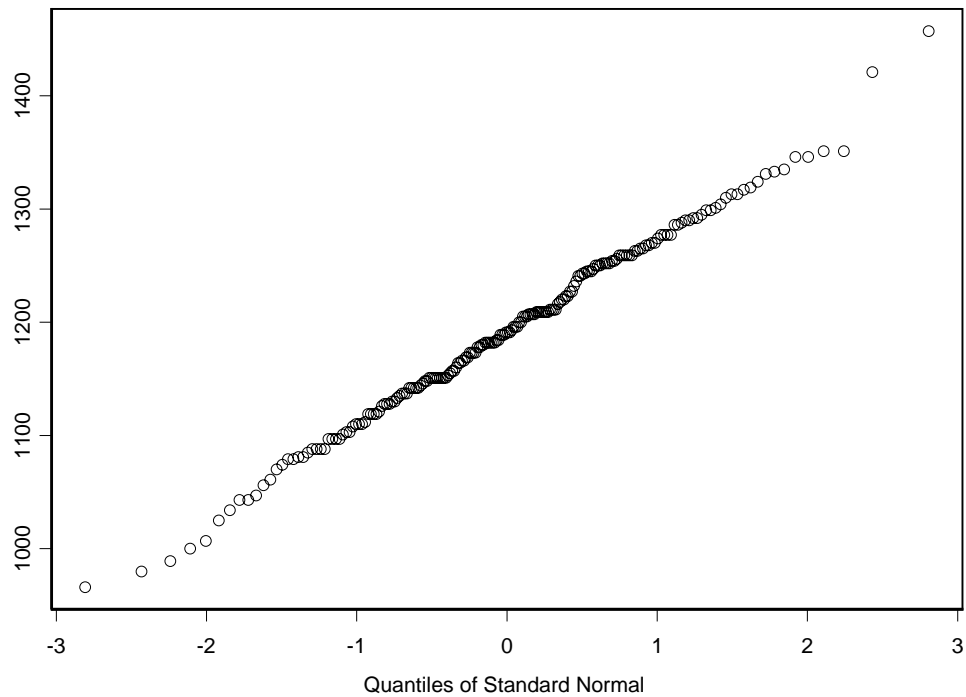
$$CEMSE_{\mathbf{x}}(t, K) = \frac{1}{K} \sum_{h=1}^K (x_{t+h} - \hat{x}_{t+h})^2. \quad (5.23)$$

O PCEMSE do modelo \mathbf{x}_t com relação ao modelo \mathbf{y}_t com origem em t e horizonte K é dado por

$$PCEMSE_{\mathbf{x}, \mathbf{y}}(t, K) = \frac{CEMSE_{\mathbf{x}}(t, K) - CEMSE_{\mathbf{y}}(t, K)}{CEMSE_{\mathbf{y}}(t, K)}. \quad (5.24)$$

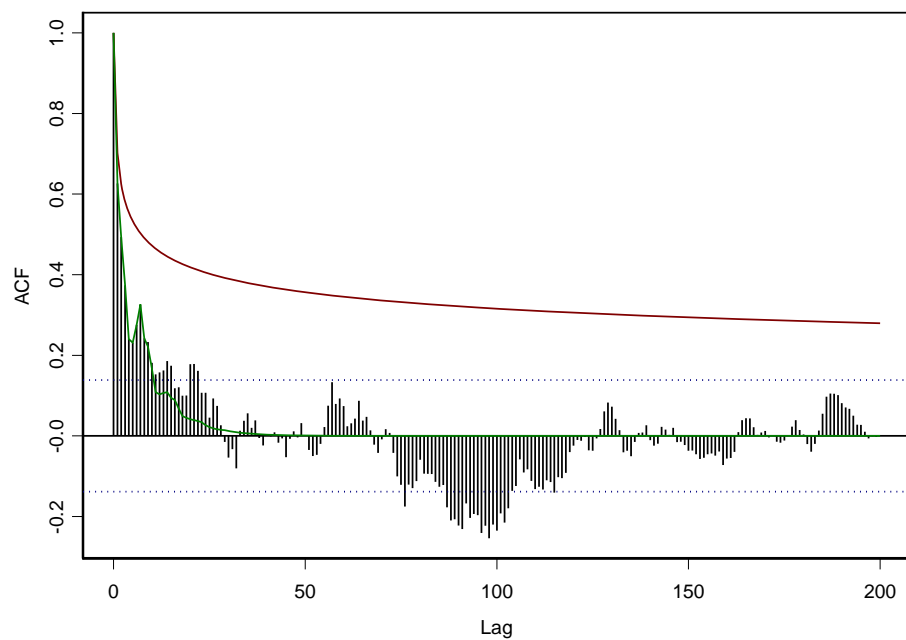
Diz-se que \mathbf{x}_t é um modelo de previsão mais preciso (ou melhor) do que \mathbf{y}_t se

qqplot of Nile river minima (1007-1206)



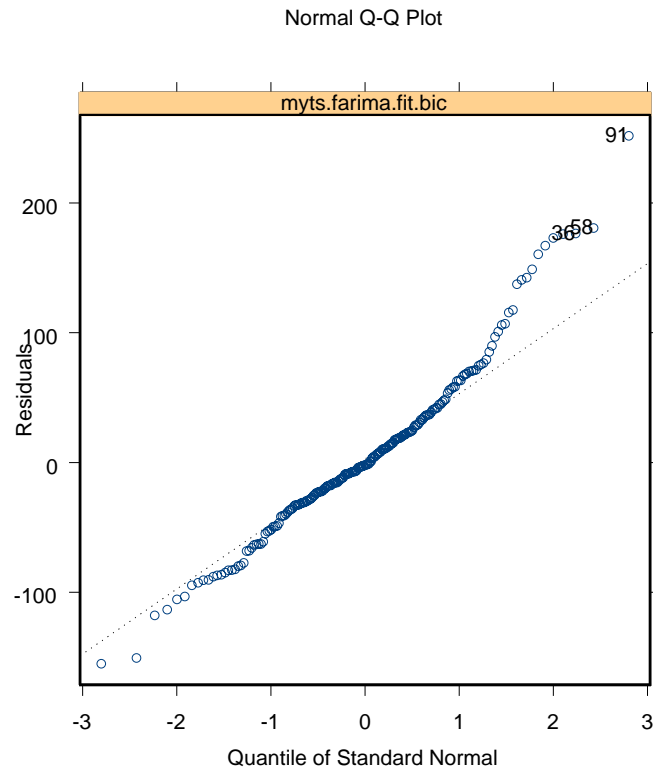
(a)

Rio Nilo

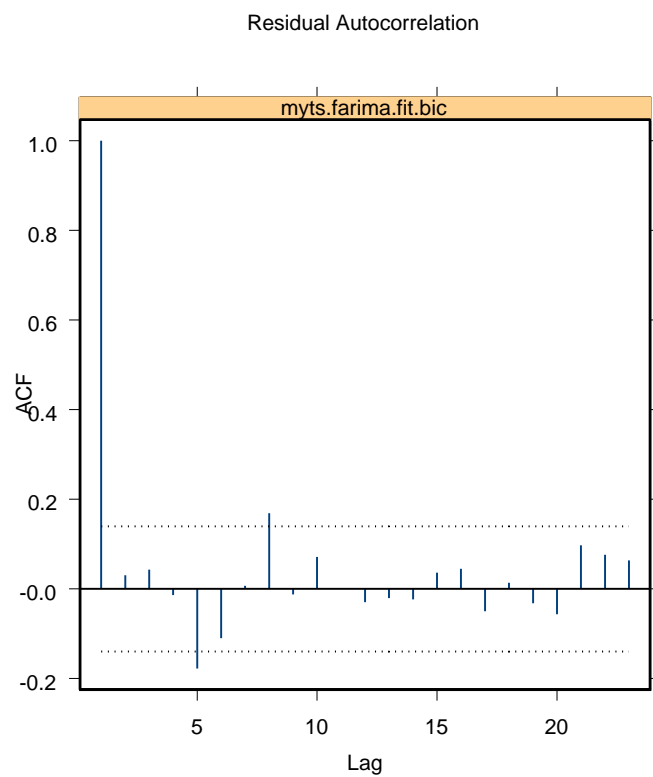


(b)

Figura 5.26: Série dos níveis anuais mínimos do rio Nilo entre os anos de 1007 e 1206: (a) QQ-plot e (b) gráfico com a SACF e ACFs teóricas dos modelos ARFIMA(0; 0, 4125; 0) (linha de cor vermelha) e AR(8) (linha de cor verde) estimadas.



(a)



(b)

Figura 5.27: Resíduos do modelo ARFIMA(0;0,4125;0) ajustado para a série dos níveis anuais mínimos do rio Nilo entre os anos de 1007 e 1206: (a) QQ-plot e (b) SACF.

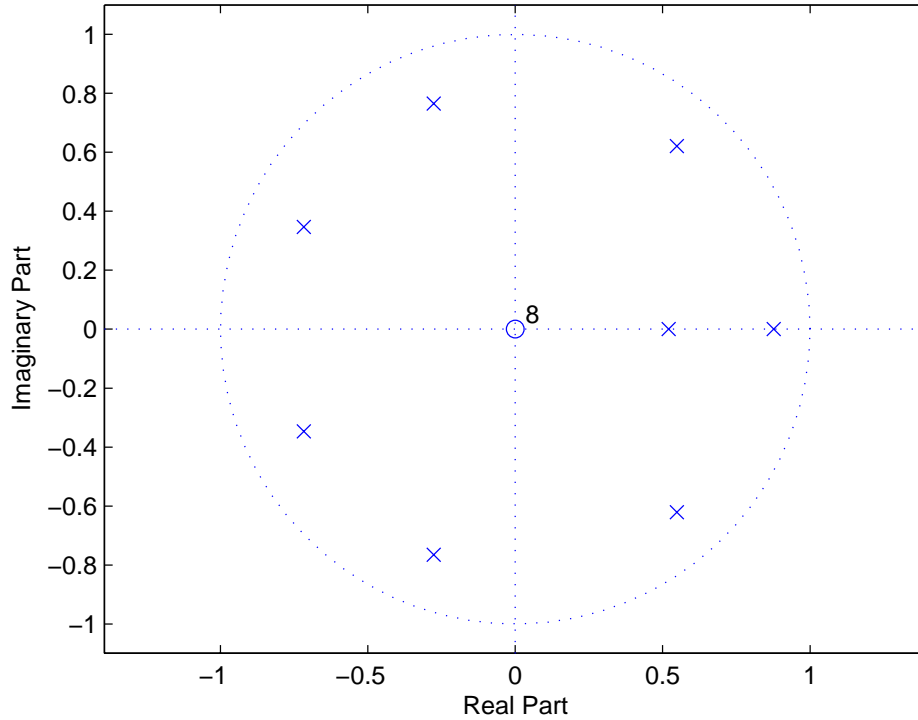


Figura 5.28: Diagrama de pólos e zeros do modelo AR(8).

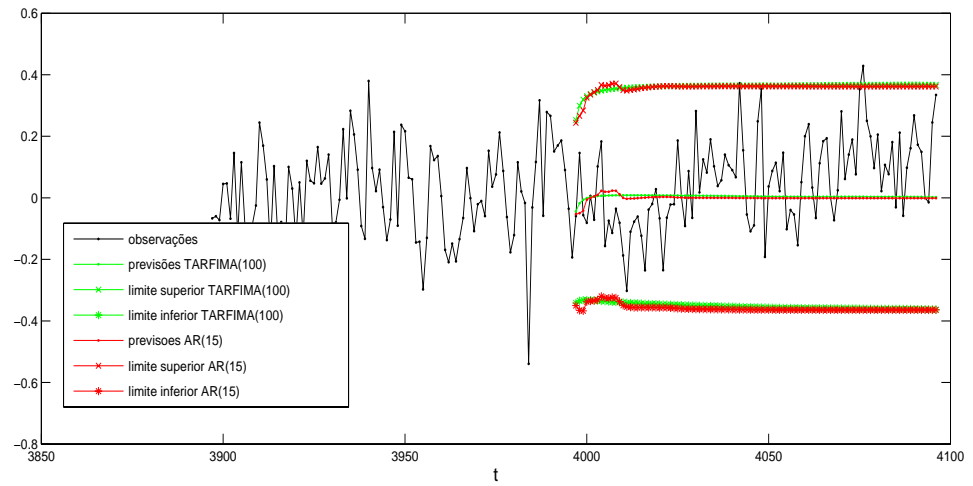
$$PCEMSE_{x,y}(t, K) < 0.$$

A Tabela 5.13 mostra que as previsões dos modelos AR(15), TARFIMA(10), TARFIMA(50) e TARFIMA(100) para a série ARFIMA(0; 0, 4; 0) são melhores do que as previsões do modelo AR(2) e que os modelos AR(15), TARFIMA(50) e TARFIMA(100) têm desempenhos similares, de acordo com o critério comparativo PCFEMSE.

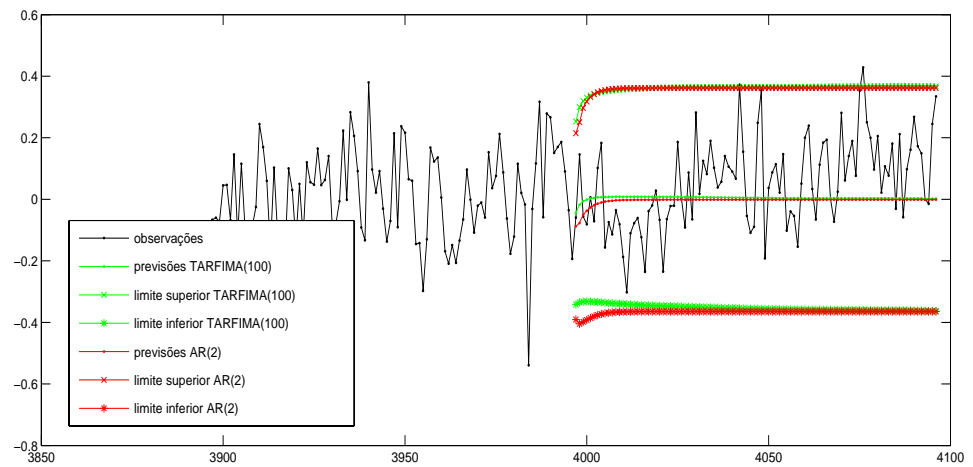
Tabela 5.13: Série ARFIMA(0; 0, 4; 0): $PCFEMSE_{x,y}(t = 3996, K)$ dos modelos estimados AR(15), TARFIMA(10), TARFIMA(50) e TARFIMA(100) (colunas em que $L = 10, 50$, e 100 , respectivamente) relativos ao modelo AR(2). Os modelos TARFIMA(L) utilizaram $\hat{d} = 0,3684$.

K	AR(15)	L=10	L=50	L=100
7	-0,2807	-0,0956	-0,2839	-0,2882
10	-0,3451	-0,1529	-0,3401	-0,3438
20	-0,4203	-0,2187	-0,4068	-0,4092
50	-0,0517	-0,0262	-0,0917	-0,0840
100	0,0000	-0,0095	-0,0145	-0,0239

A Fig. 5.31 ilustra as previsões de valores futuros da série ARFIMA(0; 0, 3; 0) com origem em $t = 3996$, horizontes $h = 1, 2, \dots, 100$ e intervalos de 95% de confiança que foram obtidas por meio dos modelos TARFIMA(100) e AR(16) (Fig. 5.31a) e TARFIMA(100) e AR(2) (Fig. 5.31b). O modelo TARFIMA(100)

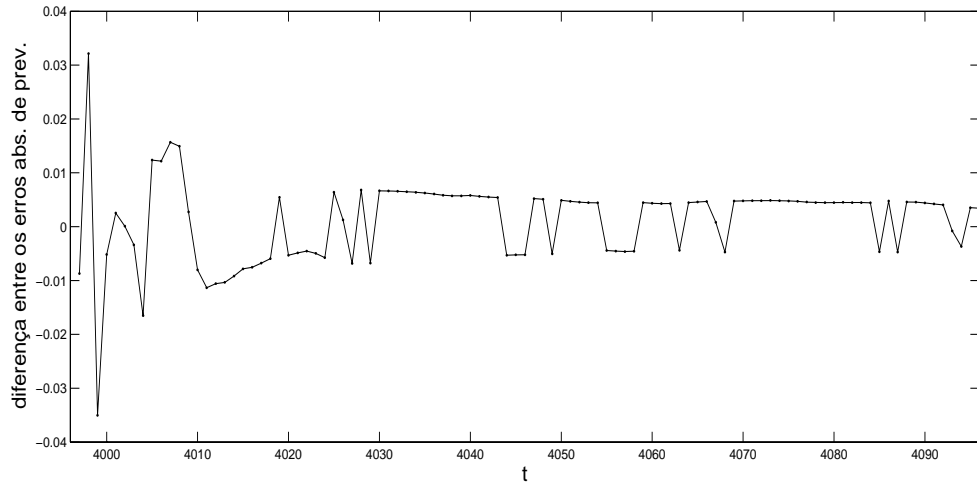


(a)

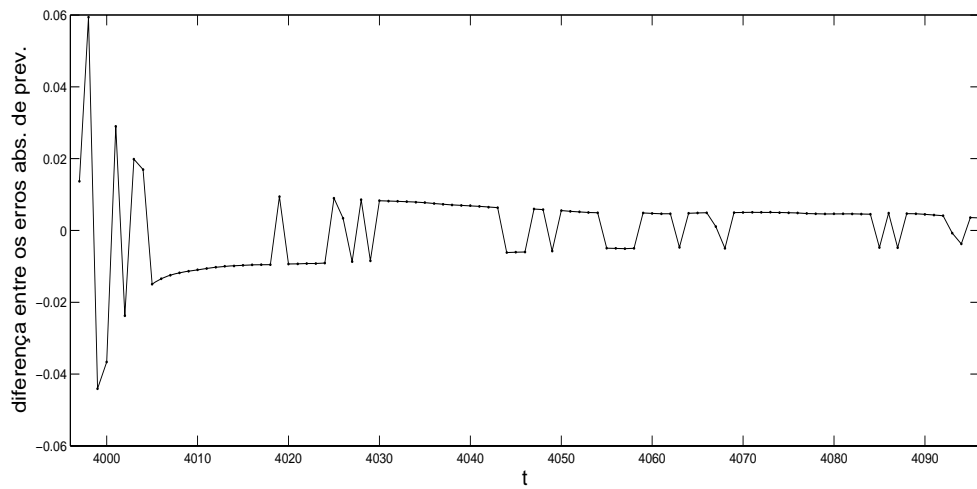


(b)

Figura 5.29: Série ARFIMA(0; 0, 4; 0): (a) observações, previsões de passo- h e intervalos de 95% de confiança para as previsões dos modelos TARFIMA(100) e AR(15) e (b) observações, previsões de passo- h e intervalos de 95% de confiança para as previsões dos modelos TARFIMA(100) e AR(2) .



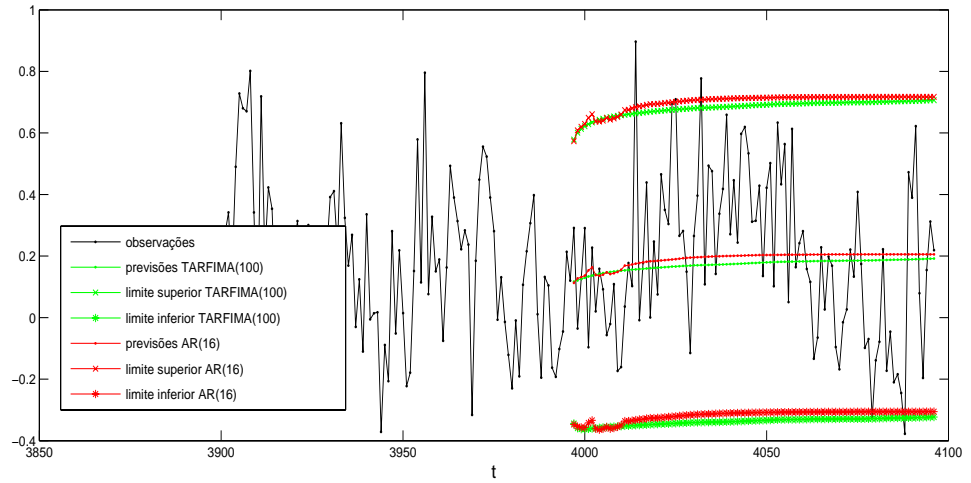
(a)



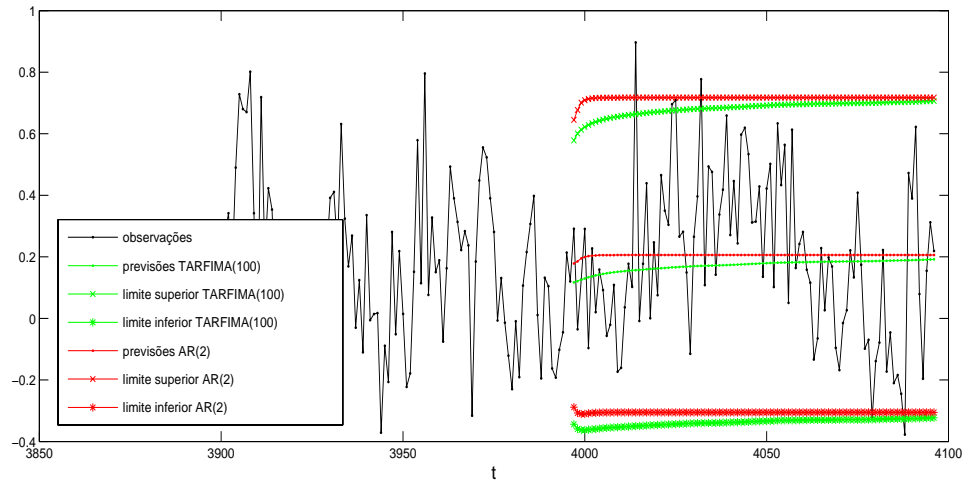
(b)

Figura 5.30: Série ARFIMA(0; 0, 4; 0): (a) diferença entre os erros absolutos de previsão dos modelos AR(15) e TARFIMA(100) (gráfico $(\Delta_h(1) - \Delta_h(2))$ vs. t , em que $\Delta_h(1)$ é o erro absoluto de previsão do modelo AR(15) e $\Delta_h(2)$ é o erro absoluto de previsão do modelo TARFIMA(100)) e (b) diferença entre os erros absolutos de previsão dos modelos AR(2) e TARFIMA(100) (gráfico $(\Delta_h(1) - \Delta_h(2))$ vs. t , em que $\Delta_h(1)$ é o erro absoluto de previsão do modelo AR(2) e $\Delta_h(2)$ é o erro absoluto de previsão do modelo TARFIMA(100)).

utilizou a estimativa $\hat{d} = 0,2895$ (vide Tabela 5.7).



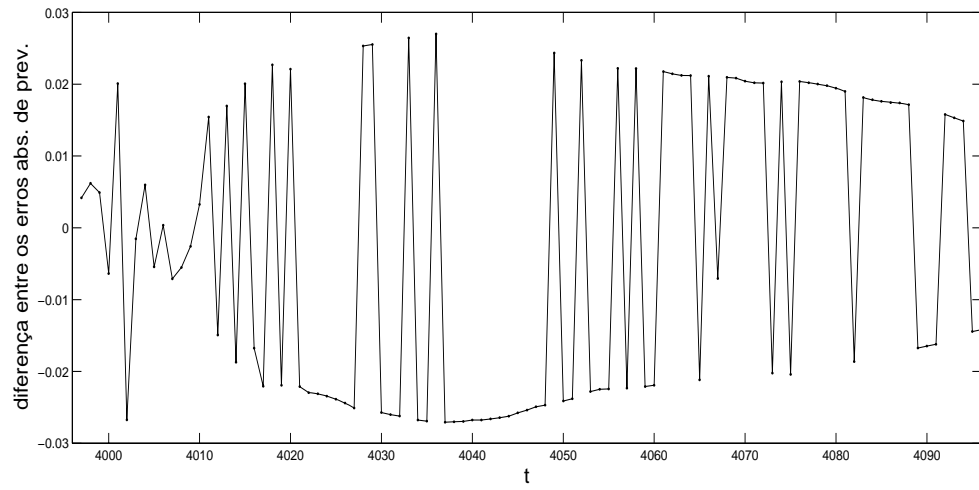
(a)



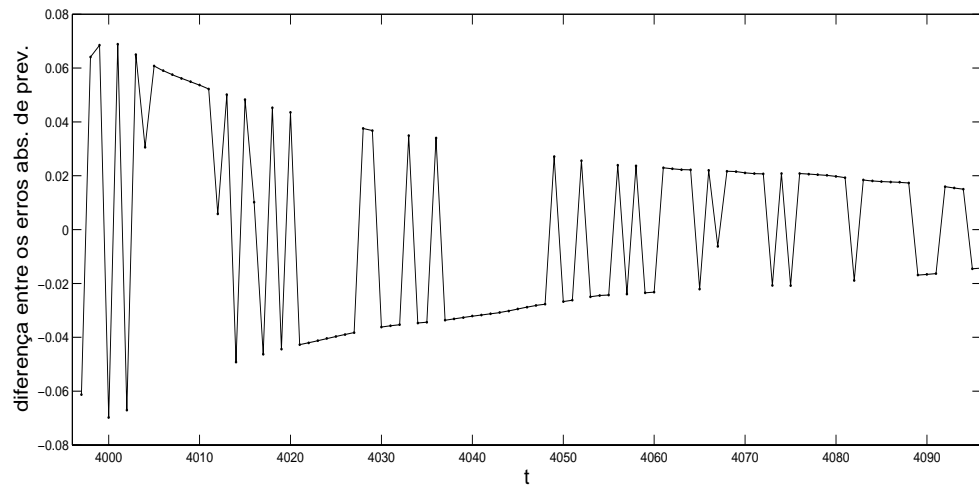
(b)

Figura 5.31: Série ARFIMA(0; 0, 3; 0): (a) observações, previsões de passo- h e intervalos de 95% de confiança para as previsões dos modelos TARFIMA(100) e AR(16) e (b) observações, previsões de passo- h e intervalos de 95% de confiança para as previsões dos modelos TARFIMA(100) e AR(2).

A Fig. 5.32a mostra as diferenças entre os erros absolutos das previsões dos modelos AR(16) e TARFIMA(100) para a série ARFIMA(0; 0, 3; 0), enquanto que a Fig. 5.32b mostra as diferenças entre os erros absolutos das previsões dos modelos AR(2) e TARFIMA(100) para a mesma série. O gráfico da Fig. 5.32a ilustra que o desempenho do modelo TARFIMA(100) é “quase sempre” pior do que o do modelo AR(16) até o instante $t = 4060$ e que depois esta tendência se inverte. O gráfico da Fig. 5.32b indica que o TARFIMA(100) “tende” a ser melhor do que o AR(2) até $t = 4020$, que o AR(2) é superior para $4021 \leq t \leq 4060$ e que depois o TARFIMA(100) é melhor.



(a)



(b)

Figura 5.32: Série ARFIMA(0;0,3;0): (a) diferença entre os erros absolutos de previsão dos modelos AR(16) e TARFIMA(100) e (b) diferença entre os erros absolutos de previsão dos modelos AR(2) e TARFIMA(100).

A Tabela 5.14 mostra que as previsões dos modelos AR(16), TARFIMA(50) e TARFIMA(100) para a série ARFIMA(0; 0, 3; 0) para $K = \{7, 10, 50\}$ são piores do que as previsões do modelo AR(2) (sendo que os modelos TARFIMA(50) e TARFIMA(100) apresentam desempenhos inferiores ao do AR(16)), conforme o critério comparativo PCEMSE. Qual é o significado desse resultado? Será que ele é uma evidência de que “não vale a pena” estimar modelos de memória longa (pelo menos quando o objetivo é a previsão de valores futuros) mesmo que a série em questão seja LRD? A resposta é **não**. Modelos SRD podem apresentar, **eventualmente**, em determinados trechos da série sob análise, um desempenho de previsão superior ao de modelos LRD. Contudo, deve-se ter em mente que modelos LRD têm um desempenho superior “na média” (porque levam em conta todas as possíveis realizações do processo estocástico) quando aplicados às séries LRD, conforme justificado pelo item 5.3. Observe-se que o desempenho superior do AR(2) é justificado pela ocorrência de uma inversão “local” de tendência na série ARFIMA(0; 0, 3; 0) (vide Fig. 5.31b) por volta de $t = 4010$ (veja que os valores da série tendem a diminuir para $3960 \leq t \leq 4010$ e a crescer para $4010 < t \leq 4060$), que não é bem modelada pelos modelos TARFIMA e pelo AR de ordem alta.

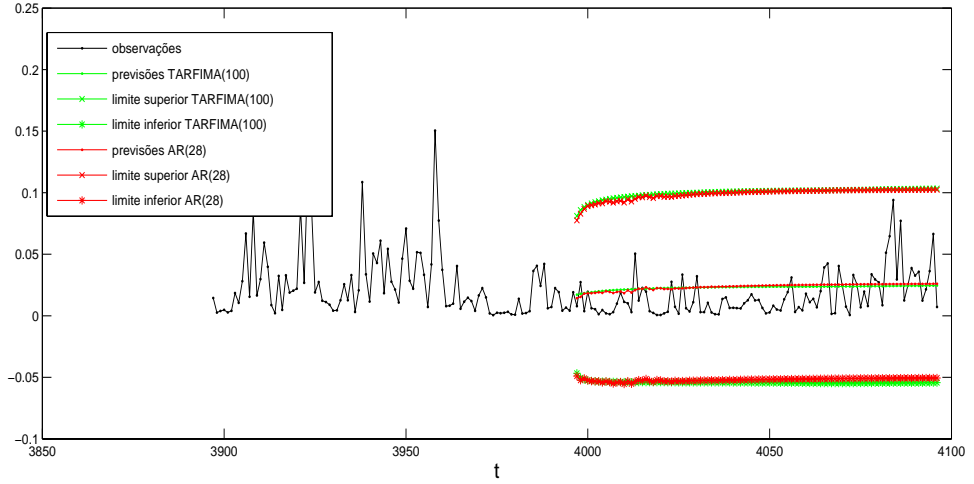
Tabela 5.14: Série ARFIMA(0; 0, 3; 0): $PCEMSE_{x,y}(t = 3996, K)$ dos modelos estimados AR(16), TARFIMA(10), TARFIMA(50) e TARFIMA(100) relativos ao modelo AR(2). Os modelos TARFIMA(L) utilizaram $\hat{d} = 0,2895$.

K	AR(16)	L=10	L=50	L=100
7	0,2186	0,1793	0,2464	0,2355
10	0,1118	0,0878	0,1309	0,1242
20	-0,1296	-0,0359	-0,0249	-0,0324
50	0,0144	-0,0311	0,1075	0,0733
100	-0,0104	-0,0107	0,0173	0,0167

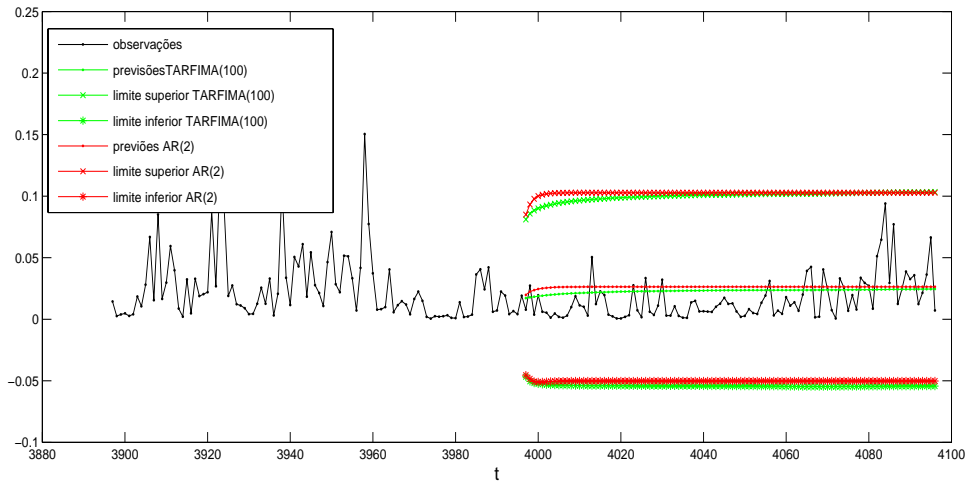
5.5.2 Séries MWM

A Fig. 5.33 mostra as previsões de valores futuros da série MWM($H = 0,9$) com origem em $t = 3996$, horizontes $h = 1, 2, \dots, 100$ e intervalos de 95% de confiança que foram obtidas por meio de modelos TARFIMA(100) e AR(28) (Fig. 5.33a) e TARFIMA(100) e AR(2) (Fig. 5.33b). As previsões do modelo TARFIMA(100) na Fig. 5.33 utilizaram a estimativa $\hat{d} = 0,3572$ (vide Tabela 5.5). A Fig. 5.34 ilustra as previsões de valores futuros do sinal MWM($H = 0,8$) com origem em $t = 3996$, horizontes $h = 1, 2, \dots, 100$ e intervalos de 95% de confiança que foram obtidas por meio de modelos TARFIMA(100) e AR(29) (Fig. 5.34a) e

TARFIMA(100) e AR(2) (Fig. 5.34b). As previsões do modelo TARFIMA(100) na Fig. 5.34 utilizaram a estimativa $\hat{d} = 0,3164$ (vide Tabela 5.9).



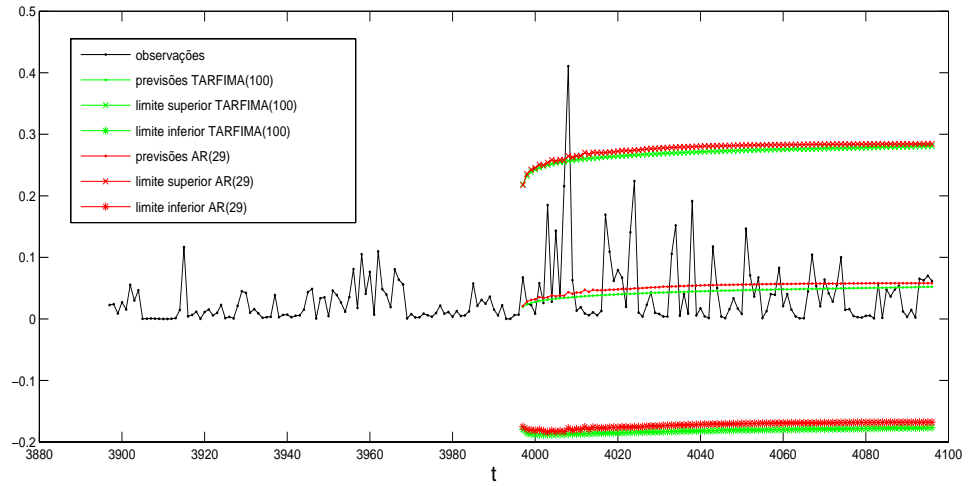
(a)



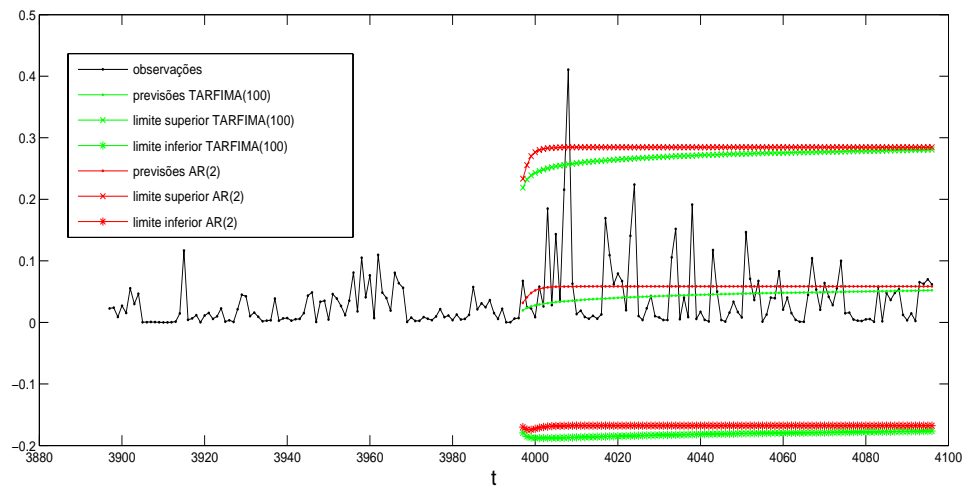
(b)

Figura 5.33: Série MWM($H = 0,9$): (a) observações, previsões de passo- h e intervalos de 95% de confiança para as previsões dos modelos TARFIMA(100) e AR(28) e (b) observações, previsões de passo- h e intervalos de 95% de confiança para as previsões dos modelos TARFIMA(100) e AR(2) .

A Fig. 5.35 ilustra a diferença entre os erros absolutos de previsão de modelos AR (de ordens alta e baixa) e do modelo TARFIMA(100) para a realização MWM($H = 0,9$). O gráfico da Fig. 5.35b mostra que as previsões do modelo TARFIMA(100) para a série MWM($H = 0,9$) tendem a ser melhores do que as previsões realizadas com o modelo AR(2). Os gráficos da Fig. 5.36 indicam que as previsões do modelo TARFIMA(100) para a série MWM($H = 0,8$) tendem a ser melhores do que as previsões realizadas com modelos AR de ordens baixa e alta (vide Figs. 5.36a e 5.36b).

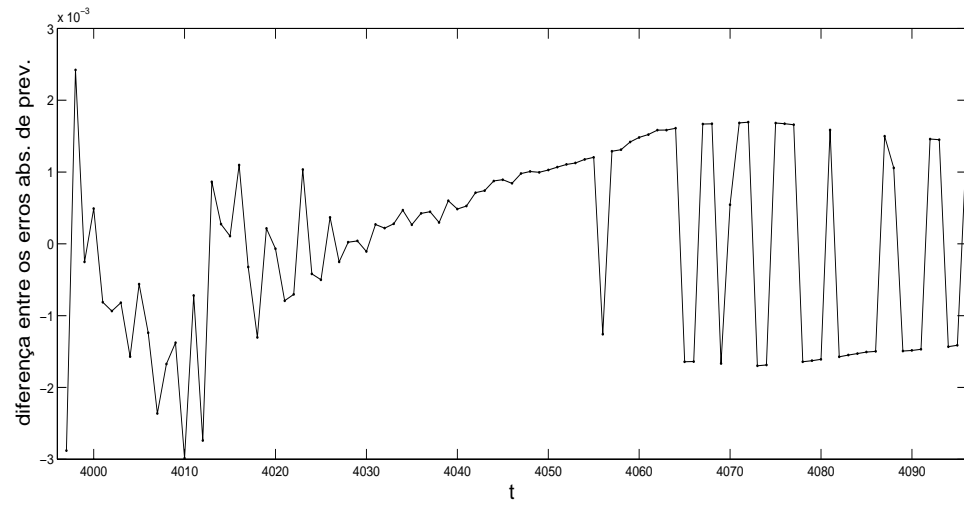


(a)

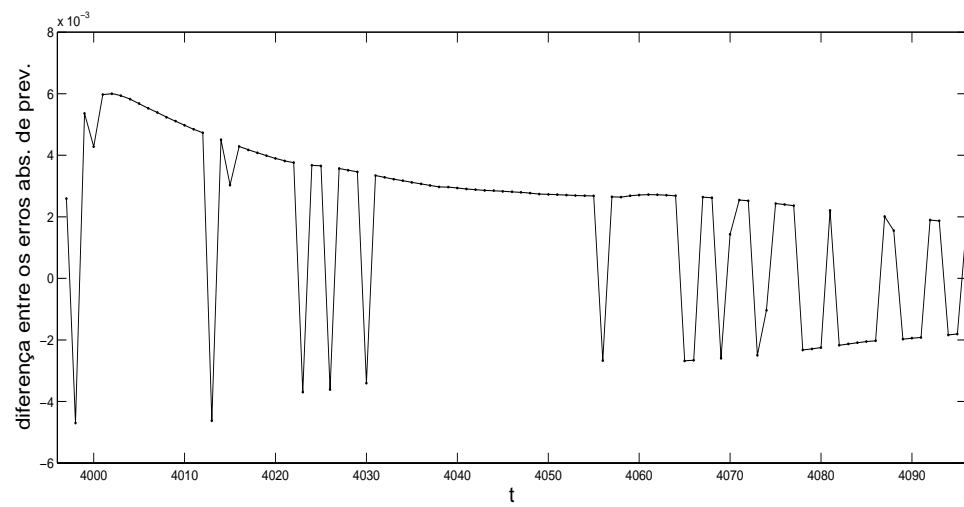


(b)

Figura 5.34: Série MWM($H = 0, 8$): (a) observações, previsões de passo- h e intervalos de 95% de confiança para as previsões dos modelos TARFIMA(100) e AR(29) e (b) observações, previsões de passo- h e intervalos de 95% de confiança para as previsões dos modelos TARFIMA(100) e AR(2) .

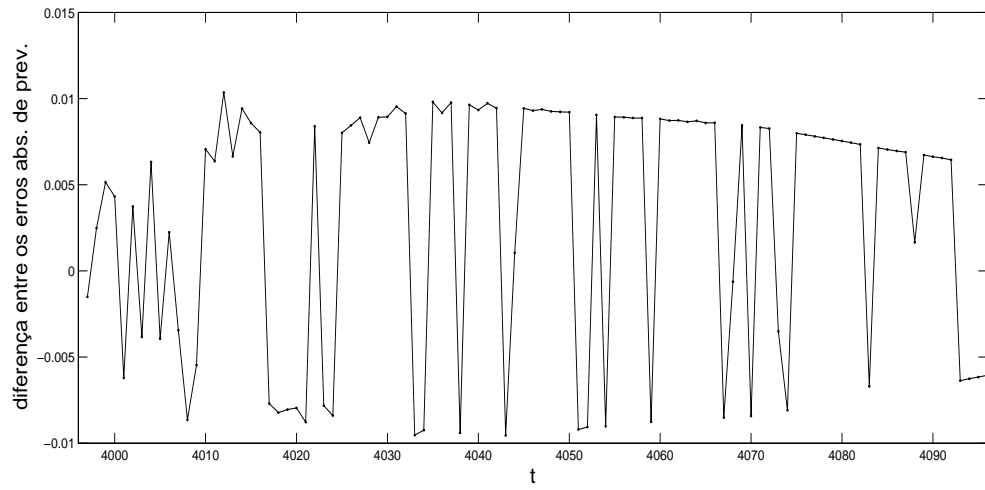


(a)

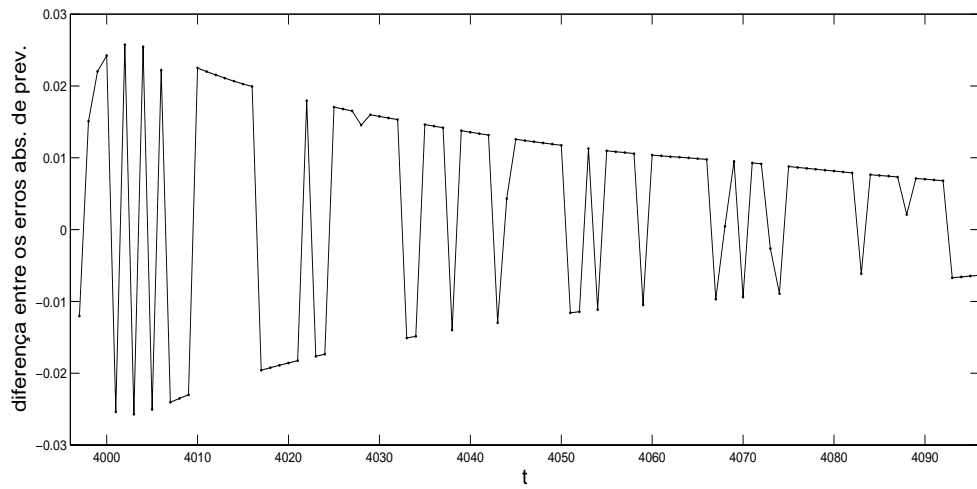


(b)

Figura 5.35: Série MWM($H = 0,9$): (a) diferença entre os erros absolutos de previsão dos modelos AR(28) e TARFIMA(100) e (b) diferença entre os erros absolutos de previsão dos modelos AR(2) e TARFIMA(100).



(a)



(b)

Figura 5.36: Série MWM($H = 0, 8$): (a) diferença entre os erros absolutos de previsão dos modelos AR(29) e TARFIMA(100) e (b) diferença entre os erros absolutos de previsão dos modelos AR(2) e TARFIMA(100).

A Tabela 5.15 ($PCEMSE_{x,y}(t = 3996, K)$) para a série MWM com $H = 0,9$ mostra que os modelos AR(28), TARFIMA(10), TARFIMA(50) e TARFIMA(100) apresentam desempenhos superiores aos do modelo AR(2) para $K = \{50, 100\}$, de acordo com o critério comparativo PCEMSE. Note também que os modelos TARFIMA(50) e TARFIMA(100) são levemente superiores ao AR(2) para $K = 7$. Os modelos AR(28), TARFIMA(50) e TARFIMA(100) apresentam desempenhos similares para $K = 100$. A Tabela 5.16 ($PCEMSE_{x,y}(t = 3996, K)$) para a série MWM com $H = 0,8$ mostra que as previsões dos modelos AR(29), TARFIMA(10), TARFIMA(50) e TARFIMA(100) para a série MWM($H = 0,8$) são melhores do que as previsões do modelo AR(2) para $K = \{7, 10, 20, 50\}$. As previsões dos modelos TARFIMA(50) e TARFIMA(100) são melhores do que as do AR(29) para $K = \{10, 20, 50\}$. Por outro lado, as previsões dos modelos AR(29), TARFIMA(10), TARFIMA(50) e TARFIMA(100) para $K = 100$ são um pouco piores do que as do modelo AR(2).

Tabela 5.15: Série MWM($H = 0,9$): $PCEMSE_{x,y}(t = 3996, K)$ dos modelos estimados AR(28), TARFIMA(10), TARFIMA(50) e TARFIMA(100) relativos ao AR(2). Os modelos TARFIMA(L) utilizaram $\hat{d} = 0,3572$.

K	AR(28)	L=10	L=50	L=100
7	0,0858	0,0118	-0,0148	-0,0148
10	0,1119	0,1891	0,0871	0,0473
20	0,1863	0,0668	0,2205	0,2888
50	-0,0865	-0,0673	-0,0625	-0,0024
100	-0,1671	-0,0430	-0,1797	-0,1646

Tabela 5.16: Série MWM($H = 0,8$): $PCEMSE_{x,y}(t = 3996, K)$ dos modelos estimados AR(29), TARFIMA(10), TARFIMA(50) e TARFIMA(100) relativos ao AR(2). Os modelos TARFIMA(L) utilizaram $\hat{d} = 0,3164$.

K	AR(29)	L=10	L=50	L=100
7	-0,0609	-0,0443	-0,0041	-0,0041
10	-0,2881	-0,2557	-0,3058	-0,3038
20	-0,2010	-0,0809	-0,2559	-0,2604
50	-0,0500	-0,0233	-0,1160	-0,1453
100	0,0500	0,0207	0,0447	0,0399

5.5.3 Série do Rio Nilo entre os anos de 1007 e 1206

A Fig. 5.37 mostra as previsões para a série do rio Nilo entre os anos de 1007 e 1206 com origem em $t = 100$ (ano 1106) e $K = 100$ obtidas por meio de modelos TARFIMA(100) e AR(8) (vide Fig. 5.37a) e TARFIMA(100) e AR(2) (vide Fig. 5.37b). O modelo TARFIMA(100) utilizou a estimativa $\hat{d} = 0,4125$, conforme a

Tabela 5.11 (essa estimativa é conservadora, dado que Beran [58] estima $\hat{d} = 0,463$ e Percival e Walden [67] estimam $\hat{d} = 0,4532$ para a mesma série). A Fig. 5.37b mostra que as previsões do modelo AR(2) convergem rapidamente para a média amostral da série e que a probabilidade de cobertura dos seus intervalos de previsão é pior do que a do modelo TARFIMA(100), uma vez que algumas amostras entre $t = 100$ e $t = 130$ têm valores maiores do que os limites superiores dos intervalos de confiança obtidos.

As Figs. 5.38a e 5.38b mostram as diferenças entre os erros absolutos de previsão dos modelos AR(8) e TARFIMA(100) e as diferenças entre os erros absolutos de previsão dos modelos AR(2) e TARFIMA(100), respectivamente. Observe que o desempenho do modelo TARFIMA(100) é superior, haja vista que os erros absolutos de previsão dos modelos AR tendem a ser maiores do que os do processo TARFIMA(100) ajustado.

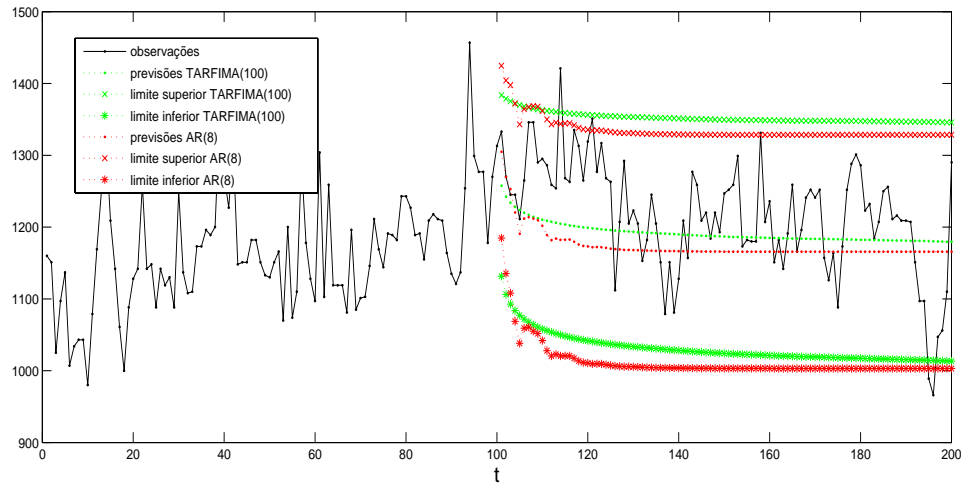
A Tabela 5.17 indica que os modelos TARFIMA(50) e TARFIMA(100) são melhores do que os modelos AR(2) e AR(8) para $K = 100$, conforme o critério PCEMSE.

Tabela 5.17: Série do rio Nilo: $PCEMSE_{x,y}(t = 1106, K)$ dos modelos estimados AR(8), TARFIMA(10), TARFIMA(50) e TARFIMA(100) relativos ao AR(2). Os modelos TARFIMA(L) utilizaram $\hat{d} = 0,4125$.

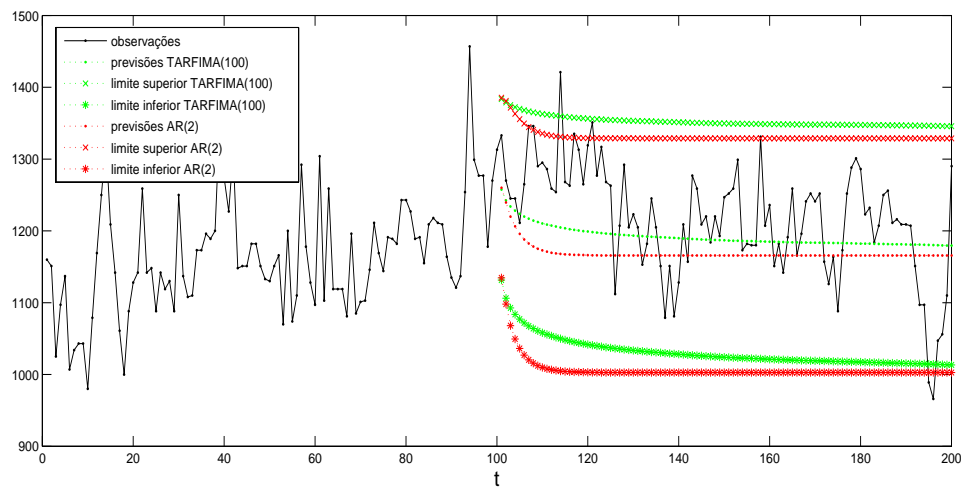
K	AR(8)	L=10	L=50	L=100
7	0,0796	0,1331	0,1852	0,2128
10	0,0925	0,0649	0,1332	0,1687
20	0,0516	0,0871	0,1864	0,2570
50	-0,0145	-0,0061	0,0098	0,0559
100	-0,1024	-0,1033	-0,2602	-0,2617

5.6 Conclusão

Este Capítulo apresentou um novo modelo de teletráfego LRD no espaço de estados denominado TARFIMA(L), que é uma versão aproximada do processo ARFIMA(p, d, q). O estudo apresentado neste Capítulo valida a aplicação do modelo de memória longa TARFIMA(L) em sinais (Gaussianos/não-Gaussianos) LRD. As simulações sugerem que a modelagem de sinais com memória longa por meio de processos AR é viável, pelo menos do ponto de vista teórico. Os experimentos computacionais com as séries ARFIMA e MWM demonstram que modelos AR de ordem alta podem oferecer um desempenho de previsão competitivo. O procedimento de inferência estatística “funciona” porque o número N de

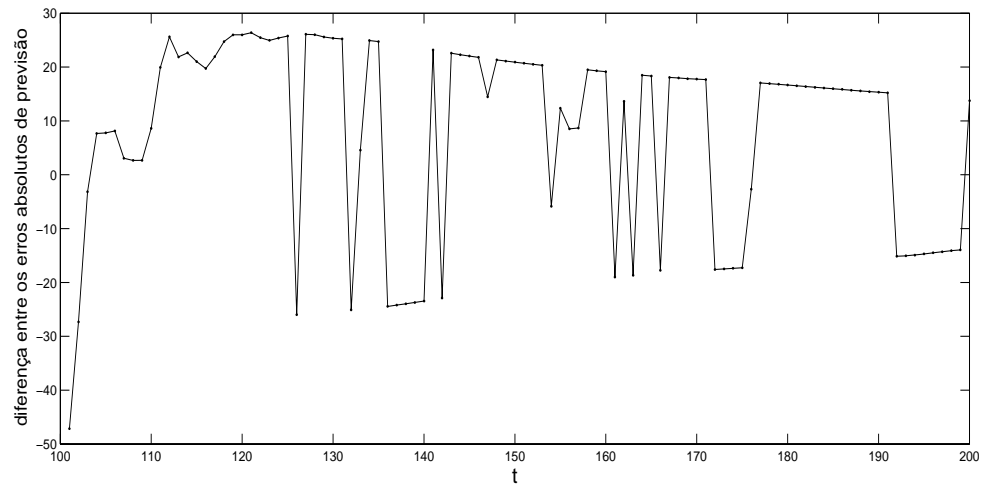


(a)

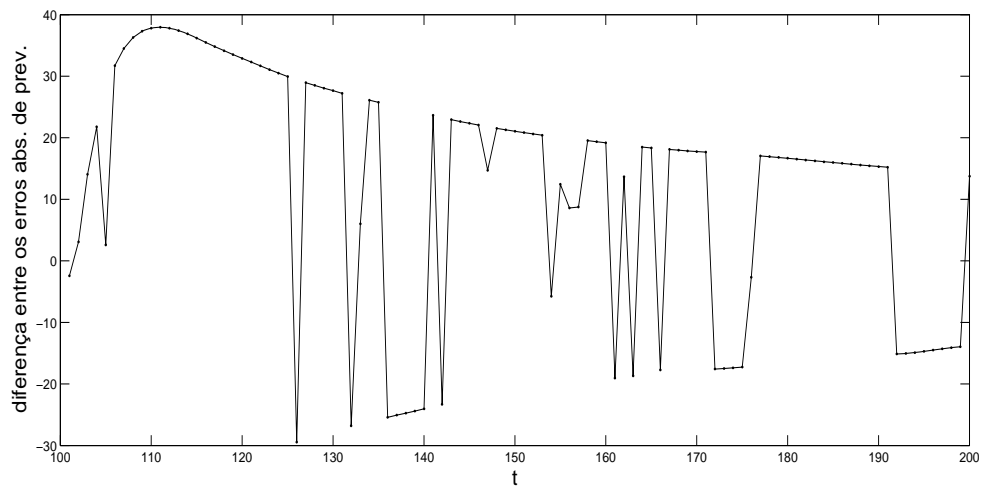


(b)

Figura 5.37: Série do rio Nilo: (a) observações, previsões de passo- h e intervalos de 95% de confiança para as previsões dos modelos TARFIMA(100) e AR(8) e (b) observações, previsões de passo- h e intervalos de 95% de confiança para as previsões dos modelos TARFIMA(100) e AR(2). O modelo TARFIMA(100) utilizou $\hat{d} = 0,4125$.



(a)



(b)

Figura 5.38: Série do rio Nilo: (a) diferença entre os erros absolutos de previsão dos modelos AR(8) e TARFIMA(100) e (b) diferença entre os erros absolutos de previsão dos modelos AR(2) e TARFIMA(100). O modelo TARFIMA(100) utilizou $\hat{d} = 0,4125$.

pontos do sinal é suficientemente grande (ou seja, o fato do modelo AR estimado não ser parcimonioso não é um problema).

Não obstante, observe-se que surgem algumas questões práticas relevantes como:

- a modelagem da memória longa pode ser obtida às custas do posicionamento de um ou mais pólos nas proximidades do círculo unitário (vide a Fig. 5.8, por exemplo); portanto, a implementação do modelo AR ajustado pode ser instável devido a efeitos de quantização dos coeficientes do filtro digital.
- o **problema fundamental da seleção da ordem** do modelo. Note-se que critérios de seleção tais como o AIC não levam em conta, de maneira explícita, a singularidade do tipo $1/f^\alpha$ que há na origem da DEP das séries LRD (vide (2.59)). O autor desta tese destaca o pensamento de Percival e Walden [44, pág.434] (segue transcrição abaixo):

“First, any order selection method we use should be appropriate for what we intend to do with the fitted AR model. (...) Most of the commonly used order selection criteria are geared toward selecting a low-order AR model that does well for one-step-ahead predictions”.

Em seguida, Percival e Walden descrevem quatro métodos de seleção (sendo um deles o critério AIC visto no Cap. 2). Finalmente, segue-se abaixo a citação de duas frases importantes do último parágrafo da mesma seção [44, pág.437]:

*“How well do these order selection criteria work in practice? (...) They conclude¹¹ that **subjective** judgment is still needed to select the AR order for actual time series.”.*

Depreende-se que a aplicação de modelos AR em esquemas de previsão *on-line* de teletráfego é fortemente prejudicada pelo problema da identificação da ordem do modelo. Por outro lado, o processo TARFIMA(L) tem as seguintes vantagens:

1. modela a região das baixas frequências do espectro de sinais LRD;
2. pode ser implementado em aplicações práticas tais como CAC e alocação dinâmica de banda, porque é finito-dimensional (ele não exige que a memória seja infinita, como é o caso para o modelo ARFIMA);

¹¹Aqui Percival e Walden estão se referindo ao artigo [152], de Kay e Marple.

3. a representação no espaço de estados viabiliza a utilização do filtro de Kalman.

A metodologia de análise estatística de sinais LRD empregada na seção 5.4 (sobre análise exploratória de séries LRD) é uma outra contribuição deste Capítulo. Os resultados obtidos sugerem que o desempenho do método (de estimativa do parâmetro H) R/S não é bom, confirmando o que afirmam Bardet *et al* na introdução do artigo [132].

6 Modelagem de Tráfego de Redes

Este Capítulo apresenta um estudo de caso baseado na análise do *trace* de tráfego Internet 2002_Apr_13_Sat_1930.7260.sk1.1ms.B_P.ts.gz [153], [70]. Esta série, que faz parte de um conjunto de vinte *traces* denominado “UNC02”, contém uma coleta de 2 horas de tráfego IP unidirecional na escala de 1 milissegundo de um enlace Gigabit Ethernet¹ existente entre o campus da University of North Carolina (UNC) at Chapel Hill e o seu *Internet Service Provider* (ISP) em 13/04/2002 no horário 19:30-21:30. Este *trace*, que consiste no registro das séries temporais de contagem de pacotes e de *bytes* no sentido *inbound* (ou seja, tráfego entrante no campus da UNC) por *bin* de 1 milissegundo, foi coletado pelo programa *tcpdump*, originalmente desenvolvido por Van Jacobson, Leres e McCanne [154], e encontra-se à disposição no endereço <http://www-dirt.cs.unc.edu/ts/>. De acordo com os autores do estudo [70], a coleta foi implementada com uma precisão de 1 microsegundo, sendo que foi observada uma taxa de perda de pacotes de 0,03% para o *trace* em consideração. Portanto, o efeito desta imprecisão em coletas de tráfego com *bins* de 1 ou 10 ms é desprezível. O restante desta tese se refere ao *trace* 2002_Apr_13_Sat_1930.7260.sk1.1ms.B_P.ts.gz simplesmente como *trace* UNC02. O artigo [70] também analisa um outro conjunto de *traces*, ao qual esta tese se refere como UNC03, coletados em abril de 2003.

Algumas das descobertas do estudo empírico de Park *et al* [70] são extremamente relevantes e têm aplicação imediata nesta tese e por isso estão resumidas a seguir:

- Foi observado um significativo decréscimo nos valores das estimativas do parâmetro H das séries de pacotes do conjunto de *traces* UNC03 com relação ao conjunto UNC02. Por outro lado, as estimativas de H para as séries de contagem de *bytes* mantiveram-se praticamente idênticas. Note-se que alguns autores, como Stoev *et al* [131], limitam-se à investigação das séries

¹ “Banda” ou vazão de 10^9 bits por segundo (bps), ou seja, 1 Gbps.

de contagem de pacotes por bin^2 . Para tal, argumentam que a experiência mostra que não há diferença importante entre o estudo das séries de *bytes* e de pacotes, porque o tamanho dos pacotes tende a ser igual ao *Maximum Transfer Unit* (MTU) (unidade máxima de transmissão³). Conforme mencionado acima, Park *et al* [70] mostraram que a realidade não confirma este tipo de argumento intuitivo e que pode haver uma diferença quantitativa significativa entre as estimativas de H obtidas das análises das séries de pacotes e de *bytes*. Além disso, observe-se que os mecanismos de controle de tráfego das redes de telecomunicações levam em conta a vazão (banda ou capacidade) em bps das interfaces de rede e não a vazão em pacotes por segundo⁴. Portanto, a estimativa de H deve ser feita sobre a série de *bytes* quando a finalidade é o controle do tráfego.

- Várias séries (das 56 que foram analisadas) não são Gaussianas.
- Estimativas do parâmetro H para as séries de contagem de *bytes* são tipicamente superiores a 0,8, indicando que o tráfego da UNC é fortemente LRD.
- O grau de memória longa do tráfego (estimativa do parâmetro H) **não** foi influenciado pelo dia da semana ou pelo horário do dia. Isto quer dizer que as flutuações no número de usuários ativos na rede do campus da UNC (população total de 35.000 pessoas) não influenciaram o grau de LRD.
- O grau de memória longa do tráfego **não** foi influenciado pelo nível de utilização (em % da banda) do enlace de acesso ou pelo número de conexões TCP ativas. Em particular, não foi observado qualquer declínio do parâmetro H com o aumento do número de conexões TCP.

6.1 Modelagem do *Trace* UNC02

6.1.1 Análise Exploratória

Nesta seção, faz-se uma análise exploratória das séries de pacotes e de *bytes* UNC02 na escala de 1 milissegundo e na escala agregada de 1 segundo. O restante deste trabalho se refere às séries de pacotes UNC02 nas escalas de 1 milissegundo e 1

²Stoev *et al* também analisaram os *traces* UNC02 e UNC03.

³Em redes de computadores, MTU corresponde ao tamanho do maior datagrama que uma camada de um protocolo de comunicação pode transmitir.

⁴Exemplos: a) uma interface E1 (da hierarquia digital plesiócrona) tem uma vazão de 2.048 kbps e b) uma interface *Synchronous Transport Module* (STM)-1, da hierarquia digital síncrona, trabalha a uma taxa de 155 Mbps.

segundo como UNC02Pktsbin1ms e UNC02Pktsbin1s, respectivamente, e às séries de *bytes* UNC02 nas escalas de 1 milissegundo e 1 segundo como UNC02bin1ms e UNC02bin1s, respectivamente.

Um dos objetivos fundamentais da análise de séries temporais é a construção de um modelo estocástico da série que está sendo investigada. O modelo pode ser estacionário ou não-estacionário, conforme visto no Cap. 2. Modelos ARMA ou ARFIMA, por exemplo, supõem que a série seja estacionária. Entretanto, a maioria das séries empíricas pode apresentar alguma forma de estacionariedade. Uma série pode ser estacionária durante um período muito longo ou pode ser estacionária apenas em períodos bastante curtos, mudando de nível e/ou inclinação. Sendo assim, é importante tentar responder às seguintes questões preliminares antes de se proceder à uma análise exploratória mais detalhada [155], [156]:

- A série sob investigação é **globalmente** estacionária?
- Caso a série não seja globalmente estacionária, pode-se identificar um encadeamento de sub-séries estacionárias, que podem ser geradas a partir de modelos (estacionários)?
- A série apresenta tendências determinísticas e/ou estocásticas, SRD e/ou LRD?
- Como detectar e estimar as mudanças de regime⁵ (conhecido na literatura como *change-point problem* ou problema do ponto de mudança) de uma série não estacionária?

A última questão, em especial, envolve o desenvolvimento de novas técnicas estatísticas, que incorporem a memória longa [131], [155]. De acordo com Kokoszka e Leipus [155], o problema do ponto de mudança tem sido extensivamente estudado para o caso de séries pouco correlacionadas. Somente mais recentemente (desde meados da década de 1990) é que o problema da mudança de regime no contexto da LRD tem recebido mais atenção. Kokoszka e Leipus [155] também afirmam que o estudo da LRD em conjunto com a análise de pontos de mudança deve ser realizado quando se tem uma série temporal LRD suficientemente grande (isto é, com um número muito grande de amostras), pois a possibilidade de ocorrência de pontos de mudança é maior do que nas séries pequenas, o que é o caso do tráfego das redes.

A **não-estacionariedade** do tráfego das redes para escalas de tempo da ordem de horas é intuitiva. Considere, por exemplo, a utilização de um enlace de

⁵Por exemplo, mudanças na média e na variância da série.

acesso à Internet numa rede corporativa como a da USP. Não há dúvida de que a utilização desse enlace durante os horários de pico da manhã e da tarde deva ser maior do que a do período da madrugada. Também é esperado que o tráfego apresente algum tipo de tendência determinística do tipo “rampa de subida” logo após o começo do expediente, que se dá por volta das 8h da manhã. Portanto, faz sentido afirmar que o tráfego das redes é não-estacionário quanto à média e/ou inclinação em escalas de horas. Por outro lado, vários estudos [1], [2], [70] (dentre inúmeros outros) mostram que é razoável caracterizar o tráfego Internet como sendo auto-similar e LRD (lembre-se de que o conceito de memória longa do Cap. 3 só foi definido para processo estacionário em sentido amplo) numa faixa de escalas que pode se estender de centenas de milissegundos a até centenas de segundos (esses limites superior e inferior podem variar em função da rede em questão). Veja, por exemplo, a Fig. 6.1, que ilustra uma coleta de 1 hora de tráfego na rede do LCS com *bins* de 100 milissegundos. Observe que esse sinal é “visualmente” não-estacionário quanto à média e variância (a sua potência varia com o tempo). Contudo, tem-se a impressão de que a série é **localmente estacionária** (flutuações homogêneas em torno de uma média local) em alguns trechos como aquele em que $1.700 \leq t \leq 10.000$ (limites aproximados).

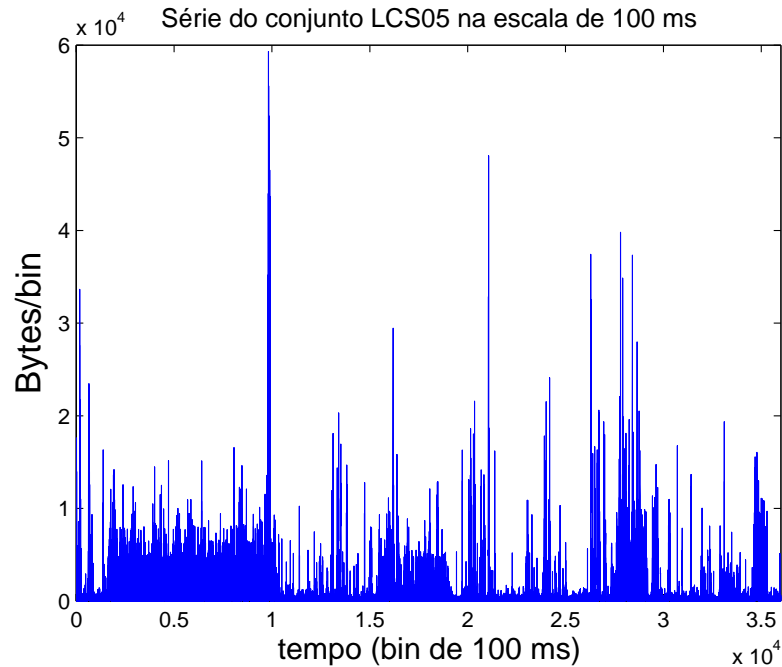


Figura 6.1: Tráfego coletado na rede local do LCS: série de contagem de *bytes* por *bin* de 100 milissegundos.

Stoev *et al* [131] afirmam que os sinais `UNC02Pktsbin1ms` e `UNC02Pktsbin1s` das Fig. 6.2a e 6.2b, respectivamente, podem ser consideradas estacionárias. A Fig. 6.3 mostra os QQ-plots de `UNC02Pktsbin1ms` e `UNC02Pktsbin1s`. O QQ-plot da Fig. 6.3a indica que a PDF de primeira ordem da série `UNC02Pktsbin1ms` é

assimétrica, não-Gaussiana e de cauda pesada (observe que os quantis empíricos nos extremos superior direito e inferior esquerdo da Fig. 6.3a não acompanham a linha vermelha com ponto e traço e que os quantis empíricos da extremidade superior são maiores do que os teóricos). Observe como o formato do QQ-plot da Fig. 6.3a tem uma certa semelhança com o QQ-plot da realização MWM com $H = 0,9$ da Fig. 5.11a. O QQ-plot da Fig. 6.3b sugere que a distribuição marginal de probabilidades de `UNC02Pktsbin1s` também é assimétrica, não-Gaussiana e com cauda superior pesada. De fato, o teste de normalidade de Jarque-Bera [157] (baseado nas medidas de assimetria e curtose) rejeitou a hipótese nula de normalidade de `UNC02Pktsbin1s`. A estatística desse teste é definida como

$$JB = \frac{N}{6} \left(\hat{A}^2 + \frac{(\hat{K} - 3)^2}{4} \right), \quad (6.1)$$

em que N é o número de amostras, \hat{A} é o estimador de assimetria

$$\hat{A} = \frac{1}{Ns_x^3} \sum_{t=1}^N (\mathbf{x}_t - \hat{\mu})^3 \quad (6.2)$$

em que $\hat{\mu}$ denota a média amostral, s_x^2 é a variância amostral e \hat{K} é o estimador de curtose

$$\hat{K} = \frac{1}{Ns_x^4} \sum_{t=1}^N (\mathbf{x}_t - \hat{\mu})^4. \quad (6.3)$$

Sob a hipótese nula de que os dados sejam normalmente distribuídos, espera-se que $JB \sim \chi^2(2)$ (chi-quadrada com dois graus de liberdade).

O programa `STABLE4.0` para `MATLAB` [95], que foi cedido pelo Dr. John P. Nolan (Robust Analysis, Inc.) para utilização nesta pesquisa, ajustou uma distribuição estável $S_\alpha(\sigma, \eta, \mu) = S_{1,92}(495; 0, 86; 2, 15 \times 10^4)$ para `UNC02Pktsbin1s`. A Fig. 6.4 ilustra os gráficos produzidos pela função `stablediag`, a saber: ajuste de PDF, QQ-plot (quantis da distribuição estável ajustada *vs* quantis empíricos), PP(*Probability-Probability*)-plot e ZZ-plot. O PP-plot é construído utilizando-se a CDF teórica, $F_x(x)$, do modelo ajustado. Os valores das amostras de dados, ordenadas em ordem crescente, são denotadas por $x(1), x(2), \dots, x(n)$. Para $i = 1, 2, \dots, n$, $F_x(x(i))$ (valor teórico) é representado contra $q_i = (i - 0,5)/n$ (valor empírico da CDF) [158]. O ZZ-plot é o gráfico do inverso da CDF teórica *vs* o inverso da CDF empírica (ou seja, pode ser obtido a partir do PP-plot). De acordo com Nolan [159], o diagnóstico de dados estáveis por meio do QQ-plot apresenta os seguintes problemas práticos (vide o QQ-plot da Fig. 6.4): a maior parte dos dados fica comprimida numa pequena região e b) há um desvio razoável dos dados das extremidades do QQ-plot com relação à linha teórica. É por isso

que Nolan recomenda o uso das ferramentas de diagnóstico PP-plot e ZZ-plot (vide [159] para maiores informações sobre o diagnóstico de dados que possuam distribuição com cauda pesada). Sendo assim, a série `UNC02Pktsbin1ms` possui cauda pesada e é bem modelada por uma distribuição estável, de acordo com o *toolbox* STABLE4.0.

A Fig. 6.5 ilustra os espectros *wavelet* de `UNC02Pktsbin1ms` e `UNC02Pktsbin1s`. Os gráficos 6.5a e 6.5b representam a estatística S_j como uma função da escala j para as séries `UNC02Pktsbin1ms` e `UNC02Pktsbin1s`, respectivamente. Conforme visto na seção 4.5.4, a estatística S_j quantifica a fração da potência da série na faixa de frequências $[1/2^{j+1}, 1/2^j]$ (lembre que um coeficiente *wavelet* (DWT) $w_{j,k}$ é proporcional à diferença entre duas médias adjacentes de um sinal x_t na escala de tempo $\tau_j = 2^{j-1}\Delta t$, em que Δt denota o intervalo de tempo entre duas observações sucessivas). Grandes valores de j correspondem às baixas frequências (escalas dilatadas ou mais lentas de tempo) e pequenos valores correspondem às altas frequências (escalas rápidas ou mais refinadas de tempo). Observe que o espectro *wavelet* da série `UNC02Pktsbin1ms` (Fig. 6.5a) nas escalas $1 \leq j \leq 5$ é quase constante. Esta característica é uma forte evidência de que flutuações de `UNC02Pktsbin1ms` nas escalas $\tau_j = 2^{j-1}\Delta t = \{1, 2, 4, 8, 16\}$ ms, em que $\Delta t = 1\text{ms}$, são pouco correlacionadas⁶. Este comportamento também foi observado em *traces* de tráfego capturados em enlaces de alta capacidade *Synchronous Optical network* (SONET)⁷ *Optical Carrier* (OC)-3, OC-12 e OC-48 (155 Mbps, 622 Mbps e 2,5 Gbps, respectivamente) da rede do ISP *tier-1* (nível 1) norte-americano Sprint⁸ [161]. Na Fig. 6.5a, observe que o espectro é aproximadamente linear para escalas $j \geq 10$ (ou $\tau_j \geq 512\text{ms}$). Portanto, `UNC02Pktsbin1ms` é LRD⁹ e assintoticamente auto-similar.

O espectro *wavelet* da série agregada `UNC02Pktsbin1s` (Fig. 6.5b) corresponde aos componentes do espectro da Fig. 6.5a para as escalas dilatadas $11 \leq$

⁶A estrutura de dependência de curta duração de um sinal está relacionada à forma da sua DEP na região das altas frequências, que, por sua vez, está relacionada às pequenas escalas do espectro *wavelet* [67], [131].

⁷Observe-se que enlaces SONET possuem um alto nível de multiplexação ou agregação de tráfego.

⁸A Internet é uma rede de redes. Na Internet pública, redes de acesso situadas na borda da Internet são conectadas ao restante segundo uma hierarquia de níveis de ISPs. Os ISPs de acesso estão no topo mais baixo da hierarquia. No topo dela está um número relativamente pequeno de ISPs (norte-americanos em sua maioria) denominados ISPs de nível 1, os quais apresentam as seguintes características: a) conectam-se diretamente a cada um dos outros ISPs de nível 1, conectam-se a um grande número de ISPs de nível 2 e têm cobertura internacional [160].

⁹É comum a literatura mencionar que sinais de tráfego podem apresentar, simultaneamente, características SRD e LRD. O espectro *wavelet* da Fig. 6.5a mostra que `UNC02Pktsbin1ms` é um exemplo de sinal que só apresenta a característica LRD, uma vez que a SRD se manifestaria por meio de um *spike* na região das escalas rápidas do espectro *wavelet*.

$j \leq 20$. Note-se que o espectro da Fig. 6.5b também é aproximadamente linear. Os parâmetros de Hurst de `UNC02Pktsbin1ms` e `UNC02Pktsbin1s`, estimados por meio do ajuste de uma linha ao espectro *wavelet* da Fig. 6.5a entre as escalas $j = 10, 11, \dots, 20$ e de uma linha ao espectro *wavelet* da Fig. 6.5b entre as escalas $j = 1, 2, \dots, 10$, é igual a $\hat{H} \approx 0,854$ e $\hat{H} \approx 0,922$, respectivamente.

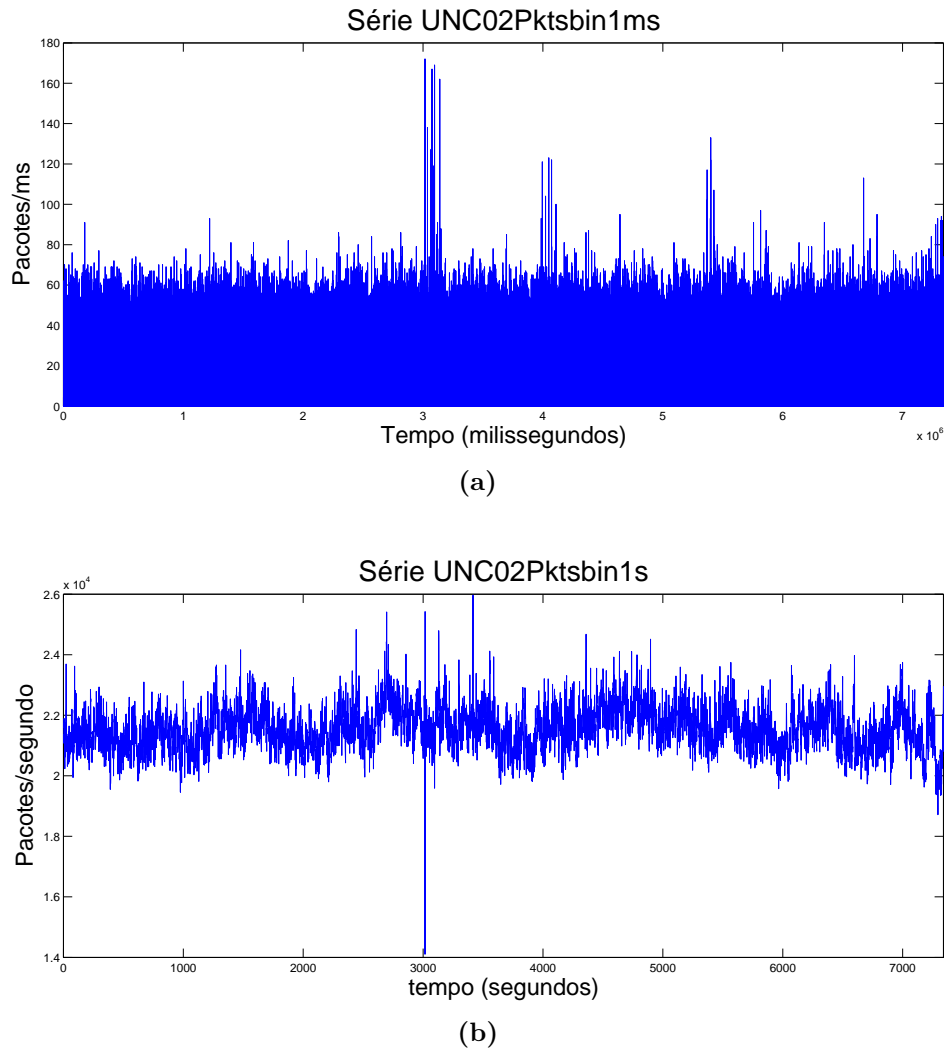
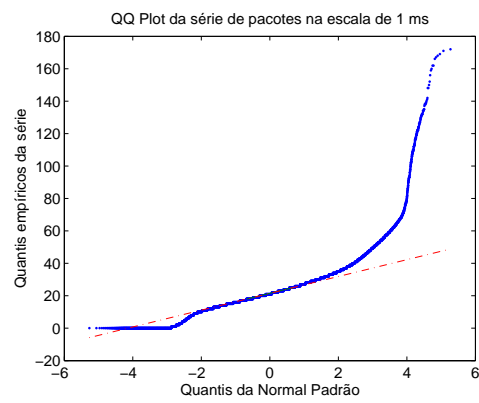
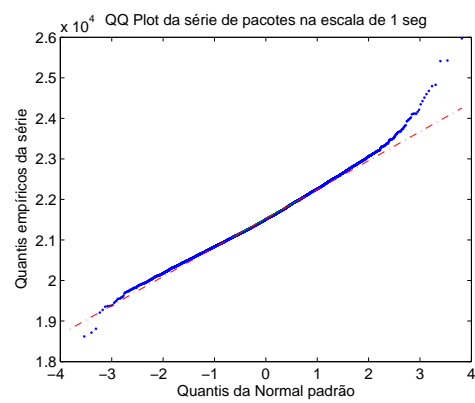


Figura 6.2: Tráfego entrante (série de contagem de pacotes por *bin*) num enlace de acesso Gigabit Ethernet da UNC em 13/04/2002, no horário 19:30-21:30, na escala de 1 milissegundo (`UNC02Pktsbin1ms`) (a) e na escala de 1 segundo (`UNC02Pktsbin1s`) (b).

A Fig. 6.6 sugere que é razoável admitir que as séries de contagem de *bytes* `UNC02bin1ms` e `UNC02bin1s`, que não foram analisadas pelo estudo de Stoev *et al* [131], sejam estacionárias. Como o foco deste trabalho é a modelagem de séries de *bytes* em escalas temporais maiores ou iguais a 1 segundo (com fins à implementação de mecanismos de controle preventivo de congestionamento em redes multimídia), o restante desta seção se concentra na análise da série agregada `UNC02bin1s`.



(a)



(b)

Figura 6.3: (a) QQ-plot de UNC02Pktsbin1ms e (b) QQ-plot de UNC02Pktsbin1s.

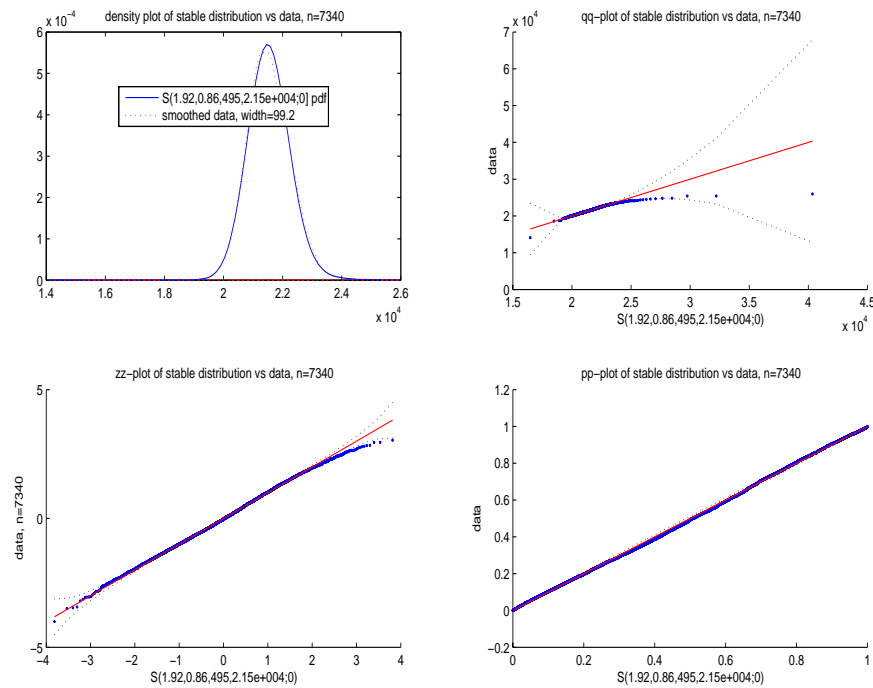


Figura 6.4: Ajuste de uma distribuição estável com parâmetros $\{\alpha = 1, 92, \sigma = 495, \eta = 0, 86, \mu = 2, 15 \times 10^4\}$, à série UNC02Pktsbin1s (gráfico do canto superior esquerdo), QQ-plot da distribuição estável *vs* os dados empíricos (canto superior direito), ZZ-plot da distribuição estável *vs* os dados empíricos (canto inferior esquerdo) e PP-plot da distribuição estável ajustada *vs* os dados empíricos (canto inferior direito). Os gráficos ZZ-plot e PP-plot sugerem que o ajuste da distribuição $S_{1,92}(495; 0, 86; 2, 15 \times 10^4)$ aos dados empíricos é bastante satisfatório.

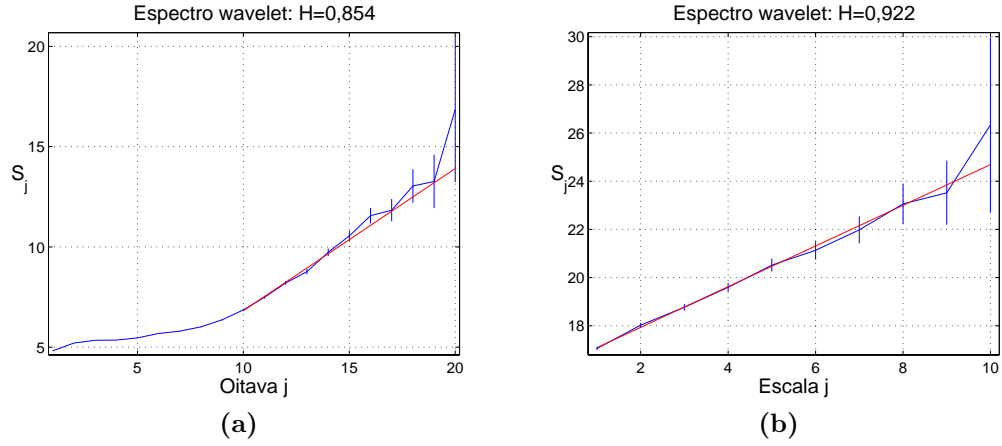


Figura 6.5: Espectros *wavelet* do *trace* UNC02 (em pacotes/*bin*) estimados por meio da série UNC02Pktsbin1ms (a) e por meio da série UNC02Pktsbin1s (b). Em (a), observe que o espectro é aproximadamente linear para escalas $j \geq 10$ (ou seja, para $\text{bins} \geq 1$ seg.). Os parâmetros de Hurst de UNC02Pktsbin1ms e UNC02Pktsbin1s, estimados por meio do ajuste de uma linha ao espectro *wavelet* (a) entre as escalas $j = 10, 11, \dots, 20$ e de uma linha ao espectro *wavelet* (b) entre as escalas $j = 1, 2, \dots, 10$, é igual a $\hat{H} \approx 0,854$ e $\hat{H} \approx 0,922$, respectivamente.

O resultado do teste de estacionariedade KPSS aplicado à série UNC02bin1s (vide Tabela 6.1) não rejeita a hipótese de estacionariedade. A Tabela 6.1 também mostra que UNC02bin1s não possui raízes unitárias, é linear, não-Gaussiana e possui memória longa de acordo com os testes R/S e GPH. A Fig. 6.7 mostra: a) o ajuste de uma distribuição estável com parâmetros $\{\alpha = 1,79, \sigma = 4,09 \times 10^5, \eta = 0,99, \mu = 6,35 \times 10^6\}$ à série UNC02bin1s no gráfico do canto superior esquerdo, b) o QQ-plot da distribuição estável ajustada *vs* os dados empíricos no canto superior direito, c) o ZZ-plot da distribuição estável estimada *vs* os dados empíricos no canto inferior esquerdo e d) o PP-plot da distribuição estável estimada *vs* os dados empíricos no canto inferior direito. Os gráficos ZZ-plot e PP-plot indicam que a distribuição marginal de probabilidades da série UNC02bin1s é bem modelada pela distribuição estável $S_{1,79}(4,09 \times 10^5; 0,99; 6,35 \times 10^6)$, o que explica o fato deste sinal ser linear porém não-Gaussiano.

O espectro *wavelet* da Fig. 6.8a é aproximadamente linear para as escalas $j \geq 7$ (ou seja, para $\tau_j \geq 64$ ms) e aproximadamente constante nas escalas $1 \leq j \leq 6$, o que quer dizer que flutuações de UNC02bin1ms nas escalas de $1, 2, \dots, 32$ ms são pouco correlacionadas (comportamento similar foi observado no espectro *wavelet* da Fig. 6.5a). Por outro lado, o espectro *wavelet* da Fig. 6.8b é aproximadamente linear em todas as escalas. Os parâmetros de Hurst de UNC02bin1ms e UNC02bin1s, estimados por meio do ajuste de uma linha ao es-

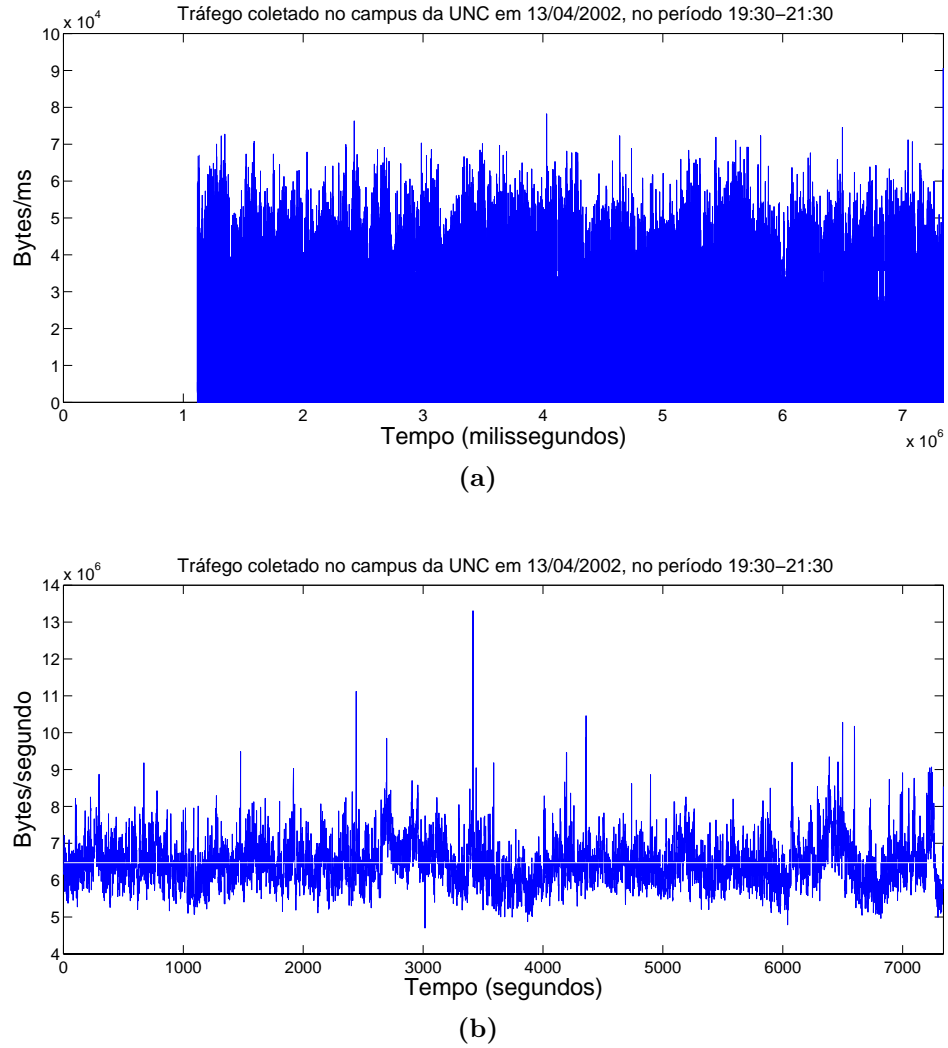


Figura 6.6: Tráfego entrante (série de contagem de *bytes* por *bin*) num enlace de acesso Gigabit Ethernet da UNC em 13/04/2002, no horário 19:30-21:30, na escala de 1 milissegundo (UNC02bin1ms) (a) e na escala de 1 segundo (UNC02bin1s) (b). O tráfego médio é igual a 51,623 Mbps (ou 6,4529 Mbytes/s) e é indicado pela linha branca em (b), o que implica uma utilização do enlace de aproximadamente 5,16%.

Tabela 6.1: Série UNC02bin1s: resultados de testes estatísticos (estacionariedade, memória longa, normalidade e linearidade) e quantificação do número de raízes unitárias (coluna \hat{m}). Os testes de Hinich, que utilizam a FFT (*Fast Fourier Transform*), foram aplicados ao primeiro bloco de 4096 amostras de UNC02bin1s (que contém 7340 amostras, um número que não é uma potência de 2).

x_t é $I(0)$?	x_t é LRD?		Hinich		\hat{m}
KPSS	R/S	GPH	x_t é normal?	x_t é linear?	
SIM	SIM	SIM	NÃO	SIM	0
$p_{af} > 0,05$	$p_{af} < 0,01$	$p_{af} < 0,01$	$p_{af} = 0$		

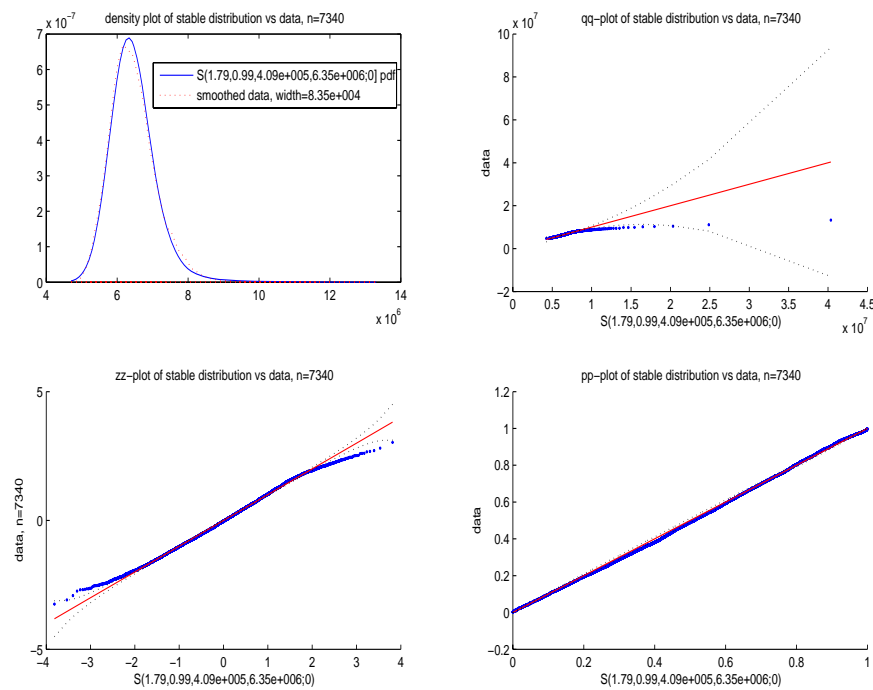


Figura 6.7: Ajuste de uma PDF estável com parâmetros $\{\alpha = 1,79, \sigma = 4,09 \times 10^5, \eta = 0,99, \mu = 6,35 \times 10^6\}$, à série `UNC02bin1s` (gráfico do canto superior esquerdo), QQ-plot da distribuição estável estimada *vs* os dados empíricos (canto superior direito), ZZ-plot da distribuição estável ajustada *vs* os dados empíricos (canto inferior esquerdo) e PP-plot da distribuição estável estimada *vs* os dados empíricos (canto inferior direito). Os gráficos ZZ-plot e PP-plot sugerem que os dados da série são bem modelados pela distribuição $S_{1,79}(4,09 \times 10^5; 0,99; 6,35 \times 10^6)$.

pectro *wavelet* 6.8a entre as escalas $j = 10, 11, \dots, 20$ e de uma linha ao espectro *wavelet* 6.8b entre as escalas $j = 1, 2, \dots, 10$, é igual a $\hat{H} \approx 0,839$ e $\hat{H} \approx 0,906$, respectivamente.

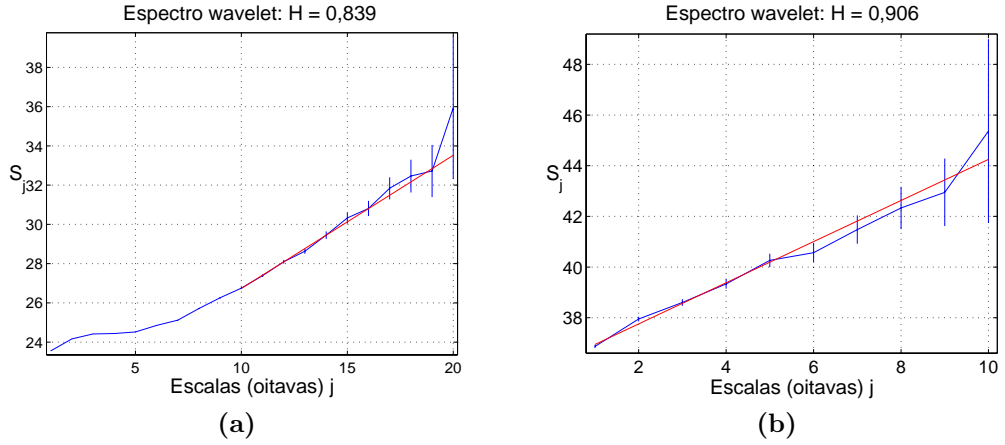


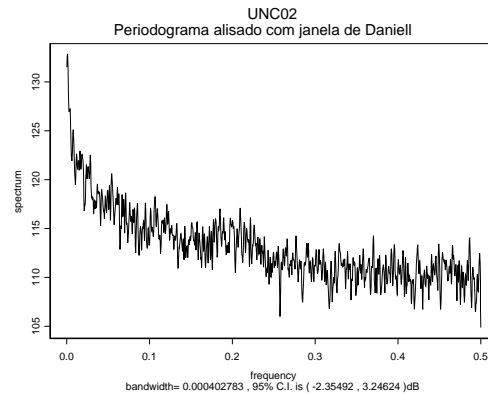
Figura 6.8: Espectros *wavelet* do trace UNC02 (em *bytes/bin*) estimados por meio da série UNC02bin1ms (a) e por meio da série UNC02bin1s (b). Em (a), observe que o espectro é aproximadamente linear para escalas $j \geq 7$ (ou seja, para *bins* > 128 ms). O parâmetro de Hurst, estimado por meio do ajuste de uma linha ao espectro *wavelet* (a) entre as escalas $j = 10, 11, \dots, 20$ e de uma linha ao espectro *wavelet* (b) entre as escalas $j = 1, 2, \dots, 10$, é igual a $\hat{H} \approx 0,839$ e $\hat{H} \approx 0,906$, respectivamente.

A Tabela 6.2 resume as estimativas do parâmetro d para a série UNC02bin1s obtidas pelos métodos R/S, periodograma (com janela de Daniell), Whittle, Haslett-Raftery e Abry-Veitch. Observe que o resultado da análise R/S é discordante dos demais, os quais estão situados na faixa $0,37 \leq \hat{d} \leq 0,43$. A Fig. 6.9 mostra o periodograma alisado com janela de Daniell (Fig. 6.9a), o periodograma alisado pelo método WOSA (Fig. 6.9b) e a SACF da série UNC02bin1s (Fig. 6.9c, que também mostra as ACFs teóricas dos modelos AR(18) e ARFIMA(0;0,3717;1) estimados para a série). A Fig. 6.9c mostra que, diferentemente do modelo ARFIMA(0;0,3717;1), o processo AR(18) não modela o decaimento lento da SACF da série UNC02bin1s.

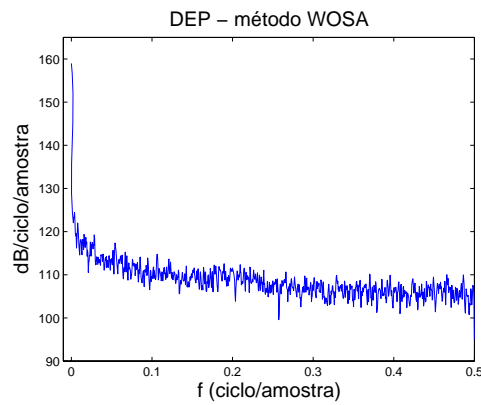
Tabela 6.2: Estimativas do parâmetro d para a série UNC02bin1s.

análise	periodograma	estimador	estimador de MV	estimador
R/S	com janela	de Whittle	de Haslett-Raftery	<i>wavelet</i>
	de Daniell			Abry/Veitch
0,3125	0,3854	0,4225	0,3717	0,406

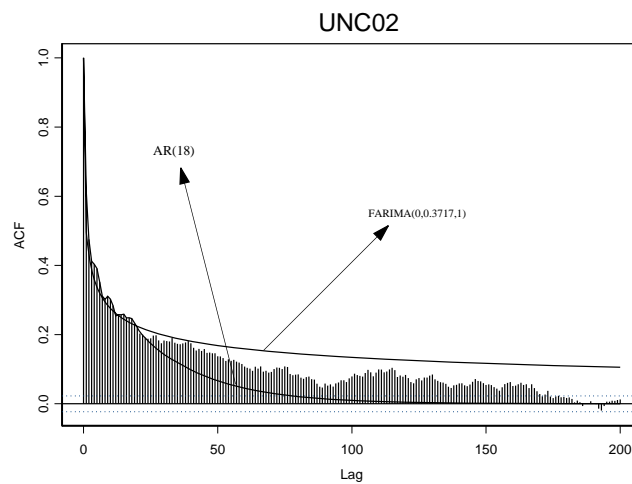
As Figs. 6.10a, 6.10b e 6.10c ilustram o QQ-plot, o periodograma alisado e a SACF dos resíduos do modelo ARFIMA(0;0,3717;1) estimado para a série



(a)



(b)



(c)

Figura 6.9: Série UNC02bin1s: (a) periodograma alisado com janela de Daniell, (b) periodograma alisado pelo método WOSA e (c) SACF e ACFs teóricas dos modelos AR(18) e ARFIMA(0; 0, 3717; 1) estimados.

UNC02bin1s. As Figs. 6.11a e 6.11b mostram o periodograma alisado e a SACF dos resíduos do modelo TARFIMA(100) ajustado à série UNC02bin1s (foi utilizado o valor $\hat{d} = 0,3717$ estimado pelo método de Haslett-Raftery). As Figs. 6.12a e 6.12b mostram as SACFs dos resíduos dos modelos AR(18) e AR(2) estimados pelo critério AIC, respectivamente. As Figs. 6.13a e 6.13b mostram os periodogramas alisados dos resíduos dos modelos AR(18) e AR(2) ajustados, respectivamente.

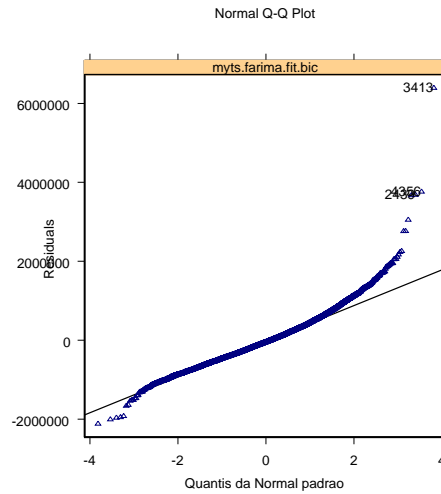
De acordo com as Figs. 6.10c, 6.11b, 6.12a (SACFs dos resíduos dos modelos ARFIMA, TARFIMA(100) e AR(18), respectivamente) e 6.10b, 6.11a e 6.13a (periodogramas dos resíduos dos modelos ARFIMA, TARFIMA(100) e AR(18), respectivamente), os resíduos dos modelos ARFIMA, TARFIMA(100) e AR(18) se comportam como RB. Por outro lado, as Figs. 6.12b e 6.13b mostram que a série dos resíduos do modelo AR(2) estimado não é um RB (veja a singularidade na origem do periodograma da Fig. 6.13b).

A Fig. 6.14 mostra os diagramas de pólos e zeros dos modelos AR(18) (Fig. 6.14a) e AR(2) (Fig. 6.14b). Note que o pólo real positivo do processo AR(18) estimado está bastante próximo do círculo unitário (este comportamento também foi observado para as séries MWM com $H = \{0,9; 0,8; 0,7\}$ e ARFIMA com $d = \{0,4; 0,3\}$ do Cap. 5).

6.1.2 Análise Local da Memória Longa do *Trace* UNC02

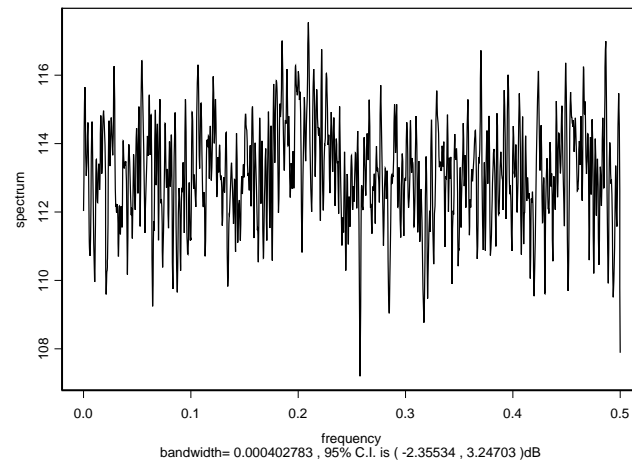
Na prática, os métodos de estimação do parâmetro de Hurst podem ser severamente afetados por não-estacionariedades existentes numa determinada série, tais como mudanças bruscas do seu nível médio. De acordo com a literatura [131], [162], este tipo de mudança de regime tende a provocar uma sobre-estimativa do parâmetro H ($\hat{H} \geq 1$) efetuada pelo método *wavelet* Abry-Veitch (portanto, a obtenção de uma estimativa $\hat{H} \geq 1$ é uma indicação da presença de não-estacionariedade). Sendo assim, a premissa, freqüentemente adotada, de que sinais de tráfego LRD podem ser modelados por meio de um processo estacionário caracterizado por um parâmetro H **global** pode não ser realista para muitos casos encontrados na prática.

Abry e Veitch [122] efetuaram um estudo preliminar da estacionariedade de três séries (**pAug**, **pOct** e **OctExt**) do conjunto de *traces* de tráfego Ethernet coletados por Leland *et al* [1] na rede local da Bellcore no início da década de 1990 por meio do exame da variação temporal do parâmetro de Hurst. Os *traces* **pAug** e **pOct** têm menos de 1 hora de duração. O *trace* **OctExt** tem 35 horas

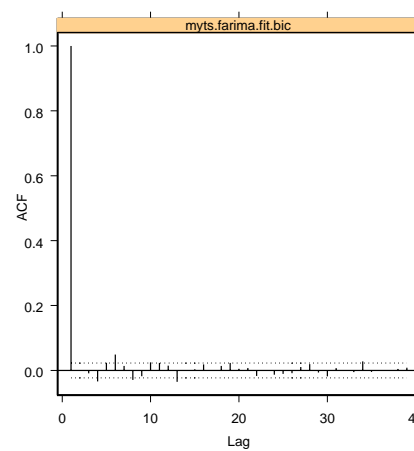


(a)

Periodograma dos resíduos do modelo ARFIMA(0;0,3717;1)

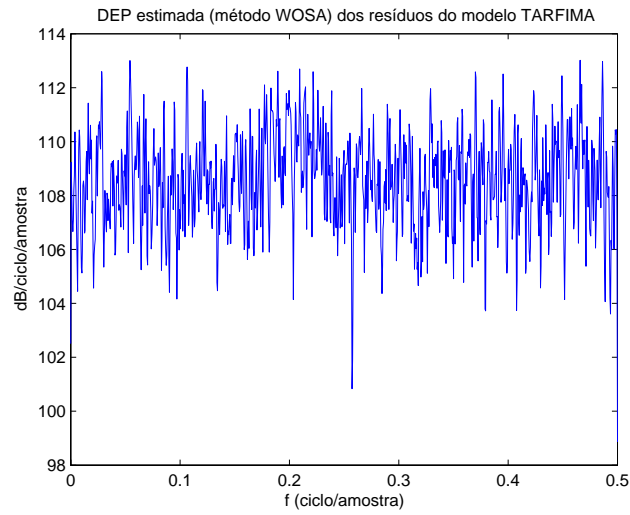


(b)



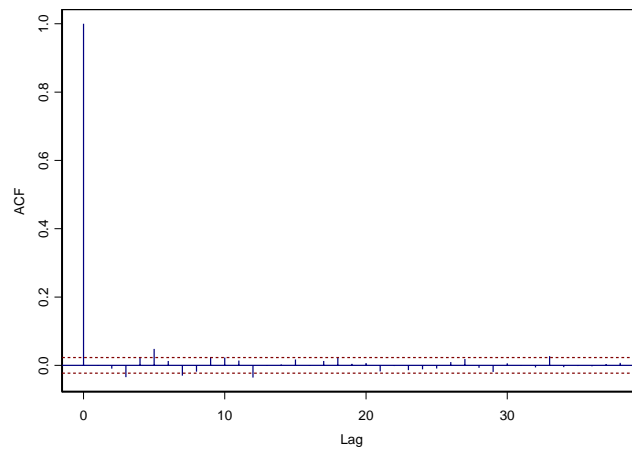
(c)

Figura 6.10: Resíduos do modelo ARFIMA(0;0,3717;1) estimado para a série UNC02bin1s: (a) QQ-plot, (b) periodograma e (c) SACF.



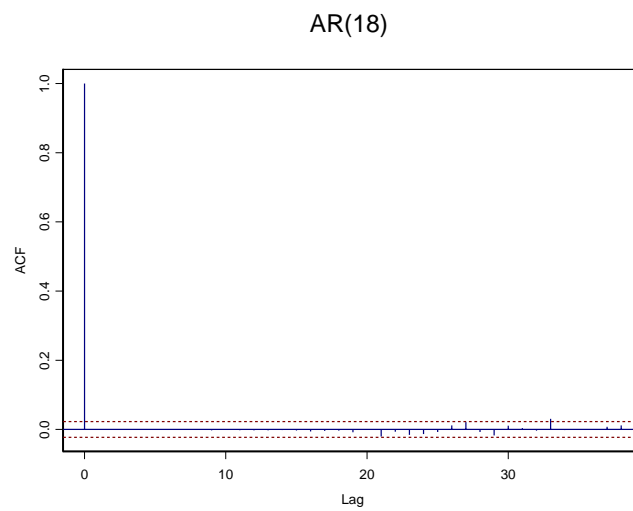
(a)

TARFIMA(100)

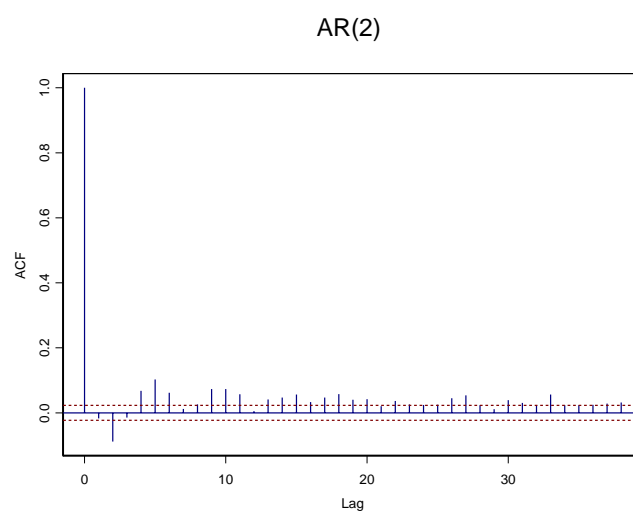


(b)

Figura 6.11: Resíduos do modelo TARFIMA(100) ajustado à série UNC02bin1s: (a) periodograma e (b) SACF.

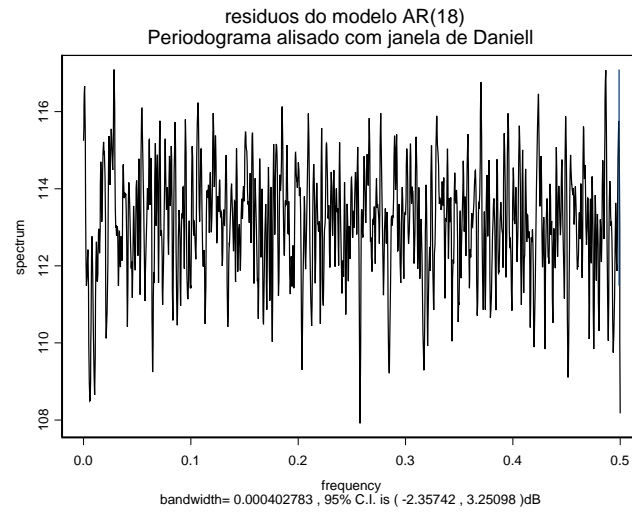


(a)

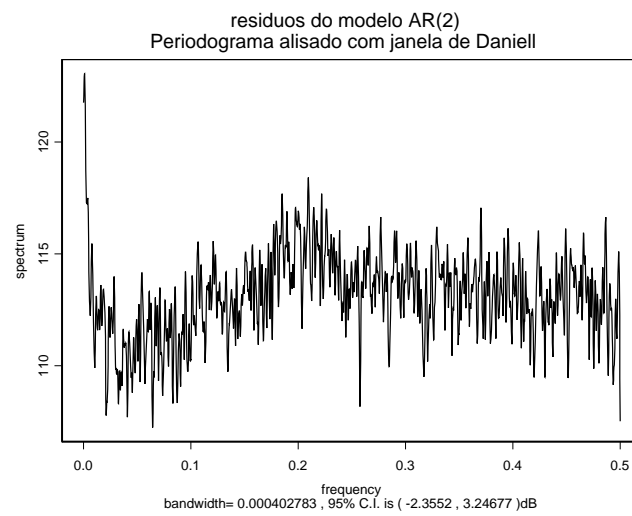


(b)

Figura 6.12: SACF dos resíduos dos modelos AR(18) (a) e AR(2) (b) ajustados à série *UNC02bin1s*.

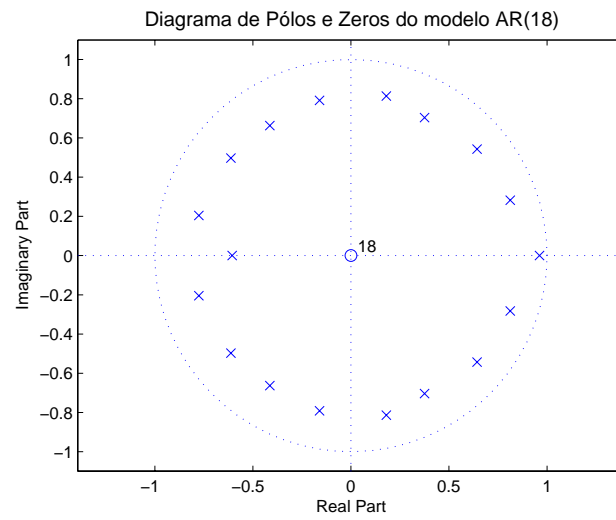


(a)

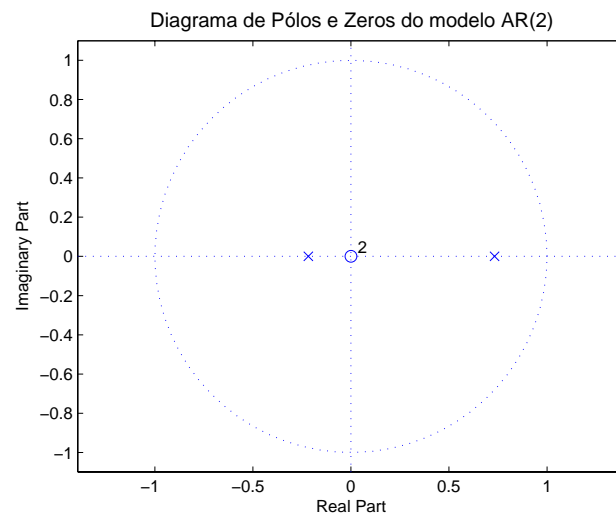


(b)

Figura 6.13: Periodogramas: (a) resíduos do modelo AR(18) e (b) resíduos do modelo AR(2).



(a)



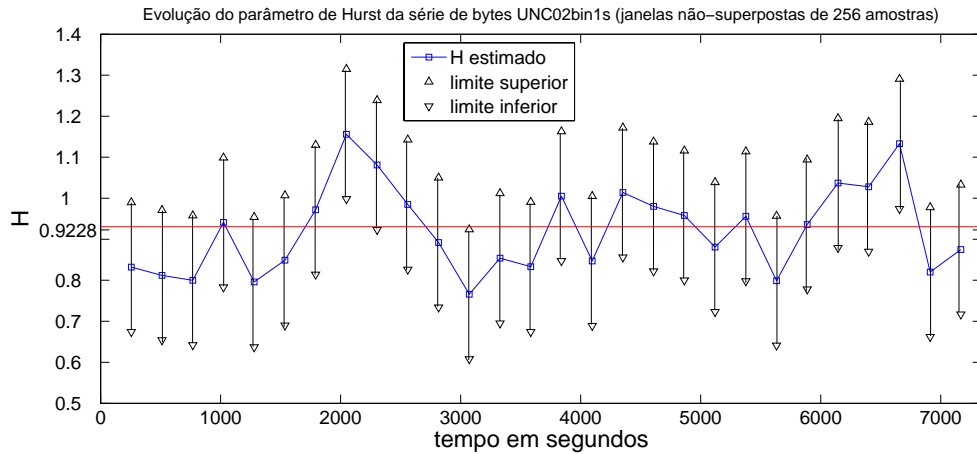
(b)

Figura 6.14: Diagrama de pólos e zeros dos modelos AR(18) (a) e AR(2) (b).

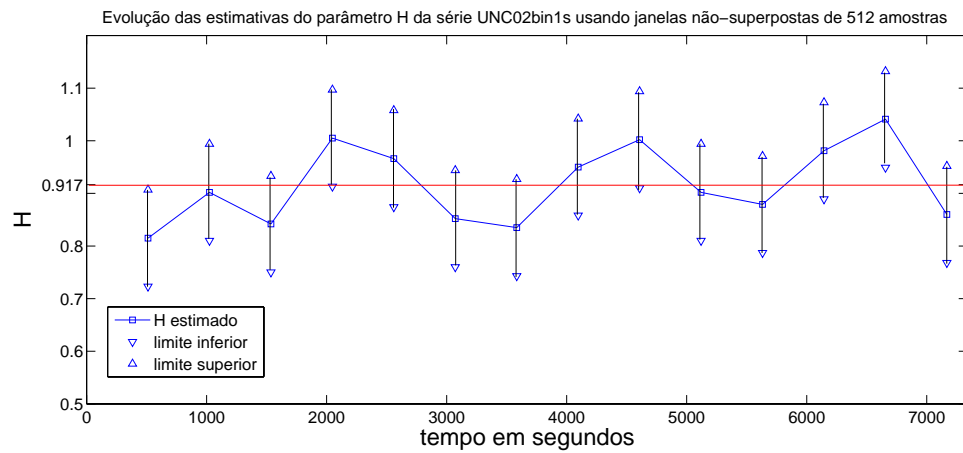
de duração. Abry e Veitch dividiram as séries em 32 segmentos contíguos sobre os quais foram calculadas estimativas de H . Este procedimento é justificado pelo fato dos coeficientes *wavelet* de sinais LRD serem aproximadamente não-correlacionados no plano tempo-escala (mais precisamente, pode-se afirmar que não há correlação entre os coeficientes *wavelet* de uma mesma escala e que a correlação entre escalas diferentes é fraca) [67], [110]. Para os *traces* **pAug** e **pOct**, Abry e Veitch observaram que as médias das 32 estimativas ficaram próximas dos valores estimados para H quando foram utilizadas todas as amostras das séries. De acordo com esses pesquisadores, esta é uma boa evidência heurística de que H se mantém constante ao longo daqueles *traces*. Por outro lado, as estimativas de H para o *trace* **OctExt** apresentam uma grande variabilidade, o que indica de forma clara a sua não-estacionariedade (a média das 32 estimativas difere significativamente da estimativa de H obtida sobre todo o *trace*). Entretanto, esta observação não surpreende, haja vista o ciclo diurno exibido pela série (conforme [122]).

As Figs. 6.15a e 6.15b ilustram a variação temporal do parâmetro de Hurst (gráficos das estimativas de H vs tempo em segundos) da série **UNC02bin1s** da Fig. 6.15c. As estimativas de H em 6.15a e 6.15b foram obtidas por meio da aplicação do estimador Abry-Veitch sobre segmentos de dados sucessivos e não-superpostos de 256 e 512 amostras, respectivamente. Os limites superior e inferior dos intervalos de confiança a 95% também estão representados nas Figs. 6.15a e 6.15b. Observe-se que a variabilidade da estimativa de H diminui com o aumento do número de amostras do segmento de dados. As médias das estimativas nas Figs. 6.15a e 6.15b estão indicadas pelas linhas horizontais vermelhas e correspondem a 0,9228 e 0,9166, respectivamente. Note-se que elas estão razoavelmente próximas da estimativa $\hat{H} = 0,906$, que foi obtida utilizando-se todas as amostras do sinal. As Figs. 6.15a e 6.15b indicam que o sinal **UNC02bin1s** pode ser considerado estacionário.

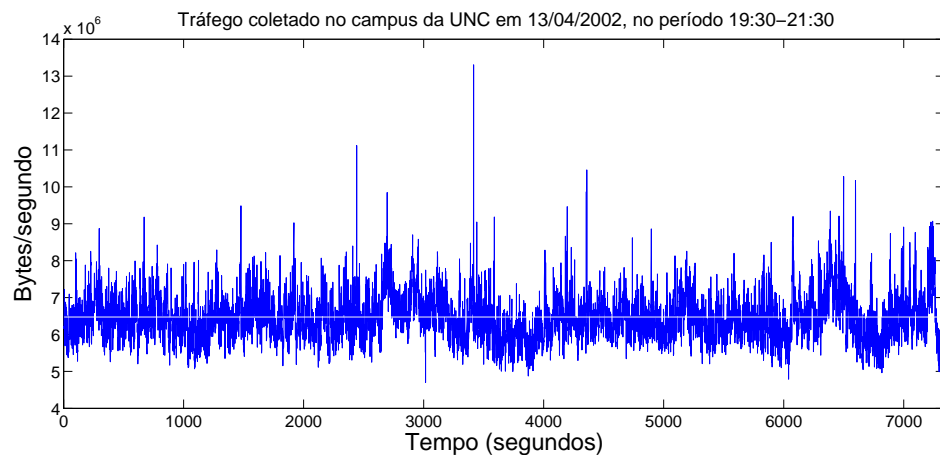
A análise local do parâmetro de Hurst ilustrada pela Fig. 6.15 é **cega** porque: a) não se sabe, a priori, a localização dos pontos de mudança de modelo gerador e b) não há um critério para a escolha do tamanho da janela de dados. Não obstante, essa **análise empírica** sugere que o janelamento com um número relativamente pequeno de amostras (como a que se fez com janelas de tamanho igual a 256) pode provocar um aumento da variância do estimador, o que não é desejável em aplicações práticas. Por outro lado, a análise local empírica com janelas de 512 pontos parece produzir resultados mais consistentes. Percival e Walden [67], por exemplo, analisam a variância *wavelet* (é o termo usado por es-



(a)



(b)



(c)

Figura 6.15: As Figs. (a) e (b) mostram análises locais do parâmetro de Hurst da série UNC02bin1s da Fig. (c). As estimativas de H em (a) e (b) foram obtidas por meio da aplicação do estimador Abry-Veitch sobre janelas de dados sucessivas e não-superpostas de 256 e 512 amostras, respectivamente. As médias das estimativas em (a) e (b) estão indicadas pelas linhas horizontais vermelhas e correspondem a 0,9228 e 0,9166, respectivamente.

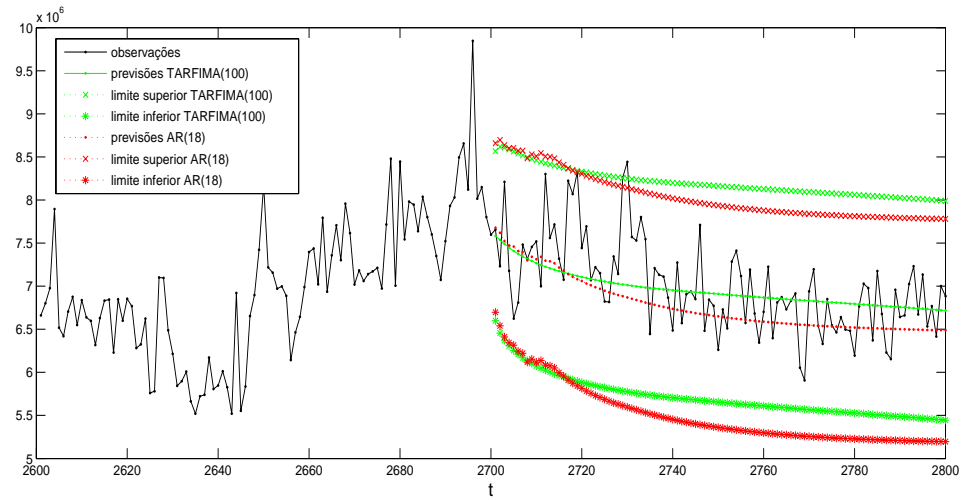
ses autores para denominar o espectro *wavelet*) da série do Rio Nilo entre os anos de 773 e 1284 (análise *wavelet* com 512 pontos). Conforme explicado pela seção 6.1.1, a detecção/estimação dos pontos de mudanças de regime de séries LRD ainda é um assunto pouco entendido. Essa questão ainda parece ser mais complicada no contexto de modelagem de teletráfego, onde se requer que a estimação seja *on-line*.

6.1.3 Previsão Empírica com o Modelo TARFIMA

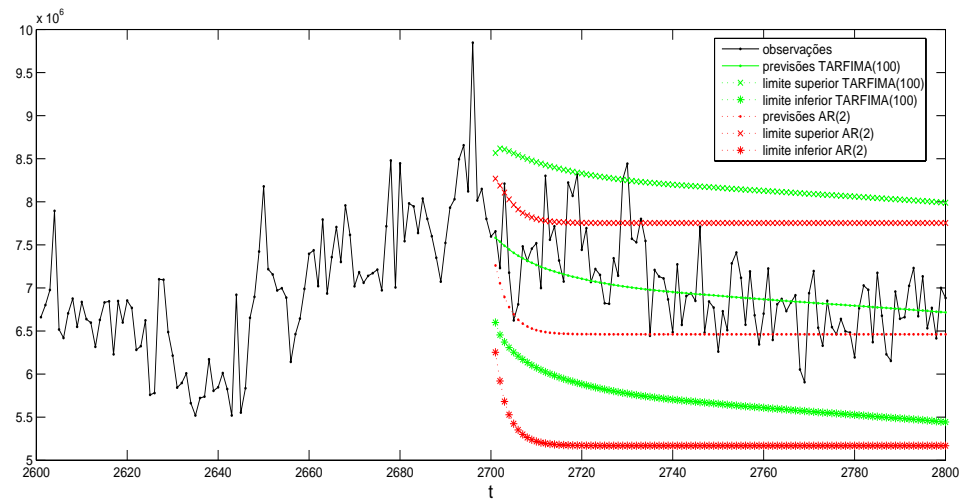
A Fig. 6.16 ilustra previsões para a série `UNC02bin1s` com origem em $t = 2700$ que foram obtidas por meio de modelos TARFIMA(100) e AR(18) (vide Fig. 6.16a) e TARFIMA(100) e AR(2) (vide Fig. 6.16b). O modelo TARFIMA(100) utilizou a estimativa (conservadora) $\hat{d} = 0,3717$ (vide Tabela 6.2). O processo AR(18) é o melhor modelo AR de acordo com o critério AIC. Os modelos AR(2) e AR(18) foram estimados pela função **S-PLUS**[®] `AR` (método de Yule-Walker [151, pág.94]).

A Fig. 6.16b mostra que as previsões do modelo AR(2) convergem rapidamente para a média amostral da série e que os seus intervalos de previsão não têm uma boa probabilidade de cobertura, uma vez que várias amostras entre $t = 2700$ e $t = 2740$ têm valores maiores do que os limites superiores dos intervalos de confiança obtidos. Isto também é ilustrado pela Fig. 6.17b que mostra a diferença entre os erros absolutos de previsão dos modelos AR(2) e TARFIMA(100) (observe que o erro absoluto de previsão do modelo AR(2) é quase sempre maior do que o do modelo TARFIMA(100)). Por outro lado, as Figs. 6.16b e 6.17b indicam que os intervalos de previsão do modelo TARFIMA(100) têm uma probabilidade de cobertura satisfatória. As Figs. 6.16a e 6.17a sugerem que os modelos TARFIMA(100) e AR(18) têm desempenhos similares. De fato, a Fig. 6.17a sugere que o modelo TARFIMA(100) tem um desempenho ligeiramente superior, haja vista que o erro absoluto de previsão do modelo AR(18) tende a ser maior do que o do processo TARFIMA(100) para $h \leq 40$ ($2701 \leq t \leq 2740$).

A Tabela 6.3 contém os valores de $PCEMSE_{x,y}(t = 2700, K)$ para $K = \{7, 10, 20, 50, 100\}$ dos processos AR(18) e TARFIMA(L), $L = 10, 50, 100$ (utilizou-se $\hat{H} = 0,8717$), relativos ao modelo AR(2). A Tabela 6.3 mostra que a) o modelo TARFIMA(100) apresenta o melhor desempenho do ponto de vista do critério PCEMSE e que b) os modelos AR(18), TARFIMA(50) e TARFIMA(100) têm um desempenho similar. Note-se que os resultados obtidos confirmam a teoria exposta pela seção 5.3: a precisão das previsões aumenta com o aumento da

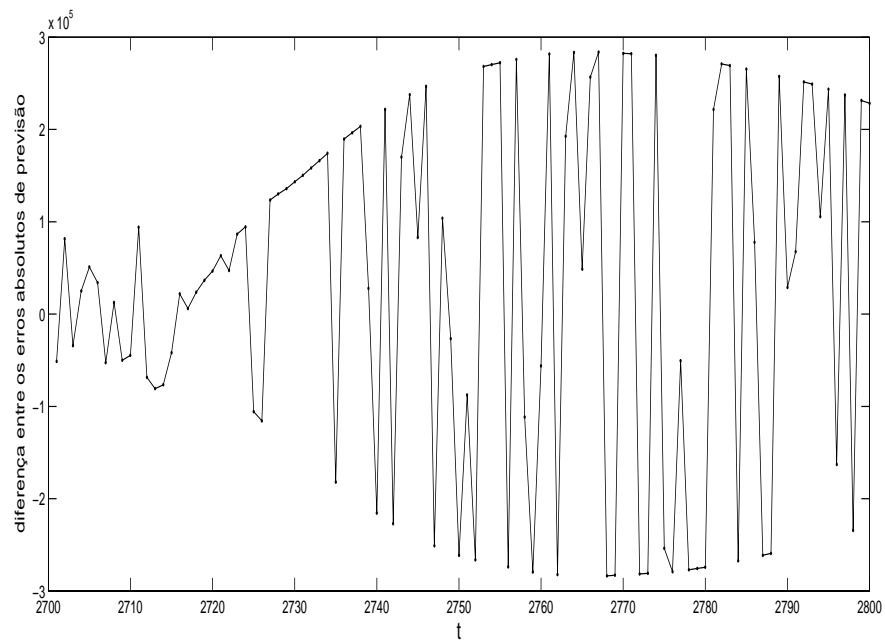


(a)

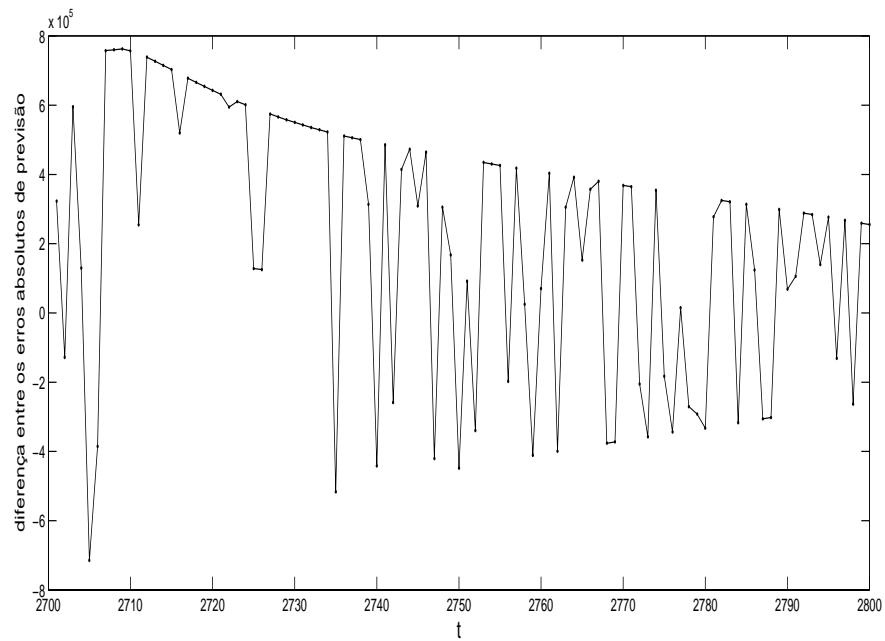


(b)

Figura 6.16: Previsão da série UNC02bin1s com origem em $t = 2700$ ($\hat{d} = 0, 3717$): (a) observações, previsões de passo- h e intervalos de 95% de confiança para as previsões dos modelos TARFIMA(100) e AR(18) e (b) observações, previsões de passo- h e intervalos de 95% de confiança para as previsões dos modelos TARFIMA(100) e AR(2).



(a)



(b)

Figura 6.17: Previsão da série UNC02bin1s com origem em $t = 2700$ ($\hat{d} = 0,3717$): (a) diferença entre os erros absolutos de previsão dos modelos AR(18) e TARFIMA(100) (gráfico $(\Delta_h(1) - \Delta_h(2))$ vs. h , em que $\Delta_h(1)$ é o erro absoluto de previsão do modelo AR(18) e $\Delta_h(2)$ é o erro absoluto de previsão do modelo TARFIMA(100)) e (b) diferença entre os erros absolutos de previsão dos modelos AR(2) e TARFIMA(100) (gráfico $(\Delta_h(1) - \Delta_h(2))$ vs. h , em que $\Delta_h(1)$ é o erro absoluto de previsão do modelo AR(2) e $\Delta_h(2)$ é o erro absoluto de previsão do modelo TARFIMA(100)).

ordem de truncamento L , dada uma “boa” estimativa de H .

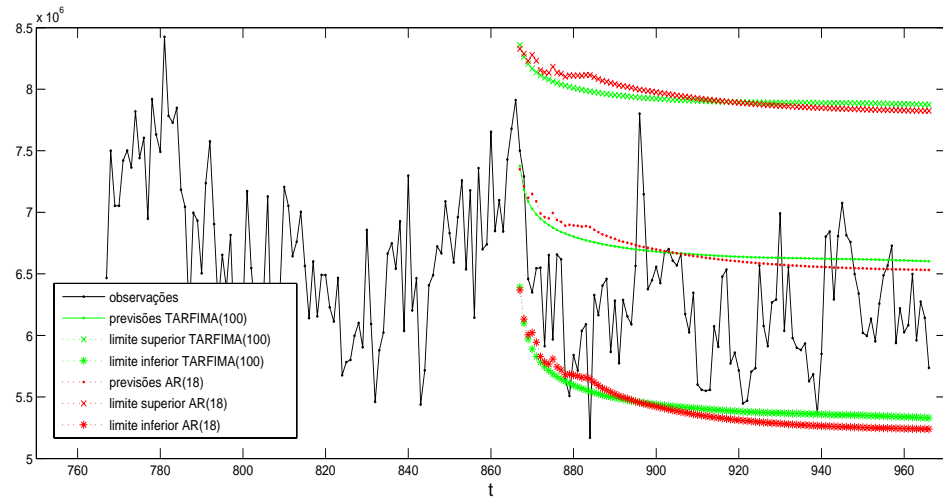
Tabela 6.3: Série `UNC02bin1s` ($\hat{d} = 0,3717$): $PCEMSE_{x,y}(t = 2700, K)$ dos modelos estimados AR(18), TARFIMA(10), TARFIMA(50) e TARFIMA(100) relativos ao modelo AR(2).

K	AR(18)	L=10	L=50	L=100
7	-0,3933	-0,4892	-0,4365	-0,4365
10	-0,6611	-0,5307	-0,6778	-0,6778
20	-0,6885	-0,2885	-0,6858	-0,6912
50	-0,6160	-0,1791	-0,6532	-0,6655
100	-0,5448	-0,1474	-0,5989	-0,6082

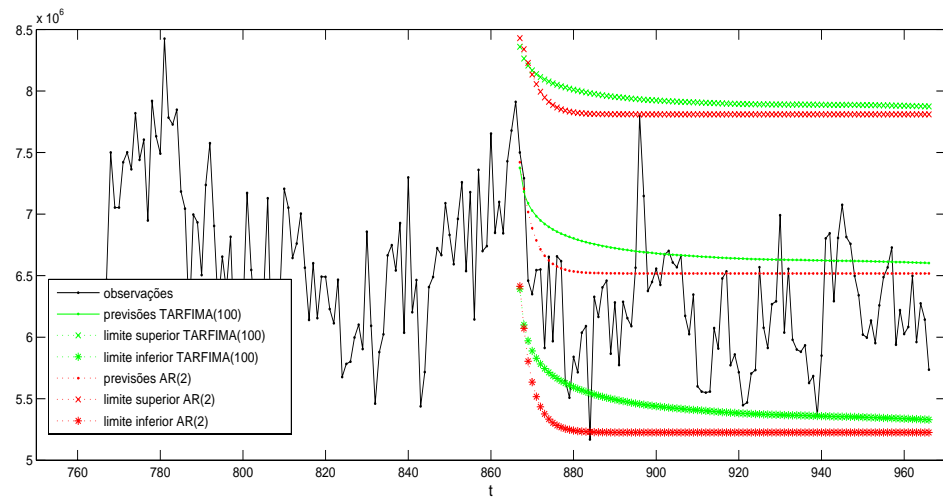
A Fig. 6.18 mostra previsões para a série `UNC02bin1s` com origem em $t = 866$ obtidas por meio dos modelos TARFIMA(100) e AR(18) (vide Fig. 6.18a) e TARFIMA(100) e AR(2) (vide Fig. 6.18b). A Fig. 6.19a mostra que o modelo TARFIMA(100) é superior ao AR(18) para $h < 40$ (a Fig. 6.18a ilustra que as probabilidades de cobertura desses modelos é similar para $t = 866$). Por outro lado, a Fig. 6.19b indica que o modelo AR(2) apresenta um desempenho superior ao do modelo de memória longa TARFIMA(100) (uma vez que o erro absoluto de previsão do modelo AR(2) é quase sempre menor do que o do modelo TARFIMA(100)). Este fato é justificado pela existência de uma inversão “local” de tendência na série `UNC02bin1s` por volta de $t = 866$ (veja que os valores da série tendem a aumentar entre $825 \leq t < 866$ e a decrescer entre $866 \leq t \leq 885$), a qual não é bem modelada pelo modelo TARFIMA(100). Este resultado não surpreende. Conforme visto na seção 5.5.1, modelos SRD podem apresentar, em determinados trechos da série LRD sob análise, um desempenho de previsão superior ao de modelos LRD. Não obstante, modelos de memória longa têm um desempenho superior “na média” quando aplicados às séries LRD.

6.2 Conclusão

O estudo de caso apresentado neste Capítulo mostra que é possível aplicar o modelo de memória longa TARFIMA(L) em trechos estacionários de sinais (Gaussianos/não-Gaussianos) de teletráfego fractal. Além disso, os resultados da seção 6.1.2 sugerem que a dinâmica do parâmetro de Hurst de sinais de teletráfego seja bastante **lenta** na prática (lembre que o parâmetro de Hurst do *trace* `UNC02` mantém-se constante por um período de 2 horas), no sentido de que se o valor de H varia (e ele deve variar um pouco na prática), então espera-se que essa variação ocorra muito lentamente com o tempo (numa escala de horas, por exemplo). Portanto, o novo modelo proposto pode ser usado em esquemas de previsão *on-line*



(a)



(b)

Figura 6.18: Previsão da série UNC02bin1s com origem em $t = 866$ ($\hat{d} = 0,3717$): (a) observações, previsões de passo- h e intervalos de 95% de confiança para as previsões dos modelos TARFIMA(100) e AR(18) e (b) observações, previsões de passo- h e intervalos de 95% de confiança para as previsões dos modelos TARFIMA(100) e AR(2).

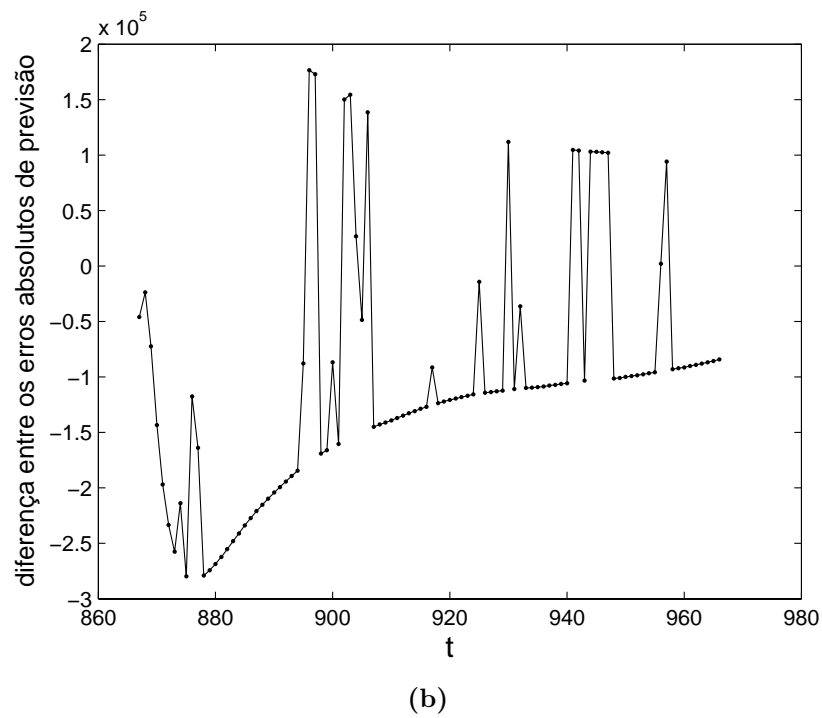
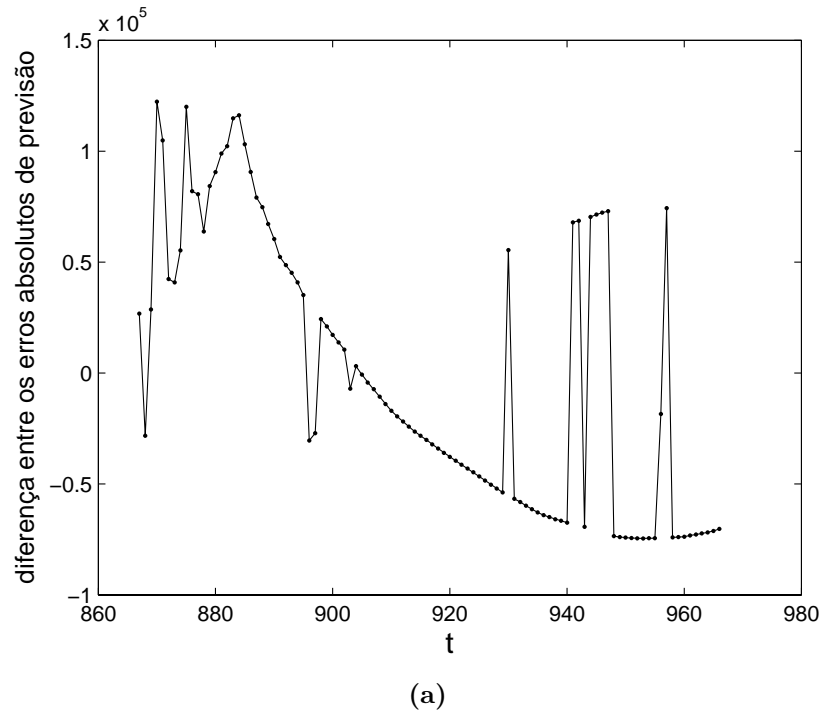


Figura 6.19: Previsão da série `UNC02bin1s` com origem em $t = 866$ ($\hat{d} = 0,3717$): (a) diferença entre os erros absolutos de previsão dos modelos `AR(18)` e `TARFIMA(100)` e (b) diferença entre os erros absolutos de previsão dos modelos `AR(2)` e `TARFIMA(100)`.

de tráfego, tais como CAC e alocação dinâmica de banda, dado que o parâmetro H pode ser estimado em tempo real por meio da aplicação da DWT¹⁰.

¹⁰O *script* MATLAB de Abry-Veitch (estimador *wavelet* de H) levou aproximadamente 13 segundos para estimar o valor do parâmetro de Hurst da série UNC02bin1ms, que possui $N \approx 7,34 \times 10^6$ amostras, num *Personal Computer* (PC) com sistema operacional Windows®XP, processador AMD Athlon de 64 bits e 1 GB de memória RAM. É interessante observar que a mesma máquina “travou” quando o autor desta tese tentou obter estimativas de H para a mesma série utilizando os métodos disponíveis no *software* S+FinMetrics® (análise R/S, Whittle, etc.)!

7 Conclusões e Trabalhos Futuros

Este Capítulo resume as principais conclusões desta tese e também apresenta algumas sugestões para pesquisas futuras.

7.1 Sumário

Esta tese propôs um novo modelo de tráfego com memória longa, para uso em esquemas de previsão Tipo-II, que é construído a partir de uma versão truncada, com ordem L , da representação $AR(\infty)$ (5.1) do processo $ARFIMA(p, d, q)$ (4.74). A representação aproximada $AR(L)$ (5.10) é mapeada para uma representação no espaço de estados conforme a transformação de Harvey [129], dando então origem ao modelo de espaço de estados $TARFIMA(L)$, com as equações do estado e da observação (5.11) e (5.12), respectivamente. O diferencial do $TARFIMA(L)$ é a sua possibilidade de implementação prática, pois não requer memória infinita. A modelagem por meio do processo $TARFIMA(L)$ também oferece outras vantagens como: a) modelagem (explícita) da região das baixas frequências do espectro de sinais LRD e b) viabilização da utilização do filtro de Kalman.

A aplicação de modelos AR de ordem alta em esquemas de previsão *on-line* de teletráfego é fortemente prejudicada pelo problema da identificação da ordem do modelo. Além disso, as simulações do Capítulo 5 mostraram que a modelagem da memória longa pode ser obtida às custas do posicionamento de um ou mais pólos nas proximidades do círculo de raio unitário. Portanto, a implementação do modelo AR ajustado pode ser instável devido a efeitos de quantização dos coeficientes do filtro digital.

O estudo de caso apresentado no Capítulo 6 mostra que é possível aplicar o modelo de memória longa $TARFIMA(L)$ em trechos estacionários de sinais de teletráfego LRD. Os resultados obtidos mostram que a dinâmica do parâmetro H de Hurst de sinais de teletráfego pode ser bastante lenta na prática. Portanto, o

novo modelo proposto é adequado para esquemas de previsão *on-line* de tráfego, tais como CAC e alocação dinâmica de banda, dado que o parâmetro H pode ser estimado em tempo real por meio da aplicação da DWT.

7.2 Trabalhos Futuros

Como o tráfego das redes de comunicações não é estacionário, a detecção e a estimação dos pontos de mudança de regime das séries é uma linha promissora de pesquisa, a qual por sua vez está relacionada à funcionalidade de rastreamento, em tempo real, do modelo de previsão. Conforme alertam Kokoszka e Leipus no artigo [155], o estudo da memória longa em conjunto com a análise do ponto de mudança do modelo gerador da série é viável para sinais de teletráfego porque há um grande número de amostras à disposição das ferramentas estatísticas. Sendo assim, sugere-se que a questão do rastreamento seja abordada em pesquisas futuras. Além disso, note-se que esta é uma área de investigação praticamente inexplorada (vide a revisão bibliográfica efetuada pelo autor desta tese) pelos pesquisadores da área de redes. A tese de doutorado de Brandon Whitcher [156] (“*Assessing Nonstationary Time Series Using Wavelets*”) talvez seja um bom ponto de partida para trabalhos futuros na área de avaliação de teletráfego não-estacionário.

Pode-se classificar os métodos de rastreamento do modelo da mesma maneira que os métodos de previsão, ou seja, em métodos de rastreamento Tipo-I (adequado para modelagem por meio de processos AR de ordem alta) e Tipo-II (adequado para modelos como o TARFIMA).

De acordo com as literaturas das áreas de controle [163], processamento de sinais [52], [124], [164] e séries temporais [42], [146], o rastreamento Tipo-I poderia ser implementado, por exemplo, a partir:

- de um procedimento de estimação de MV que utilize o algoritmo *Expectation-Maximization* (EM) [47] em conjunto com o filtro de Kalman.
- de filtros adaptativos *Recursive Least Squares* (RLS) ou *Least Mean Squares* (LMS).

O rastreamento Tipo-II poderia ser feito por meio da estimação do parâmetro de memória longa H baseada em *wavelets* e da estimação dos parâmetros de um modelo residual de memória curta $\text{ARMA}(p, q)$, que seria obtido por meio da filtragem da memória longa da série original. A identificação dos parâmetros desse

modelo ARMA(p, q) poderia então ser implementada com alguma das técnicas para rastreamento Tipo-I que foram mencionadas anteriormente.

Outra linha de pesquisa bastante interessante (e relevante) é a análise de variância (*Analysis of Variance* (ANOVA)) de sinais de teletráfego baseada em *wavelets*. Conforme explicam Percival e Walden em [67, Cap.8], um uso importante da DWT, e também da sua variante *Maximal Overlap Discrete Wavelet Transform* (MODWT), é a decomposição da variância amostral de um sinal em várias escalas. Percival e Walden [67, Cap.8] mostram como uma quantidade teórica denominada *wavelet variance* (variância *wavelet*) pode ser estimada pela DWT ou MODWT e como estimar a DEP de um sinal LRD a partir da variância *wavelet*. Note-se que o estimador *wavelet* AV faz algo parecido; a diferença é que o método AV está embasado na DWT de tempo contínuo, enquanto que a variância *wavelet* de Percival e Walden está baseada na MRA de tempo discreto.

Por último, mas nem por isso menos importante, destaca-se que a análise estatística da série UNC02 (vide Cap. 6) evidencia a importância da agregação de distribuições estáveis ao gerador de tráfego baseado em *wavelets* desenvolvido por Mello *et al* [77], [78], o qual foi usado para gerar as simulações ARFIMA e MWM do Capítulo 5.

Referências

- [1] LELAND, W. et al. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, v. 2, n. 1, p. 1–15, Feb. 1994.
- [2] PAXSON, V.; FLOYD, S. Wide-area traffic: The failure of Poisson modeling. *IEEE/ACM Transactions on Networking*, v. 3, n. 3, p. 226–244, June 1995.
- [3] FRAME, M.; MANDELBROT, B. B.; NEGER, N. *Natural Fractals, Class on Fractal Geometry offered by Yale University*. 2006. Disponível em: <<http://classes.yale.edu/fractals/>>.
- [4] CAPPE, O. et al. Long-range dependence and heavy-tail modeling for tele-traffic data. *IEEE Signal Processing Magazine*, v. 19, p. 14–27, May 2002.
- [5] INSIGHTFUL. *S-PLUS software*. 2007. Disponível em: <<http://www.insightful.com/products/>>.
- [6] AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In: PETROV, B. N.; CSAKI, F. (Ed.). *2nd International Symposium on Information Theory*. Akademia Kiado, Budapest: [s.n.], 1973. p. 267–281.
- [7] WIKIPEDIA. *Levy Skew alpha-stable Distribution*. Disponível em: <http://en.wikipedia.org/wiki/Levy_skew_alpha-stable_distribution>.
- [8] COEURJOLLY, J.-F. *Inference statistique pour les mouvements browniens fractionnaires et multifractionnaires*. Tese (PhD Thesis) — Université Joseph Fourier, Grenoble I, Institut National Polytechnique de Grenoble, INPG, France, 2000.
- [9] COEURJOLLY, J.-F. *Códigos Matlab para simulação e identificação de fractional Brownian motion e multifractional Brownian motion*. Disponível em: <<http://www-lmc.imag.fr/SMS/software.html>>.
- [10] WAVELAB 850. Disponível em: <http://www-stat.stanford.edu/wavelab/Wavelab_850/index_wavelab850.html>.
- [11] RIEDI, R. H. et al. A multifractal wavelet model with application to network traffic. *IEEE Transactions on Information Theory*, v. 45, n. 3, p. 992–1018, April 1999.
- [12] BARRA, C. R. *Caracterização experimental e por simulação e modelagem da qualidade de serviço obtida na transmissão de áudio e vídeo em tempo real*. Tese (Tese de Doutorado) — Escola Politécnica da USP, São Paulo, 2005.
- [13] DAVIE, B. S.; REKHTER, Y. *MPLS: Technology and Applications*. first. [S.l.]: Morgan Kaufmann, 2000.

- [14] ROSEN, E.; VISWANATHAN, A.; CALLON, R. *Internet RFC 3031: Multiprotocol Label Switching Architecture*. 2001. Disponível em: <<http://www.ietf.org/rfc/rfc3031.txt>>.
- [15] MINEI, I.; LUCEK, J. *MPLS-Enabled Applications: Emerging Developments and New Technologies*. [S.l.]: John Wiley & Sons, 2005.
- [16] BRADEN, R.; CLARK, D.; SHENKER, S. *Internet RFC 1633: Integrated Services in the Internet Architecture*. 1994. Disponível em: <<http://www.ietf.org/rfc/rfc1633.txt>>.
- [17] NICHOLS, K. *Internet RFC 2474: Definition of the Differentiated Services Field (DS field) in the IPv4 and IPv6 Headers*. 1998. Disponível em: <<http://www.ietf.org/rfc/rfc2474.txt>>.
- [18] BLAKE, S. *Internet RFC 2475: An Architecture for Differentiated Services*. 1998. Disponível em: <<http://www.ietf.org/rfc/rfc2475.txt>>.
- [19] ARMITAGE, G. *Quality of Service in IP Networks*. [S.l.]: MacMillan Technical Publishing, 2000.
- [20] WANG, Z. *Internet QoS: Architectures and Mechanisms for Quality of Service*. first. [S.l.]: Morgan Kaufmann, 2001.
- [21] LIMA, A. B. de. *Proposta de uma Estratégia para Controle de Admissão de Conexões Baseado em Medições de Tráfego Agregado e Caracterização de Redes IP*. Tese (Dissertação de Mestrado) — Escola Politécnica da USP, São Paulo, 2002.
- [22] ERLANG, A. K. The theory of probabilities and telephone conversation. *Nyt Tidsskrift for Matematik B (first publication)*, v. 20, 1909.
- [23] GUBNER, J. A. *Probability and Random Processes for Electrical and Computer Engineers*. Cambridge, UK: Cambridge University Press, 2006.
- [24] MANDELBROT, B. B. The variation of certain speculative prices. *J. Business*, v. 36, p. 394–419, 1963.
- [25] BATES, S.; MCLAUGHLIN, S. Testing the gaussian assumption for self-similar teletraffic models. In: *IEEE Signal Processing Workshop on Higher-Order Statistics, 1997*. [S.l.: s.n.], 1997. p. 444–447.
- [26] KARASARIDIS, A.; HATZINAKOS, D. On the modeling of network traffic and fast simulation of rare events using stable self-similar processes. In: *IEEE Signal Processing Workshop on Higher-Order Statistics (HOS)*. [S.l.: s.n.], 1997. p. 268–272.
- [27] GARROPPO, R. G. et al. Testing α -stable processes in modeling broadband teletraffic. In: *IEEE International Conference on Communications, 2000 (ICC 2000)*. [S.l.: s.n.], 2000. v. 3, p. 1615–1619.
- [28] KARASARIDIS, A.; HATZINAKOS, D. Network heavy traffic modeling using α -stable self-similar processes. *IEEE Transactions on Communications*, v. 49, n. 7, p. 1203–1214, July 2001.

- [29] KIM, S.; LEE, J. Y.; SUNG, D. K. A shifted gamma distribution model for long-range dependent internet traffic. *IEEE Communications Letters*, v. 7, n. 3, p. 124–126, March 2003.
- [30] PETROPULU, A. P.; YANG, X. Self-similarity and long-range dependence through the wavelet lens. In: BARNER, K. E.; ARCE, G. R. (Ed.). *Nonlinear Signal and Image Processing - Theory, Methods, and Applications*. [S.l.]: CRC Press, 2004.
- [31] KLEINROCK, L. *Queueing Systems*. [S.l.]: John Wiley & Sons, 1975.
- [32] ERRAMILI, A.; NARAYAN, O.; WILLINGER, W. Experimental queuing analysis with long-range dependent traffic. *IEEE/ACM Transactions on Networking*, v. 4, p. 209–223, April 1996.
- [33] YANG, X.; PETROPULU, A. P. The extended alternating fractal renewal process for modeling traffic in high-speed communications networks. *IEEE Transactions on Signal Processing*, Springer Netherlands, v. 49, n. 7, p. 1349–1363, July 2001.
- [34] GROSSGLAUSER, M.; BOLOT, J.-C. On the relevance of long-range dependence in network traffic. *IEEE/ACM Transactions on Networking*, v. 7, n. 5, p. 629–640, October 1999.
- [35] WILLINGER, W. et al. On traffic measurements that defy traffic models (and vice-versa): Self-similar traffic modeling for high-speed networks. *ConneXions*, v. 8, n. 11, p. 14–24, 1994.
- [36] QIU, J.; KNIGHTLY, E. W. Measurement-based admission control with aggregate traffic envelopes. *IEEE/ACM Transactions on Networking*, v. 9, n. 2, p. 199–210, April 2001. Disponível em: <<http://citeseer.nj.nec.com/qiu01measurementbased.htm>>.
- [37] BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. *Time Series Analysis: Forecasting and Control*. 3rd. ed. [S.l.]: Prentice Hall, 1994.
- [38] BOX, G. E. P.; JENKINS, G. M. *Time Series Analysis, Forecasting and Control*. Revised. San Francisco: Holden Day, 1976.
- [39] TSAY, R. S. *Analysis of Financial Time Series*. second. Hoboken, New Jersey: John Wiley and Sons, 2005.
- [40] BROCKWELL, P. J.; DAVIS, R. A. *Time series: theory and methods*. 2^a. ed. New York: Springer-Verlag, 1991.
- [41] BROCKWELL, P. J.; DAVIS, R. A. *Introduction to Time Series and Forecasting*. New York: Springer-Verlag, 1996.
- [42] MORETTIN, P. A.; TOLOI, C. M. C. *Análise de Séries Temporais*. São Paulo, SP: Edgard Blücher Ltda., 2004.
- [43] ZIVOT, E.; WANG, J. *Modeling Financial Time Series with S-PLUS*. [S.l.]: Springer, 2003.
- [44] PERCIVAL, D. B.; WALDEN, A. T. *Spectral Analysis for Physical Applications*. New York: Cambridge, 1993.

- [45] OPPENHEIM, A. V.; SCHAFER, R. W. *Discrete-Time Signal Processing*. N.J., U.S.A.: Prentice-Hall International, Inc., 1989.
- [46] PAPOULIS, A. *Probability, Random Variables, and Stochastic Processes*. Third. [S.l.]: McGraw-Hill, 1996.
- [47] STARK, H.; WOODS, J. W. *Probability and Random Processes with Applications to Signal Processing*. 3rd. ed. Upper Saddle River, NY: Prentice Hall, 2002.
- [48] MORETTIN, P. A. *Econometria Financeira*. [S.l.]: Associação Brasileira de Estatística, 2006.
- [49] BACCALÁ, L. A. *Notas de Aula da Disciplina PTC5779, Processamento de Sinais 3*. [S.l.]: EPUSP, 2003.
- [50] PAXSON, V. Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic. *Computer Communication review*, v. 27, p. 5–18, Oct. 1997.
- [51] SAMORODNITSKY, G.; TAQQU, M. S. *Stable non-Gaussian random processes*. London, UK: Chapman & Hall, 1994.
- [52] PROAKIS, J. G.; MANOLAKIS, D. G. *Digital Signal Processing*. Fourth. [S.l.]: Pearson Prentice Hall, 2007.
- [53] FELLER, W. *An Introduction to Probability Theory and Its Applications, Vol. 2*. 2nd. ed. New York: John Wiley, 1971.
- [54] NOLAN, J. *An Introduction to Stable Distributions*. Disponível em: <<http://academic2.american.edu/~jpnolan/stable/stable.html>>.
- [55] SHAO, M.; NIKIAS, C. Signal processing with fractional lower order moments: Stable processes and their applications. *Proceedings of the IEEE*, v. 81, n. 7, July 1993.
- [56] AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19, p. 716–723, 1974.
- [57] MORETTIN, P. A. *Econometria Financeira: Um curso em Séries Temporais Financeiras*. [s.n.], 2003. Disponível em: <<http://www.ime.usp.br/~pam>>.
- [58] BERAN, J. *Statistics for Long-Memory Processes*. [S.l.]: Chapman & Hall, 1994.
- [59] MANDELBROT, B. B. *The Fractal Geometry of Nature*. New York: WH Freeman, 1977.
- [60] FALCONER, K. *Fractal geometry, Mathematical Foundations and Applications*. England: John Wiley & Sons, Inc., 1990.
- [61] FELDMANN, A.; GILBERT, A. C.; WILLINGER, W. Data networks as cascades: investigating the multifractal nature of internet WAN traffic. *Computer Communication review*, v. 28, n. 4, p. 42–55, 1998.

- [62] PARK, K.; WILLINGER, W. Self-similar network traffic: An overview. In: PARK, K.; WILLINGER, W. (Ed.). *Self-Similar Network Traffic and Performance Evaluation*. [S.l.]: John Wiley & Sons, Inc., 2000.
- [63] ABRY, P. et al. Wavelets for the analysis, estimation, and synthesis of scaling data. In: PARK, K.; WILLINGER, W. (Ed.). *Self-Similar Network Traffic and Performance Evaluation*. [S.l.]: John Wiley & Sons, Inc., 2000.
- [64] RIEDI, R. Multifractal processes. In: DOUKAN, P.; OPPENHEIM, G.; TAQQU, M. S. (Ed.). *Theory and Applications of Long Range Dependence*. Boston, MA: Birkhäuser, 2003.
- [65] SPROTT, J. C. *Chaos and Time-Series Analysis*. New York: Oxford University Press, 2003.
- [66] KLIMKE, A. *Mandelbrot Set GUI (for MATLAB)*. 2003. Disponível em: <<http://matlabdb.mathematik.uni-stuttgart.de/index.jsp>>.
- [67] PERCIVAL, D. B.; WALDEN, A. T. *Wavelet Methods for Time Series Analysis*. [S.l.]: Cambridge University Press, 2000.
- [68] CANTOR, G. Über unendliche, lineare punktmannigfaltigkeiten v. *Mathematische Annalen*, v. 21, p. 545–591, 1883.
- [69] PEITGEN, H.-O.; JÜRGENS, H.; SAUPE, D. *Chaos and Fractals: New Frontiers of Science*. [S.l.]: Springer-Verlag, 1992.
- [70] PARK, C. et al. Long-range dependence in a changing internet traffic mix. *Computer Networks*, v. 48, p. 401–422, 2005.
- [71] HURST, H. E. Long-term storage capacity of reservoirs. *Trans. Am. Soc. Civil Engineers*, v. 116, p. 770–799, 1951.
- [72] RIEDI, R.; VÉHEL, J. L. *Multifractal properties of TCP traffic: a numerical study*, *Tech. Rep. 3129*. [S.l.], 1997. Disponível em: <<http://citeseer.ist.psu.edu/riedi97multifractal.html>>.
- [73] GILBERT, A. C.; WILLINGER, W.; FELDMANN, A. Scaling analysis of conservative cascades, with applications to network traffic. *IEEE Transactions on Information Theory*, v. 45, n. 3, p. 971–991, April 1999. Special Issue on “Multiscale Statistical Signal Analysis and its Applications”.
- [74] RIBEIRO, V. et al. Simulation of nonGaussian long-range dependent traffic using wavelets. In: *SIGMETRICS'99*. [S.l.: s.n.], 1999.
- [75] GRANGER, C. W. J. The typical spectral shape of an economic variable. *Econometrica*, v. 34, p. 150–161, 1966.
- [76] MANDELBROT, B. B. When can price be arbitrated efficiently? a limit to the validity of the random walk and martingale models. *Review of Economics and Statistics*, v. 56, p. 225–236, 1971.
- [77] MELLO, F. L. de et al. Geração de séries auto-similares Gaussianas via wavelets para uso em simulações de tráfego. *IEEE Latin America Transactions*, v. 5, n. 1, Março 2007.

- [78] LIMA, A. B. de et al. A generator of teletraffic with long and short-range dependence. In: *12th Computer Aided Modeling and Design of Communication Links and Networks (CAMAD07) Workshop, part of the 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC07)*. Athens, Greece: [s.n.], 2007.
- [79] DUFFIELD, N. G. Economies of scale for long-range dependent traffic in short buffers. *Telecommunication Systems*, Springer Netherlands, v. 7, n. 1-3, p. 267–280, March 1997.
- [80] MA, S.; JI, C. Modeling heterogeneous network traffic in wavelet domain. *IEEE Transactions on Networking*, v. 9, n. 5, p. 634–649, Oct. 2001.
- [81] GRANGER, C. W. J.; JOYEUX, R. An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, v. 1, p. 15–29, Oct. 1980.
- [82] HOSKING, J. R. M. Fractional differencing. *Biometrika*, v. 68, p. 165–176, Oct. 1981.
- [83] JAIN, R. *The Art of Computer Systems Performance Analysis - Techniques for Experimental Design, Simulation, and Modeling*. [S.l.]: John Wiley & Sons, Inc., 1991.
- [84] TAQQU, M. S. Fractional brownian motion and long-range dependence. In: DOUKAN, P.; OPPENHEIM, G.; TAQQU, M. S. (Ed.). *Theory and Applications of Long Range Dependence*. Boston, MA: Birkhäuser, 2003.
- [85] MANDELBROT, B. B.; NESS, J. V. Fractional brownian motions, fractional noises and applications. *SIAM Rev.*, v. 10, p. 422–437, Feb. 1968.
- [86] BÄCKAR, J.-A. *A Framework for Implementing Fractal Traffic Models in Real Time*. Tese (Master Thesis) — SERC, Melbourne, 2000.
- [87] TSYBAKOV, B.; GEORGANAS, N. D. On self-similar traffic in ATM queues: definitions, overflow probability bound, and cell delay distribution. *IEEE Trans. Networking*, v. 5, p. 397–409, June 1997.
- [88] PILLAI, S. S.; HARISANKAR, M. Simulated performance of a DS spread spectrum system in impulsive atmospheric noise. *IEEE Trans. Electromag. Compat.*, EMC-29, p. 80–82, 1987.
- [89] BOUVET, M.; SCHWARTZ, S. C. Comparison of adaptive and robust receivers for signal detection in ambient underwater noise. *IEEE Trans. Acoust., Speech, Signal Processing*, v. 37, p. 621–626, 1989.
- [90] MIDDLETON, D. Statistical physical models of electromagnetic interference. *IEEE Trans. Electromag. Compat.*, EMC-19, p. 106–127, 1977.
- [91] PIERCE, R. D. Application of the positive alpha-stable distribution. In: *IEEE Signal Processing Workshop on Higher-Order Statistics*. Alberta, Canada: [s.n.], 1997. p. 420–424.
- [92] BERGER, J.; MANDELBROT, B. B. A new model for error clustering in telephone circuits. *IBM J. Res. Develop.*, p. 224–236, 1963.

- [93] LÉVY, P. *Calcul des Probabilités*. Paris: Gauthier-Villars, 1925.
- [94] WIKIPEDIA. *Pareto Distribution*. Disponível em: http://en.wikipedia.org/wiki/Pareto_distribution.
- [95] NOLAN, J. *S-TABLE Library (MATLAB version)*. Disponível em: <http://www.robustanalysis.com>.
- [96] PETROPULU, A. P. et al. Power-law shot noise and relationship to long-memory processes. *IEEE Trans. Signal Processing*, v. 48, n. 7, July 1989.
- [97] CROVELLA, M. E.; BESTAVROS, A. Self-similarity in world wide web traffic - evidence and possible causes. In: *Proceedings of the ACM Sigmetrics'96*. [S.l.: s.n.], 1996. p. 160–169.
- [98] WILLINGER, W. et al. Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level. *IEEE/ACM Trans. on Networking*, v. 5, p. 71–86, 1997.
- [99] WILLINGER, W. et al. Scaling phenomena in the internet: Critically examining critically. In: *National Academy of Sciences, Colloquium*. [s.n.], 2002. v. 99, p. 2573–2580. Disponível em: <http://www.pnas.org/cgi/doi/10.1073/pnas.012583099>.
- [100] GONG, W.-B. et al. Self-similarity and long range dependence on the internet: a second look at the evidence, origins and implications. *Computer Networks*, v. 48, p. 377–399, 2005.
- [101] BARDET, J.-M. et al. Generators of long-range dependent processes: a survey. In: DOUKAN, P.; OPPENHEIM, G.; TAQQU, M. S. (Ed.). *Theory and Applications of Long Range Dependence*. Boston, MA: Birkhäuser, 2003.
- [102] DAVIES, R. B.; HARTE, D. S. Tests for Hurst effect. *Biometrika*, v. 74, p. 95–101, 1987.
- [103] PERCIVAL, D. B. Simulating Gaussian random processes with specified spectra. *Computing Science and Statistics*, v. 24, p. 534–538, 1992.
- [104] BERAN, J. et al. Long-range dependence in variable-bit-rate video traffic. *IEEE Trans. Commun.*, v. 43, n. 234, p. 1566–1579, Feb./March/Apr. 1995.
- [105] CHANG, P.-R.; HU, J.-T. Optimal nonlinear adaptive prediction and modeling of MPEG video in ATM networks using pipelined recurrent neural networks. *IEEE J. Select. Areas Commun.*, v. 15, n. 8, p. 1087–1100, 1997.
- [106] LIMA, A. B. de; AMAZONAS, J. R. de A. State-space modeling of long-range dependent teletraffic. In: *20th International Teletraffic Congress (ITC 2007)*. Ottawa, Canada: [s.n.], 2007. p. 260–271.
- [107] ANDERSEN, A. T.; NIELSEN, B. F. An application of superpositions of two state Markovian sources to the modeling of self-similar behavior. In: *Sixteenth Annual Joint Conf. of the IEEE Comp. and Comm. Societies, INFOCOM 1997*. [S.l.: s.n.], 1997. p. 196–204.
- [108] GABOR, D. Theory of communication. *J. Inst. Electr. Eng.*, v. 93, n. III, p. 429–457, 1946.

- [109] KAISER, G. *A Friendly Guide to Wavelets*. Boston, Mass.: Birkhäuser, 1994.
- [110] GENÇAY, R.; SELÇUK, F.; WHITCHER, B. *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. [S.l.]: Academic Press, 2001.
- [111] BACCALÁ, L. A. *Notas de Aula da Disciplina PTC5772, Processamento de Sinais no Domínio Tempo-Frequência*. [S.l.]: EPUSP, 2004.
- [112] GROSSMANN, A.; MORLET, J. Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM J. Math.*, v. 15, p. 723–736, 1984.
- [113] MALLAT, S. G. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 11, p. 674–693, 1989.
- [114] MALLAT, S. G. Multiresolution approximations and wavelet orthonormal bases of $l^2(\mathbb{R})$. *Transactions of the American Mathematical Society*, v. 315, p. 69–87, 1989.
- [115] MALLAT, S. G. Multifrequency channel decompositions of images and wavelet models. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 37, p. 2091–2110, 1989.
- [116] DAUBECHIES, I. Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, v. 41, p. 909–996, 1988.
- [117] DAUBECHIES, I. *Ten Lectures on Wavelets*. Philadelphia: SIAM, 1992.
- [118] STEIN, E. M.; SHAKARCHI, R. *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*. [S.l.]: Princeton University Press, 2005.
- [119] KOLMOGOROV, A. N.; FOMIN, S. V. *Introductory Real Analysis (revised English Edition translated and edited by R. A. Silverman)*. [S.l.]: Dover Publications, Inc., 1975.
- [120] VEITCH, D.; TAQQU, M. S.; ABRY, P. Meaningful MRA initialization for discrete time series. *Signal Processing*, v. 80, p. 1971–1983, 2000.
- [121] MALLAT, S. *A Wavelet Tour of Signal Processing*. second. [S.l.]: Academic Press, 1999.
- [122] ABRY, P.; VEITCH, D. Wavelet analysis of long-range dependent traffic. *IEEE Transactions on Information Theory*, v. 4, n. 1, p. 2–15, 1998.
- [123] CROCHIERE, R. E.; RABINER, L. R. *Multirate Digital Signal Processing*. [S.l.]: Prentice Hall, 1983.
- [124] SAYED, A. H. *Fundamentals of Adaptive Filtering*. Hoboken, NJ: John Wiley & Sons, Inc., 2003.
- [125] LO, A. W. Long term memory in stock market prices. *Econometrica*, v. 59, n. 5, p. 1279–1313, 1991.

- [126] HAMILTON, J. D. *Time Series Analysis*. New Jersey, U.S.A.: Princeton University Press, 1994.
- [127] GEWEKE, J.; PORTER-HUDAK, S. The estimation and application of long memory time series models. *Journal of Time Series Analysis*, v. 4, p. 221–237, 1983.
- [128] TAQQU, M.; TEVEROVSKY, V.; WILLINGER, W. Estimators for long-range dependence: An empirical study. *Fractals*, v. 3, p. 785–798, 1995.
- [129] HARVEY, A. C. *Time Series Models*. second. [S.l.]: MIT Press, 1993.
- [130] HASLETT, J.; RAFTERY, A. E. Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. *Journal of Royal Statistical Society Series C*, v. 38, p. 1–21, 1989.
- [131] STOEV, S. et al. On the wavelet spectrum diagnostic for Hurst parameter estimation in the analysis of internet traffic. *Computer Networks*, v. 48, p. 423–444, 2005.
- [132] BARDET, J.-M. et al. Semi-parametric estimation of the long-range dependence parameter. In: DOUKAN, P.; OPPENHEIM, G.; TAQQU, M. S. (Ed.). *Theory and Applications of Long Range Dependence*. Boston, MA: Birkhäuser, 2003.
- [133] VEITCH, D.; ABRY, P. *Second Order Estimation Code*. Disponível em: <<http://www.cubinlab.ee.mu.oz.au/~darryl/>>.
- [134] BRILLINGER, D. R. An introduction to polyspectra. *Ann. Math. Statist.*, v. 36, p. 1351–1374, 1965.
- [135] ROSENBLATT, M. *Handbook of Statistics, 3, D. R. Brillinger and P. R. Krishnaiah (Eds.)*. [S.l.]: Elsevier Science Publishers B. V., 1983. 369–382 p.
- [136] NIKIAS, C. L.; PETROPULU, A. P. *Higher-order spectra analysis: a non-linear signal processing framework*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [137] HINICH, M. J. Testing for gaussianity and linearity of a stationary time series. *J. Time Series Analysis*, v. 3, p. 169–176, 1982.
- [138] SWAMI, A.; MENDEL, J. M.; NIKIAS, C. L. *Higher-Order Spectral Analysis Toolbox User's Guide, V. 2*. [S.l.]: United Signals & Systems, Inc., 2000.
- [139] KWIATKOWSKI, D. et al. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, v. 54, p. 159–178, 1992.
- [140] SHU, Y. et al. Traffic prediction using FARIMA models. In: *International Conference on Communications 1999 (ICC'99)*. [S.l.: s.n.], 1999. v. 2, p. 891–895.
- [141] SHU, Y. et al. Prediction-based admission control using FARIMA models. In: *International Conference on Communications 2000 (ICC'00)*. [S.l.: s.n.], 2000. v. 2, p. 1325–1329.

- [142] ILOW, J. Forecasting network traffic using FARIMA models with heavy tailed innovations. In: *International Conference on Acoustic, Speech and Signal Processing 2000* (ICASSP'00). [S.l.: s.n.], 2000. v. 6, p. 3814–3817.
- [143] CORRADI, M. et al. Analysis of FARIMA processes in the modeling of broadband traffic. In: *International Conference on Communications 2000* (ICC'01). [S.l.: s.n.], 2001. v. 3, p. 964–968.
- [144] SADEK, N.; KHOTANZAD, A.; CHEN, T. ATM dynamic bandwidth allocation using F-ARIMA prediction model. In: *12th International Conference on Computer Communications and Networks (ICCCN) 2003*. [S.l.: s.n.], 2003. p. 359–363.
- [145] KALMAN, R. E. A new approach to linear filtering and prediction problems. *Trans. ASME, J. Basic Eng.*, v. 82, p. 35–45, 1960.
- [146] DURBIN, J.; KOOPMAN, S. J. *Time Series Analysis by State Space Models*. [S.l.]: Oxford University Press, 2001.
- [147] BHANSALI, R. J.; KOKOSZKA, P. S. Prediction of long-memory time series. In: DOUKAN, P.; OPPENHEIM, G.; TAQQU, M. S. (Ed.). *Theory and Applications of Long Range Dependence*. Boston, MA: Birkhäuser, 2003.
- [148] KOLAROV, A.; ATAI, A.; HUI, J. Application of Kalman filter in high-speed networks. In: *Global Telecommunications Conference 1994* (GLOBECOM'94). [S.l.: s.n.], 1994. v. 1, p. 624–628.
- [149] LIM, A. O.; AB-HAMID, K. Kalman predictor method for congestion avoidance in ATM networks. In: *TENCON 2000*. [S.l.: s.n.], 2000. p. I346–I351.
- [150] ZHINJUN, F.; YUANHUA, Z.; DAOWEN, Z. Kalman optimized model for MPEG-4 VBR sources. *IEEE Transactions on Consumer Electronics*, v. 50, n. 2, p. 688–690, May 2004.
- [151] STOICA, P.; MOSES, R. *Spectral Analysis of Signals*. [S.l.]: Pearson Prentice Hall, 2005.
- [152] KAY, S. M.; JR., S. L. M. Spectrum analysis - a modern perspective. In: *Proceedings of the IEEE*. [S.l.: s.n.]. v. 69, p. 1380–1419.
- [153] PARK, C. et al. *Trace UNC02*. Disponível em: <<http://www-dirt.cs.unc.edu/ts/>>.
- [154] JACOBSON, V.; LERES, C.; MCCANNE, S. *TCPDUMP Public Repository*. Disponível em: <<http://www.tcpdump.org/>>.
- [155] KOKOSZKA, P.; LEIPUS, R. Detection and estimation of changes in regime. In: DOUKAN, P.; OPPENHEIM, G.; TAQQU, M. S. (Ed.). *Theory and Applications of Long Range Dependence*. [S.l.: s.n.]. p. 325–337.
- [156] WHITCHER, B. J. *Assessing Nonstationary Time Series Using Wavelets*. Tese (Doctoral Thesis) — University of Washington, 1998.
- [157] JARQUE, C. M.; BERA, A. K. A test for normality of observations and regression residuals. *International Statistical Review*, v. 55, n. 5, p. 163–172, 1987.

-
- [158] NOLAN, J. P. *User Manual for STABLE 4.0 Matlab version*. [S.l.].
- [159] NOLAN, J. P. Fitting data and assessing goodness-of-fit with stable-distributions. In: *Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics*. Washington, DC, USA: [s.n.], 1999.
- [160] KUROSE, J. F.; ROSS, K. W. *Redes de Computadores e a Internet*. [S.l.]: Pearson Addison Wesley, 2006.
- [161] ZHANG, Z. et al. Modeling video traffic in the wavelet domain. In: *22nd Annual Joint Conf. of the IEEE Comp. and Comm. Societies, INFOCOM 2003*. San Francisco, USA: [s.n.], 2003. p. 1826–1836.
- [162] STOEV, S. et al. LASS: a tool for the local analysis of self-similarity. *Computational Statistics & Data Analysis*, v. 50, p. 2447–2471, 2006.
- [163] GRAUPE, D. *Time series analysis: identification and adaptive filtering*. [S.l.]: Robert E. Krieger Publishing Co., Inc., 1989.
- [164] KAILATH, T.; SAYED, A. H.; HASSIBI, B. *Linear Estimation*. Upper Saddle River, NY: Prentice Hall, 2000.

Apêndice A – Códigos MATLAB

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Alexandre B. de Lima
%
% 04/11/2007
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% function emse()
%
% Given a time series FARIMA(p,d,q), an estimate of the long memory
% parameter d (= H-1/2), and an adjusted AR(p2), this function returns:
% 1) "k-steps-ahead-forecasts" based on the last 100 observations
%    (k=1,...,k_max=100) given by the AR(p2) and the TARFIMA(L) models.
% 2) an Excel Table with the Empirical Mean-Square Errors (EMSE) for
%    several values of k_max (k_max= 7,10,20,50,100)
%
% Note that TARFIMA(L) is a finite-dimensional (i.e.,truncated)
% state-space representation of an FARIMA(p,d,0) model. For more details,
% please refer to the paper
% "State-Space Modeling of Long-Range Dependent Teletraffic", ITC 2007,
% LNCS 4516, pp.260-271, 2007 (20th Int'l Teletraffic Congress, Ottawa,
% Canada).
%
% inputs:
%      timeseries = name of the text file which contains the time
%                  series under analysis (ASCII column vector)
%      d          = estimated value of "d" for "timeseries" (I use
%                  the estimate given by S+FinMetrics)
%      name_ar_farima = name of a text file with the estimated AR coef.
%                  of FARIMA (if applicable)
%      name_ma_farima = name of a text file with the estimated MA coef.
%                  of FARIMA (if applicable)
%      L          = order of the finite-dimensional representation
%      namear      = name of a text file which contains the parameters
%                  of the AR fitted model
%
% outputs:
%      Table with EMSEs, Figures with forecasts
%
% Copyright(c) 2007 Universidade de São Paulo, Laboratório de Comunicações
% e Sinais
%
% Permission is granted for use and non-profit distribution providing that
% this notice be clearly maintained. The right to distribute any portion
% for profit or as part of any commercial product is specifically reserved

```

```
% for the author.

%function emse()

clear
close all

% note that : a) vectors with coefficients are column vectors, and
%              b) vetors with observations are row vectors

% =====
% INPUT DATA

% Prompt for time series
timeseries = input('file with the time series: ','s');

% Prompt for estimated FARIMA model
d = input('parameter "d": ');

yes = 'y';
yes2 = 'Y';
j1 = input('AR coefficients? Y/N [Y]: ','s');
if ( strcmp(yes,j1) | strcmp(yes2,j1) )
    name_ar_farima = input('file with the estimated AR coef. of FARIMA:
        ','s');
    ar_farima = load(name_ar_farima);
    ar_farima = -1*ar_farima;
% remember that the coefficients given by S-PLUS are the
% negative of the coefficients used in MATLAB (convention)
% See S-PLUS Guide to Statistics, Vol. 2
    ar_farima = [1; ar_farima]; % MATLAB convention
else
    ar_farima = [1];
end;

j2=input('MA coefficients? Y/N [Y]: ','s');
if ( strcmp(yes,j2) | strcmp(yes2,j2) )
    name_ma_farima = input('file with the estimated MA coef. of FARIMA:
        ','s');
    ma_farima = load(name_ma_farima);
    ma_farima = -1*ma_farima;
% remember that the coefficients given by S-PLUS are the
% negative of the coefficients used in MATLAB (convention)
    ma_farima = [1; ma_farima]; % MATLAB convention
else
    ma_farima = [1];
end;

L = input('order of the finite-dimensional representation: ');
namear = input('file with the coef. of the estimated AR model: ','s');

fullseries =load(timeseries)';
N = length(fullseries);
%mu = mean(fullseries);

% =====
% Prediction from T-FARIMA(L) model
```

```

for j = 0:100
    D(j+1) = -gamma(j-d)/(gamma(j+1)*gamma(-d));
end

D = -1*D; % D(B) = (1-B)^d = 1 -d_1B -d_2B^2 - ...
% "D" is a row vector

phi_temp = conv(ar_farima,D');

% Truncated AR(L) polynomial (1-phi_1B-phi_2B^2 - ... - phi_LB^L)
phi_B = ldiv(phi_temp,ma_farima,(L+1));
phi_B2 = -1*phi_B(2:(L+1)); %coef. of AR(L) (S+ convention)
                                %column vector
[m,n]=size(phi_B2);

if (m<n)
    phi_B2 = phi_B2';
end

k = [7 10 20 50 100];
% k denotes the number of steps to predict ahead
% k_max=100

% ma100 = ldiv(ma_farima,phi_temp,101); % ldiv(b,a,N) is a function for
                                % the power series expansion of
                                % polynomial phi_B

% Now we must invert and truncate the AR(L) representation
% Yields the same result of: ma100 = ldiv(ma_farima,phi_temp,101)
ma100 = ldiv(1,phi_B,101);
ma100 = ma100(2:length(ma100));
% Vector with the coef. of the MA(100) representation of T-FARIMA(L)
% [psi1 psi2 ... psi100], calculated by S-PLUS
% do not change signs! (as stated in section "DETAILS" of the "arma2ma()")
%                                function help )

%prev = zeros(length(fullseries)); % initialize row vector with predictions

ma100quad = ma100.^2; % coefficiente to the square
var_erro = zeros(1,k(1)); % initialize row vector with the variances
                                % of the k-step ahead prediction errors

%t = 2700;
%t=866;
% origin of prediction

for j=1:length(k)

    t=length(fullseries)-k(j);
    % origin of prediction

    % estimate mean of time series
    mu = mean(fullseries(t:-1:1));

    % Initialize row vector Z (with L previous observations)
    Z = fullseries(t:-1:t-(length(phi_B2)-1)) - mu;
    % Z = [Z_358 Z_357 ... Z_309]

```

```

% column vector I of innovations
I = filter(phi_B,1,(fullseries'-mu));
pot_inov = var(I);

for h=1:k(j)
    prev(h) = Z*phi_B2;
    Z = [prev(h) Z(1:(length(phi_B2)-1))];
end

prev = prev + mu; % predictions
x = fullseries(t+1:t+k(j)); % future observations
EQMPt(j) = sum((prev-x).^2)/k(j);
erro = x-prev;

if j == length(k)
    for h=1:k(j)
        if h==1
            var_erro(h) = 1*pot_inov; % vide Eq.(9.11) Morettin(2004)
        else
            var_erro(h) = (ma100quad(h-1)*pot_inov) + var_erro(h-1);
        end
    end
end
end

upper_bound = prev + 1.96*sqrt(var_erro);
lower_bound = prev - 1.96*sqrt(var_erro);

erroabs = abs(erro);

% =====

% Prediction from AR(p) model

ar = load(namear);
p = length(ar);
ar = -1*ar; % remember that the coefficients given by S-PLUS are the
           % negative of the coefficients used in MATLAB (convention)

ar = [1; ar]; % MATLAB convention
ARorder = strcat('AR(',num2str(p),')')
B = [1];
[z_ar,p_ar,G] = tf2zpk(B,ar);
zplane(z_ar,p_ar,'k')
title(['Diagrama de Pólos e Zeros do modelo AR(',int2str(p),')'])
figure

ar2 = -1*ar(2:length(ar)); % S-PLUS convention

ar_ma100 = ldiv(1,ar,101);
ar_ma100 = ar_ma100(2:length(ar_ma100))';
% Vector with the coef. of the MA(100) representation of AR(p)
% [psi1 psi2 ... psi100], calculated by S-PLUS
% do not change signs! (as stated in section "DETAILS" of the "arma2ma()")
%
function help )

```

```
%prev2 = zeros(length(fullseries));
% initialize row vector with predictions
ar_ma100quad = ar_ma100.^2; % coef. to the square
var_erro2 = zeros(1,k(1));
% initialize row vector with the variances
% of the k-step ahead prediction errors

for j=1:length(k)

    t=length(fullseries)-k(j); % origin of prediction

    % estimate mean of time series
    mu = mean(fullseries(t:-1:1));

    % Initialize row vector Z2 (p observations)
    Z2 = fullseries(t:-1:t-(length(ar2)-1))- mu;

    % column vector I2 of innovations
    I2 = filter(ar,1,(fullseries'-mu));
    pot_inov2 = var(I2);

    for h=1:1:k(j)
        prev2(h) = Z2*ar2;
        Z2 = [prev2(h) Z2(1:(length(ar2)-1))];
    end

    prev2 = prev2 + mu;
    x2 = fullseries(t+1:t+k(j));
    EQMPt2(j) = sum((prev2-x2).^2)/k(j);
    erro2 = x2-prev2;

    if j == length(k)
        for h=1:k(j)
            if h==1
                var_erro2(h) = 1*pot_inov2;
                % vide Eq.(9.11) Morettin(2004)
            else
                var_erro2(h) = (ar_ma100quad(h-1)*pot_inov2) +
                    var_erro2(h-1);
            end
        end
    end
end

upper_bound2 = prev2 + 1.96*sqrt(var_erro2);
lower_bound2 = prev2 - 1.96*sqrt(var_erro2);

erro2abs = abs(erro2);

delta_error_abs = erro2abs - erroabs ;
% =====
%ARorder = strcat('AR(',num2str(p),')')
TFARIMAorder = strcat('TARFIMA(',num2str(L),')')

EQM = [EQMPt2',EQMPt'];
```

```

emse1 = {'k',ARorder,TFARIMAorder; k(1),EQM(1,1),EQM(1,2); k(2),EQM(2,1),
EQM(2,2); k(3),EQM(3,1),EQM(3,2);k(4),EQM(4,1),EQM(4,2); k(5),EQM(5,1),
EQM(5,2)}

xlswrite('emse.xls', emse1, 'teste');

%=====
% Draw graphics

% Estimation of Power spectral density (PSD) using Welch's method
% Advantages: simultaneous reduction of a) bias due to leakage (via data
% tapering) and b) variability (by averaging the periodograms of
% Nb blocks of size Ns) of the periodogram.

[Pxx,w] = pwelch(fullseries);
plot(w/(2*pi),10*log10(Pxx))
xlabel('normalized frequency (cycle/sample)')
xlabel('f (ciclo/amostra)')
ylabel('power/frequency (dB/cycle/sample)')
ylabel('dB/ciclo/amostra')
title('PSD via WOSA')
title('DEP estimada (método WOSA)')

figure
[Pxx,w] = pwelch(I);
plot(w/(2*pi),10*log10(Pxx))
xlabel('f (ciclo/amostra)')
ylabel('dB/ciclo/amostra')
title('DEP estimada (método WOSA) dos resíduos do modelo TARFIMA')

t=length(fullseries)-max(k); % origin of prediction using k_max=100
                             % t=1106 for the Nile River minima series
                             % used by Beran in the first chapter of
                             % his book

%n = (t-max(k)+1):1:N;
n = (t-max(k)+1):1:(t+max(k));

figure
plot(n,fullseries(n),'k.-')
hold on
%pause
%plot((t+1:N),prev,'.:g')
plot((t+1:t+max(k)),prev,'.-g')
hold on
%pause
%plot((t+1:N),upper_bound,'x:g')
plot((t+1:t+max(k)),upper_bound,'x:g')
hold on
%pause
%plot((t+1:N),lower_bound,'*:g')
plot((t+1:t+max(k)),lower_bound,'*:g')
hold on
%pause
%plot((t+1:N),prev2,'.:r')
plot((t+1:t+max(k)),prev2,'.:r')
hold on
%pause

```

```

%plot((t+1:N),upper_bound2,'x:r')
plot((t+1:t+max(k)),upper_bound2,'x:r')
hold on
%pause
%plot((t+1:N),lower_bound2,'*:r')
plot((t+1:t+max(k)),lower_bound2,'*:r')

figure
plot((t+1:t+max(k)),delta_error_abs,'k.-')
xlabel('h')
ylabel('difference of absolute values of prediction errors')
ylabel('diferença entre os erros absolutos de previsão')

%=====
% Comparison of TARFIMAS with L=10, and 100

L2 = 10;
phi_B3 = ldiv(phi_temp,ma_farima,(L2+1));
phi_B4 = -1*phi_B3(2:(L2+1));
[m,n]=size(phi_B4);

if (m<n)
    phi_B4 = phi_B4';
end

ma100L2 = ldiv(1,phi_B3,101);
ma100L2 = ma100L2(2:length(ma100L2));

ma100quadL2 = ma100L2.^2;
var_erro3 = zeros(1,max(k));

mu = mean(fullseries(t:-1:1));
Z3 = fullseries(t:-1:t-(length(phi_B4)-1)) - mu;
I3 = filter(phi_B3,1,(fullseries'-mu));
pot_inov3 = var(I3);

for h=1:max(k)
    prev3(h) = Z3*phi_B4;
    Z3 = [prev3(h) Z3(1:(length(phi_B4)-1))];
end

prev3 = prev3 + mu; % predictions
x = fullseries(t+1:t+max(k)); % future observations

erro3 = x-prev3;

for h=1:max(k)
    if h==1
        var_erro3(h) = 1*pot_inov3; % vide Eq.(9.11) Morettin(2004)
    else
        var_erro3(h) = (ma100quadL2(h-1)*pot_inov2) + var_erro3(h-1);
    end
end

upper_bound3 = prev3 + 1.96*sqrt(var_erro3);
lower_bound3 = prev3 - 1.96*sqrt(var_erro3);

```

```

n = (t-max(k)+1):1:(t+max(k));

figure
plot(n,fullseries(n),'k.-')
hold on
%pause
%plot((t+1:N),prev,'.:g')
plot((t+1:t+max(k)),prev,'.:g')
hold on
%pause
%plot((t+1:N),upper_bound,'x:g')
plot((t+1:t+max(k)),upper_bound,'x:g')
hold on
%pause
%plot((t+1:N),lower_bound,'*:g')
plot((t+1:t+max(k)),lower_bound,'*:g')
hold on
%pause
%plot((t+1:N),prev2,'.:r')
plot((t+1:t+max(k)),prev3,'.:b')
hold on
%pause
%plot((t+1:N),upper_bound2,'x:r')
plot((t+1:t+max(k)),upper_bound3,'x:b')
hold on
%pause
%plot((t+1:N),lower_bound2,'*:r')
plot((t+1:t+max(k)),lower_bound3,'*:b')

```



```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Alexandre B. de Lima
%
% 14/10/2007
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% This script gives measures of the k-step forecastability of a time series
% as defined in p.24 of "An Introduction to Long-Memory Time Series Models
% and Fractional Differencing", Granger and Joyeux, 1980.
%
% Copyright(c) 2007 Universidade de São Paulo, Laboratório de Comunicações
% e Sinais
%
% Permission is granted for use and non-profit distribution providing that
% this notice be clearly maintained. The right to distribute any portion
% for profit or as part of any commercial product is specifically reserved
% for the author.

clear
close all

% note that : a) vectors with coefficients are column vectors, and
%              b) vetors with observations are row vectors

% =====

% Forecasting potential of the long-memory models (predictions based on
% an infinite past)

ar_farima = [1];
ma_farima = [1];
d = input('parameter "d": '); % parameter "d"

M = 200; % to derive the first M coef. of the MA(inf) representation

for j = 0:M
    D(j+1) = -gamma(j-d)/(gamma(j+1)*gamma(-d));
    % Eq.(16.30), pág. 475, "Análise de Séries Temporais",
    % Morettin & Toloi,2004
end

D = -1*D; % D(B) = (1-B)^d = 1 -d_1B -d_2B^2 - ... (estimated truncated
           % fractional difference filter)
           % "D" is a row vector

phi_B = conv(ar_farima,D'); % AR polynomial (1-phi_1B-phi_2B^2 - ...)
phi_B2 = -1*phi_B(2:length(phi_B));
%coef. of AR polynomial (S+ convention)

ma100 = ldiv(ma_farima,phi_B,M); % ldiv(b,a,N) is a function for
                                % the power series expansion of
                                % polynomial phi_B

ma100 = ma100';
ma100quad = ma100.^2; % coefficients to the square

Sk = cumsum(ma100quad); % vector with the variances of the k-step forecast

```

```

% errors (length = 100) for the ARFIMA model

% let x_t be an FARIMA(0,d,0) of the form
%      (1-B)^d{x_t} = w_t
% where it is assumed that var{w_t}=1 for convenience.
% The variance of x_t is given by
%      var{x_t} = gamma(-2d+1)/gamma^2(-d+1)
% (see, Eq.(3.2) in paper "Fractional Differencing", Hosking, 1981)
Vd = gamma(-2*d+1)/(gamma(-d+1))^2 ;

k = [1 10 20 100]; % so that you can reproduce the results presented in
% Tables 8.7 and 8.8 by Jan Beran in "Statistics for
% Long-Memory Processes", Chapman & Hall, 1994

% quantity "Rk" measures the k-step ahead forecasting potential of the LRD
% model, i.e., it's a standardized measure of how well one can predict
% x_{t+k} given the infinite past

for j=1:length(k)
    Rk(j) = 1 - Sk(k(j))/Vd;
end

%=====
% Forecasting potential of the TARFIMA(L), whose predictions are based on
% L previous observations

% INPUT DATA

L = input('order of the finite-dimensional representation: ');
%d2 = input('estimate of "d": ');

for j = 0:L
    D2(j+1) = -gamma(j-d)/(gamma(j+1)*gamma(-d));
end

D2 = -1*D2;

phi2_B = conv(ar_farima,D2'); % AR polynomial (1-phi_1.B-phi_2.B^2 - ...)
phi2_B2 = -1*phi2_B(2:length(phi2_B));
%coef. of AR polynomial (S+ convention)

ma100_2 = ldiv(ma_farima,phi2_B,M);
ma100_2 = ma100_2';
ma100quad_2 = ma100_2.^2;

Sk2 = cumsum(ma100quad_2);
% vector with the variances of the k-step forecast errors (length = 100)
% for the TARFIMA model

% let x_t be approximated by an TARFIMA(L) model as
%      (1 - phi_1.B - phi_2.B^2 - ... - phi_L.B^L)x_t = w_t
% where it is assumed that var{w_t}=1 for convenience.
% x_t has the MA(inf) representation
%      x_t = (1 - phi_1.B - phi_2.B^2 - ... - phi_L.B^L)^{-1}w_t or
%      x_t = (1 + psi_1.B + psi_2.B^2 + ... )w_t
% and the variance of x_t is given by
%      var{x_t} = 1 + sum_{j=1}^{inf}(psi_j^2)

```

```
% (see, Eq.(3.3.1) in p.91 of "Time Series: Theory and Methods", 2nd Ed.,
% Brockwell and Davis, 1991)
% Below, Sk2(M) gives an estimate of var{x_t}
Vd2 = Sk2(M);

% quantity "R2k" measures the k-step ahead forecasting potential of the
% truncated FARIMA model, i.e., it's a standardized measure of how well one
% can predict  $x_{t+k}$  given "L" (=order of TARFIMA model) past observations
for j=1:length(k)
    R2k(j) = 1 - Sk2(k(j))/Vd2;
end

% =====

% Forecasting potential for the AR(1) process with with rho(1)=1/9
% and rho(1)=2/3. The lag-1 correlation is chosen such that it is the same
% as for a ARFIMA(0,d,0) process with H=0.6 and 0.9 respectively.
rho_1 = 1/9;
rho_2 = 2/3;

% let x_t be an AR(1) of the form
%      (1-phi_1)^d{x_t} = w_t
% It is assumed that var{w_t}=1 for convenience.

Vh06_AR1 = 1/(1 - rho_1^2); % variance of AR(1) process with H=0.6;
Vh09_AR1 = 1/(1 - rho_2^2); % variance of AR(1) process with H=0.9;
                        % see Brockwell & Davis (1991),
                        % Morettin & Toloi (2004)

for j=1:100
    Skh06_AR1(j) = Vh06_AR1*(1 - rho_1^(2*j));
    Skh09_AR1(j) = Vh09_AR1*(1 - rho_2^(2*j));
end

for j=1:length(k)
    R3k(j) = 1 - Skh06_AR1(k(j))/Vh06_AR1;
    R4k(j) = 1 - Skh09_AR1(k(j))/Vh09_AR1;
end
```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Alexandre B. de Lima
%
% 28/10/2007
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

clear
close all

j=sqrt(-1);

d = input('parameter "d": ');
%d=0.3859;
H = d + 0.5;
da = 1; %potencia do ruido branco Gaussiano

yes = 'y';
yes2 = 'Y';
j1 = input('AR coefficients? Y/N [Y]: ', 's');
if ( strcmp(yes,j1) | strcmp(yes2,j1) )
    name_ar_farima = input('file with the estimated AR coef. of FARIMA:
    ', 's');
    ar_farima = load(name_ar_farima);
else
    ar_farima = [1];
end;

j2=input('MA coefficients? Y/N [Y]: ', 's');
if ( strcmp(yes,j2) | strcmp(yes2,j2) )
    name_ma_farima = input('file with the estimated MA coef. of FARIMA:
    ', 's');
    ma_farima = load(name_ma_farima);
else
    ma_farima = [1];
end;

ar_farima = -1*ar_farima;
% remember that the coefficients given by S-PLUS are the
% negative of the coefficients used in MATLAB (convention)
% See S-PLUS Guide to Statistics, Vol. 2

ar_farima = [1; ar_farima];
ma_farima = -1*ma_farima;
ma_farima = [1; ma_farima];

N=2048;
Deltaf = (0.5/N);

i=0;

% Calculo da Densidade Espectral de Potencia do modelo FARIMA(p,d,0)
for f=(0.5/N):(0.5/N):0.5
    A = [1 -exp(-j*2*pi*f)];
    B = (abs(sum(A))^(2*d)*(da^2));
    if (length(ma_farima)>1)
        A2 =

```

```

end
C = [1 -0.3014*exp(-j*2*pi*f) 0.3988*exp(-j*2*pi*f*2)];
D = (abs(sum(C)))^2;
i = i+1;

Sd(i) = B/D;
end

Pdf = Deltaf*Sd;
Pd = sum(Pdf);

% Estimação da DEP da serie simulada
se =dlmread('fGn_(4096)pts_alpha(0.8)_NrOfVM(1)_
TopLev(9)_realiz3_filtroIIR2.txt');
[Pxx,w] = pwelch(se,[],[],(2*N));

Pxx = Pxx(2:N+1);

Pf = Deltaf*Pxx;
P = sum(Pf);

% Calculo da DEP do modelo T-FARIMA
ar_farima = [0.3014; -0.3988];
% sinais dos coeficientes de acordo com a convenção do S-PLUS = -1*Matlab

% Determina Polinomio "Vphi(B)=phi(B)D(B)" de acordo com a
% Eq. (16.26), pg. 475, do Livro "Análise de Séries Temporais", de
% Morettin e Toloi,
% onde D(B) = (1-B)^d = 1 -d(1)B -d(2)B^2 - ...

ar_farima = -1*ar_farima;
ar_farima = [1; ar_farima];

L=50;
M = length(ar_farima);
K = L-(M-1);

% calcula d(j)
for j = 0:K
    D(j+1) = -gamma(j-d)/(gamma(j+1)*gamma(-d));
end

D = -1*D; % "D" é um vetor linha

Vphi_B = conv(ar_farima,D');
Vphi_B2 = -1*Vphi_B(2:length(Vphi_B));

Btfarima = 1;
Atfarima = Vphi_B;
F=(0.5/N):(0.5/N):0.5;
H = freqz(Btfarima,Atfarima,F,1);
St = abs(H).^2;

Ptf = Deltaf*St;
Pt = sum(Ptf);

% Calculo da DEP do modelo AR estimado

```

```

Bar = 1;
Aar = load('ar10h09.txt');
Aar = -1*Aar;
Aar = [1; Aar];
H2 = freqz(Bar,Aar,F,1);
Sar = abs(H2).^2;

Par = Deltaf*Sar;
Ptar = sum(Par);

R=Pt/Pd;
R1=Pt/P;
R2=Pt/Ptar;

f=(0.5/N):(0.5/N):0.5;

figure %figura 1
plot(f,10*log10(St),'r')
xlabel('frequency')
ylabel('PSD (dB)')

hold on
pause

plot(f,10*log10(R*Sd),'g')

hold on
pause

plot(f,10*log10(R2*Sar),'g')

erro = 10*log10(St)-10*log10(R*Sd);

figure %figura 2
plot(f,10*log10(St)-10*log10(R*Sd))
xlabel('frequency')
ylabel('dB')

figure %figura 3
loglog(f,R*Sd,'g')
xlabel('frequency')
ylabel('PSD')

hold on
pause
loglog(f,St,'r')

hold on
pause
loglog(f,R1*Pxx,'m')

%hold on
%pause
%loglog(f,R2*Sar,'b')

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%   Generate.m
%
%   Fernando L. de Mello
%   Alexandre B. de Lima
%
%   10/2006
%   12/2007
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%   References:
%   1) J.-A. Bäckar, "A Framework for Implementing Fractal Traffic Models
%   in Real Time", M.Sc. Thesis, SERC, Melbourne, 2000.
%   2) F. L. de Mello, "Estudo e Implementação de um Gerador de Tráfego
%   com Dependência de Longa Duração", M.Sc. Thesis, Univ. of Sao Paulo, 2006.
%   3) F. L. de Mello, A. B. de Lima, M. Lipas, J. R. A. Amazonas,
%   "Geração de Séries Auto-Similares Gaussianas via Wavelets para Uso em
%   Simulações de Tráfego", IEEE Latin America Transactions, Março, 2007.
%   4) A. B. de Lima, F. L. de Mello, M. Lipas, J. R. A. Amazonas,
%   "A Generator of Teletraffic with Long and Short-Range Dependence",
%   12th CAMAD Workshop (part of the 18th IEEE PIMRC07), Greece, 2007.
%   5) D. B. Percival and A. T. Walden, "Wavelet Methods for Time Series
%   Analysis", Cambridge University Press, 2000.
%   6) R. H. Riedi, M. S. Crouse, V. J. Ribeiro, R. G. Baraniuk,
%   "A multifractal wavelet model with application to network traffic"
%   IEEE Transactions on Information Theory, 45(3), pp.992-1018, April, 1999.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%   'Generate.m' is a wavelet-based generator of self-similar sample paths of
%   Fractional Gaussian Noise (FGN) and Multifractal Wavelet Model (MWM).
%   It works in conjunction with functions 'Model.m' and 'Recon.m'
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%   Models:
%   1 - wavelet coefficients as white noise random processes (FGN synthesis)
%   2 - MWM
%   3 - wavelet coefficients as AR(1) random processes (FGN synthesis)
%   9 - Reconstrucao Analise Matlab
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%   Parameters:
%   TopLev    % number of iterations (reconstruction using Pyramid algorithm)
%   NrOfVM    % number of vanishing moments of the Daubechies wavelet function
%   modelo    % stochastic model used
%   Lm        % number of generated samples
%   alpha     % scaling exponent; remember that H = (1+alpha)/2 and d = alpha -0.5
%             % where 'H' is the Hurst parameter and 'd' is the fractional parameter
%             % of an ARFIMA(p,d,q) model
%   p         % shape parameter of the beta probability density function
%             % In the MWM model, the variance of the multiplier R is given by
%             % var[R] = 1/(2p+1).
%   phi       % AR(1) coefficient
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%   use as: [Data, nome_arquivo]=Generate(TopLev, NrOfVM, modelo, varargin)
%           or [Data]=Generate(TopLev, NrOfVM, 9, varargin) - modelo 9
%   where:

```

```
% for 'modelo'= 1 (FGN, white noise wav. coef.), 'varargin' will be 2 parameters: "Lm, alpha"
% for 'modelo'= 2 (MWM), 'varargin' sera dois parametros: "alpha, p"
% for 'modelo'= 3 (FGN, AR(1) wav. coef.), 'varargin' will be 3 parameters: "Lm, alpha, phi"
% for 'modelo'= 9, 'varargin' will be five parameters: "Lm, vetor_coefs (C), L, TopLev, wname"
%
% Example of interactive use: generation of an FGN signal with 4096 samples
%
%             alpha = 0.6 (H=[1+alpha]/2=0.8),
%
%             from 1 root scale (approximation) coefficient
%
% >> [Data, archive_name] = Generate(12, 1, 1, 4096,0.6)
%
% wname =
% db1
% g =
% 0.7071    0.7071
% Lf =
% 2
% h =
% 0.7071   -0.7071
% Data =
% Columns 1 through 6
% [1x1 struct] [1x1 struct] [1x1 struct] [1x1 struct]
% [1x1 struct] [1x1 struct]
% Columns 7 through 12
% [1x1 struct] [1x1 struct] [1x1 struct] [1x1 struct]
% [1x1 struct] [1x1 struct]
% Column 13
% [1x1 struct]
% archive_name =
% fGn_(4096)pts_alpha(0.6)_NrOfVM(1)_TopLev(12)
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [Data, archive_name] = Generate(TopLev, NrOfVM, model, varargin)

N = int2str(NrOfVM); % N is a string
wname = 'db'; % must concatenate N obtain dbN
% Compute the corresponding scaling filter.
wname = strcat(wname,N)
% Compute the corresponding scaling filter.
g = dbwavf(wname);
g = g * sqrt(2)
Lf = length(g) % must be 2N
h = fliplr(g); h = (-1).^(0:Lf-1).*h
% Eq.(75b) of book "Wavelet Methods for Time Series Analysis"

% call function Model, which generates an approximation on scale j=0.
[Data, archive_name] = Model(TopLev, NrOfVM, h, g, model, varargin{:});
```


[illegible]

```

function approxs = appProcess(n, modelo, varargin)
    switch modelo
        case 1,
            approxs = 4+randn(1,n);
        case 2, approxs = 4+randn(1,n);
        case 3, approxs = zeros(1,n);
        case 9,
            vetor_coefs = varargin{1};
            L=varargin{2};
            TopLev = varargin{3};
            wname = varargin{4};
            coefsA = appcoef(vetor_coefs,L,wname,TopLev);
            if length(coefsA) == n ,
                approxs = coefsA;
            else
                warning('Numero de coeficientes da dwt nao bate com o
                necessario para reconstituir o sinal');
                approxs = coefsA(1:n);
            end
        otherwise, disp('Modelo nao implementado'); approxs = [];
    end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%internal:
function [Data] = detProcess(level, TopLev, Data, modelo, varargin)
    if modelo ~= 9,
        alpha = varargin{1};
        ctrvariance = 2^( (level - TopLev) * alpha/2 );
    end

    n = diff( Data{level+(1)}.kp ) + 1;

    switch modelo
        case 1,
            Data{level+(1)}.det = ctrvariance*randn(1,n);
        case 2, %MWM:
            p = varargin{2};
            if (level == TopLev) & (p == 0),
                p = (2^alpha - 1)/(2 - 2^alpha);
                Data{0+(1)}.p = p;
                for i = 1:TopLev,
                    p = (2*p+1)/2^alpha - 1;
                    Data{i+(1)}.p = p;
                end
            elseif (level == TopLev) & (p > 0),
                Data{TopLev+(1)}.p = p;
                for i = TopLev-1:-1:0,
                    p = (2^alpha*(p+1)-1)/2;
                    Data{i+(1)}.p = p;
                end
                p = Data{TopLev+(1)}.p;
            elseif level ~= TopLev,
                p = Data{level+(1)}.p;
            end
            R01 = betarnd(p,p,1,n);
            R = 2.*R01-1;
    end

```

```

        Data{level+1}.det = R.*Data{level+1}.app;
    case 3, phi = varargin{2};
        inject = ctrvariance*randn(1,n);
        Data{level+1}.det(1) = phi*0 + inject(1);
        for i=2:n,
            Data{level+1}.det(i) = phi*Data{level+1}.det(i-1) + inject(i);
        end
        clear inject;
    case 9,
        vetor_coefs = varargin{1};
        L = varargin{2};
        coefsD = detcoef(vetor_coefs,L,level);
        if length(coefsD) == n ,
            Data{level+1}.det = coefsD;
        else
            warning('Numero de coeficientes da dwt nao bate com o
necessario para reconstituir o sinal');
            Data{level+1}.det = coefsD(1:n);
        end

    otherwise, error(['Modelo ', int2str(modelo), ' nao implementado!!!']);
end

```

```
function approxs = Recon(level, NrOfVM, Data, h, g) % h will be high-pass filter coefficients.
    Lf = length(h);

    kp = Data{level + (1)}.kp;

    % Eqs. (7.11), (7.12), (7.13) and (7.14) of Backar M.Sc thesis (see p.49):
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

    ku = 2*kp;

    kc = ku + [0 Lf-1];

    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

    kp = Data{level-1 + (1)}.kp;

    appro = Data{level + (1)}.app;
    detail = Data{level + (1)}.det;

    appup = dyadup(appro,0);
    detup = dyadup(detail,0);

    appconv = conv(appup,g);
    detconv = conv(detup,h);

    indices = [kp(1)-kc(1)+(1):kp(2)-kc(1)+(1)];

    appro = appconv(indices) + detconv(indices);

    approxs = appro;
```

Apêndice B – Código em S para S-PLUS

```
#####
# author: Alexandre B. de Lima
# version of: 24/12/2007
#
# Given a long memory time series, this script performs
# - statistical tests for LRD
# - estimation of the Hurst parameter
# - estimation of an FARIMA(p,d,q), a high order AR(p) (using the AIC
# criterion), and an AR(2) model #
# - "k-steps-ahead-forecasts" computed from the Kalman filter prediction
# equations associated to a state-space representation of the AR(p)
# model
#
# Copyright(c) 2007 Universidade de São Paulo, Laboratório de
# Comunicações e Sinais
# Permission is granted for use and non-profit distribution providing that
# this notice be clearly maintained. The right to distribute any portion
# for profit or as part of any commercial product is specifically reserved
# for the author.
#####

setTextOutputRouting(normalTextWindow = "Report", errorWindow = "Report")

# Input time series
namets=scan(file="",what=character(),n=1)
# Enter the name of the ASCII file which contains the time series
# to be analyzed. Note that this file must be in your S+ home directory

myts.df = importData(file=namets)
# Load time series from the specified file

myts.matrix = as.matrix.data.frame(myts.df)

plot(myts.matrix, type="l")
title("time series")
axes(xlab="t", ylab="")
guiSave("GraphSheet", Name="GSD2",FileName=paste("myts_plot",".sgr",
sep=""))
guiClose("GraphSheet","GSD2")

# histogram of time series
hist(myts.matrix)
title("Histogram of time series")
guiSave("GraphSheet", Name="GSD2",FileName=paste("myts_hist",".sgr",
sep=""))
```

```

guiClose("GraphSheet", "GSD2")

# QQ-plot of time series
qq.plot.myts <- qqnorm(myts.matrix)
title("qqplot of time series")
guiSave("GraphSheet", Name="GSD2", FileName=paste("myts_qqplot", ".sgr",
sep=""))
guiClose("GraphSheet", "GSD2")

mu = mean(myts.matrix)
# Be careful here ... remember that the coefficients in the function "ar"
# are for the series with the mean(s) removed. You must also demean the
# series in order to have a meaningful state-space representation
# when using the "GetSsfArma" function (try not to use and you will see
# for yourself that the Kalman recursions do not work)

# -----
# Testing for stationarity (null hypothesis that series is I(0), see
# pp. 123 of "Modeling Financial Time Series with S-PLUS")
kpss.out = stationaryTest(myts.matrix, trend="c")
kpss.out

# demean the series
myts.matrix = myts.matrix - mu

N = length(myts.matrix)      # number of points of time series
h = 100                      # t(origin of forecasts) = N-h
myts = timeSeries(myts.df) # create a "timeSeries" object

# Estimate power spectrum density of demeaned time series
spec.pgram(myts.matrix, demean=T, spans=c(9,7), plot=T)
# smoothed periodogram, Daniell smoother
guiSave("GraphSheet", Name="GSD2", FileName=paste("pgram_myts", ".sgr",
sep=""))
guiClose("GraphSheet", "GSD2")
spec.pgram(myts.matrix, demean=T, spans=1, plot=T)
guiSave("GraphSheet", Name="GSD2",
FileName=paste("raw_pgram_myts", ".sgr", sep=""))
guiClose("GraphSheet", "GSD2")

# -----
# Statistical Tests for Long Memory

# R/S statistic
rosTest(myts.matrix)

# GPH test
#gph.myts = gphTest(myts.matrix, demean=T, spans=c(9,7))
gph.myts = gphTest(myts.matrix, taper=0.1, demean=F, spans=c(9,7))
gph.myts

# -----
# Estimation of Hurst Parameter

# whittle's method
d.w = d.whittle(myts.matrix, demean= F, output="d")
d.w

```

```
# periodogram method
d.smooth.pgram = d.pgram(myts.matrix,taper=0.1, demean=F, spans=c(9,7),
output="d")
d.smooth.pgram

# R/S Analysis
d.ros(myts.matrix,minK=50,k.ratio=2,minNumPoints=10,output="d",plot=T)
# because the length of time series is sufficiently large (N=4096,
# for instance)
guiSave("GraphSheet", Name="GSD2",FileName=paste("d_est_ros",".sgr",
sep=""))
guiClose("GraphSheet","GSD2")

#d.ros(myts.matrix,minK=4,k.ratio=2,minNumPoints=10,output="d",plot=T)
# because it's a small time series, like the Nile river minima
# for the years 1007 to 1206
#guiSave("GraphSheet", Name="GSD2",FileName=paste("d_est_ros",".sgr",
sep=""))
#guiClose("GraphSheet","GSD2")

# -----
# Estimate FARIMA(p,d,q) model and SEMIFARmodel

myts.farima.fit.bic = FARIMA(myts.df,p.range=c(0,2),q.range=c(0,2),
mmax=1)
# estimated model
myts.farima.fit.bic$model
# unit roots ?
myts.farima.fit.bic$m
m = myts.farima.fit.bic$m

myts.semifar.fit = SEMIFAR(myts.df,p.range=c(0,2),trace=F)
myts.semifar.fit$model
myts.semifar.fit$m

# -----
# plot sample ACF of series and theoretical ACF of the fitted model
if (m == 0)
{
# diagnostic of the fitted FARIMA model
plot(myts.farima.fit.bic)
guiSave("GraphSheet", Name="GSD2",
FileName=paste("diagnostics_farima_fit",".sgr",sep=""))
guiClose("GraphSheet","GSD2")

# plot periodogram of FARIMA residuals
psd.resid.farima.fit <- spec.pgram(myts.farima.fit.bic$residuals,
spans=c(9,7), plot=T)
guiSave("GraphSheet", Name="GSD2",FileName=paste
("pgram_residuals_farima_fit",".sgr",sep=""))
guiClose("GraphSheet","GSD2")

# plot theoretical ACF of estimated FARIMA model
farima.mod = list(ar=myts.farima.fit.bic$model$ar,
ma=myts.farima.fit.bic$model$ma, sigma2=1,
d=myts.farima.fit.bic$model$d)
```

```

arfima.acf = acf.FARIMA(farima.mod,lag.max=200)
plot(arfima.acf$lags, arfima.acf$acf/arfima.acf$acf[1],
type="h",xlab="lags", ylab="ACF")
guiSave("GraphSheet", Name="GSD2",FileName=paste("acf_FARIMA",
".sgr",sep=""))
guiClose("GraphSheet","GSD2")

# -----
# estimate an AR(p) model using AIC
myts.ar.fit.aic = ar(myts, aic=T, order.max=40)
myts.ar.fit.aic$order
myts.ar.fit.aic$ar
coef.ar.fit.aic = as.matrix(myts.ar.fit.aic$ar)
exportData(coef.ar.fit.aic, "coef_ar_fit_aic.txt",type="ASCII",
colNames=FALSE)

# -----
# plot SACF of signal, theoretical ACFs of estimated FARIMA and
# high-order AR models
sacf.myts = acf(myts.matrix,lag=200)
lines(arfima.acf$lags, arfima.acf$acf/arfima.acf$acf[1])
arp.mod = list(ar=as.vector(myts.ar.fit.aic$ar),sigma2=1,d=0)
ar.acf = acf.FARIMA(arp.mod,lag.max=200)
lines(ar.acf$lags,ar.acf$acf/ar.acf$acf[1])
guiSave("GraphSheet", Name="GSD2",FileName=paste("myts_SACF_ACF",
".sgr",sep=""))
guiClose("GraphSheet","GSD2")

# -----
# diagnostics of the AR(p) model fit
resid.ar.aic.fit = myts.ar.fit.aic$resid
acf.resid.ar.aic.fit= acf(resid.ar.aic.fit, type="correlation", plot=T)
guiSave("GraphSheet", Name="GSD2",FileName=paste("resid_ar_AIC_SACF",
".sgr",sep=""))
guiClose("GraphSheet","GSD2")
spec.pgram(resid.ar.aic.fit, demean=T, spans=c(9,7), plot=T)
guiSave("GraphSheet", Name="GSD2",FileName=paste("pgram_resid_ar_AIC",
".sgr",sep=""))
guiClose("GraphSheet","GSD2")

}
else
{
# series has unit roots!
# diagnostic of the fitted SEMIFAR model
plot(myts.semifar.fit)
guiSave("GraphSheet", Name="GSD2",
FileName=paste("diagnostics_semifar_fit",".sgr",sep=""))
guiClose("GraphSheet","GSD2")

# plot periodogram of SEMIFAR residuals
psd.resid.semifar.fit <- spec.pgram(myts.semifar.fit$residuals,
spans=c(9,7), plot=T)
guiSave("GraphSheet", Name="GSD2",FileName=
paste("pgram_residuals_semifar_fit",".sgr",sep=""))
guiClose("GraphSheet","GSD2")

```



```
# plot theoretical ACF of estimated FARIMA model (here I use
# the Whittle estimate for "d")
farima.mod = list(ar=myts.farima.fit.bic$model$ar,
ma=myts.farima.fit.bic$model$ma, sigma2=1, d=d.w)
arfima.acf = acf.FARIMA(farima.mod,lag.max=200)
plot(arfima.acf$lags, arfima.acf$acf/arfima.acf$acf[1], type="h",
xlab="lags", ylab="ACF")
guiSave("GraphSheet", Name="GSD2",FileName=paste("acf_FARIMA",
".sgr",sep=""))
guiClose("GraphSheet","GSD2")

#sacf.myts = acf(myts.matrix,lag=200)
#guiSave("GraphSheet", Name="GSD2",FileName=paste("myts_SACF",
".sgr",sep=""))
#guiClose("GraphSheet","GSD2")

# -----
# estimate an AR(p) model using AIC
myts.ar.fit.aic = ar(myts, aic=T, order.max=40)
myts.ar.fit.aic$order
myts.ar.fit.aic$ar
coef.ar.fit.aic = as.matrix(myts.ar.fit.aic$ar)
exportData(coef.ar.fit.aic, "coef_ar_fit_aic.txt",type="ASCII",
colNames=FALSE)

# -----
# plot SACF of signal, theoretical ACFs of estimated FARIMA and high-order
# AR models
sacf.myts = acf(myts.matrix,lag=200)
lines(arfima.acf$lags, arfima.acf$acf/arfima.acf$acf[1])
arp.mod = list(ar=as.vector(myts.ar.fit.aic$ar),sigma2=1,d=0)
ar.acf = acf.FARIMA(arp.mod,lag.max=200)
lines(ar.acf$lags,ar.acf$acf/ar.acf$acf[1])
guiSave("GraphSheet", Name="GSD2",FileName=paste("myts_SACF_ACF",
".sgr",sep=""))
guiClose("GraphSheet","GSD2")

# -----
# diagnostics of the AR(p) model fit
resid.ar.aic.fit = myts.ar.fit.aic$resid
acf.resid.ar.aic.fit= acf(resid.ar.aic.fit, type="correlation", plot=T)
guiSave("GraphSheet", Name="GSD2",FileName=paste("resid_ar_AIC_SACF",
".sgr", sep=""))
guiClose("GraphSheet","GSD2")
spec.pgram(resid.ar.aic.fit, demean=T, spans=c(9,7), plot=T)
# smoothed periodogram, Daniell smoother
guiSave("GraphSheet", Name="GSD2",FileName=
paste("pgram_resid_ar_AIC", ".sgr", sep=""))
guiClose("GraphSheet","GSD2")

}

myts.ar.fit.aic$order

# -----
# estimate an AR(2) model
myts.ar2.fit = ar(myts, aic=F, order=2)
```

```

myts.ar2.fit$order
myts.ar2.fit$ar
coef.ar2.fit = as.matrix(myts.ar2.fit$ar)
exportData(coef.ar2.fit, "coef_ar2_fit.txt", type="ASCII", colNames=FALSE)

# diagnostics of the AR(2) model fit
resid.ar2.fit = myts.ar2.fit$resid
acf.resid.ar2.fit= acf(resid.ar2.fit, type="correlation", plot=T)
guiSave("GraphSheet", Name="GSD2", FileName=paste("resid_ar2_SACF", ".sgr",
sep=""))
guiClose("GraphSheet", "GSD2")
spec.pgram(resid.ar2.fit, demean=T, spans=c(9,7), plot=T)
# smoothed periodogram, Daniell smoother
guiSave("GraphSheet", Name="GSD2", FileName=paste("pgram_resid_ar2", ".sgr",
sep=""))
guiClose("GraphSheet", "GSD2")

# -----
# Forecasts with the Kalman Filter
# AR(p) model using AIC
ssf.myts.ar.fit.aic = GetSsfArma(ar=myts.ar.fit.aic$ar, ma=NULL,
sigma=sqrt(myts.ar.fit.aic$var.pred))
myts.new = c(myts.matrix[1:(N-h)], rep(NA, h))
# you must append a seq of h missing values
# (see p.532 of the book "Modeling Financial Time Series with S-PLUS")

Kalman.est = SsfMomentEst(myts.new, ssf.myts.ar.fit.aic, task="STPRED")
# you must use a demeaned time series!

myts.fcst = Kalman.est$response.moment + mu
fcst.var = Kalman.est$response.variance
exportData(myts.fcst[(N-h+1):N], "fcst_Kalman_ar_aic.txt", type="ASCII",
colNames=FALSE)
upper = myts.fcst + 1.96*sqrt(fcst.var)
lower = myts.fcst - 1.96*sqrt(fcst.var)
upper[1:(N-h)] = lower[1:(N-h)] = NA
exportData(upper[(N-h+1):N], "upper_Kalman_ar_aic.txt", type="ASCII",
colNames=FALSE)
exportData(lower[(N-h+1):N], "lower_Kalman_ar_aic.txt", type="ASCII",
colNames=FALSE)

#tsplot((myts.new[(N-2*h+1):N]+mu), myts.fcst[(N-2*h+1):N],
upper[(N-2*h+1):N], lower[(N-2*h+1):N], lty=c(1,2,2,2))

tsplot((myts.matrix[(N-2*h+1):N]+mu), myts.fcst[(N-2*h+1):N],
upper[(N-2*h+1):N], lower[(N-2*h+1):N], lty=c(1,2,2,2))
guiSave("GraphSheet", Name="GSD2", FileName=
paste("myts_KalmanPred_ARaic", ".sgr", sep=""))
guiClose("GraphSheet", "GSD2")

# -----
# Forecasts with the Kalman Filter
# AR(2) model
ssf.myts.ar2.fit = GetSsfArma(ar=myts.ar2.fit$ar, ma=NULL,
sigma=sqrt(myts.ar2.fit$var.pred))
Kalman.est2 = SsfMomentEst(myts.new, ssf.myts.ar2.fit, task="STPRED")

```

```

myts.fcst2 = Kalman.est2$response.moment + mu
fcst.var2 = Kalman.est2$response.variance
exportData(myts.fcst2[(N-h+1):N], "fcst_Kalman_ar2.txt",type="ASCII",
colNames=FALSE)

upper2      = myts.fcst2 + 1.96*sqrt(fcst.var2)
lower2      = myts.fcst2 - 1.96*sqrt(fcst.var2)
upper2[1:(N-h)] = lower2[1:(N-h)] = NA
exportData(upper2[(N-h+1):N], "upper2_Kalman_ar_aic.txt",type="ASCII",
colNames=FALSE)
exportData(lower2[(N-h+1):N], "lower2_Kalman_ar_aic.txt",type="ASCII",
colNames=FALSE)

tsplot((myts.matrix[(N-2*h+1):N]+mu),myts.fcst2[(N-2*h+1):N],
upper2[(N-2*h+1):N],
lower2[(N-2*h+1):N], lty=c(1,2,2,2))
guiSave("GraphSheet", Name="GSD2",FileName=paste("myts_KalmanPred_AR2",
".sgr",sep=""))
guiClose("GraphSheet","GSD2")

#####

setTextOutputRouting(normalTextWindow = "Default", errorWindow
= "Default")

```