

SELEÇÃO DE VARIÁVEIS DE ENTRADA PARA PREVISORES NEURAIIS
DE CARGA

Roberto Miguel Gutiérrez Velásquez

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS
PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE
FEDERAL DE RIO DE JANEIRO COMO PARTE DOS REQUISITOS
NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS DA
ENGENHARIA ELÉTRICA

Aprovada por:

Prof. Alexandre P. Alves da Silva, Ph.D.

Prof. Djalma Mosqueira Falcão, Ph.D.

Prof. Luiz Pereira Calôba, Dr. Ing .

Prof. Marcelo Cunha Medeiros, D.Sc.

Prof. Carlos Eduardo Pedreira, Ph.D.

Prof. Agnaldo José de Rocha Reis, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

NOVEMBRO DE 2006

VELÁSQUEZ, ROBERTO MIGUEL GUTIÉRREZ

Seleção de Variáveis de Entrada para Previsores
Neurais de Carga [Rio de Janeiro] 2006

VII, 125 p. 29,7 cm (COPPE/UFRJ, D.Sc.,
Engenharia Elétrica, 2006)

Tese - Universidade Federal do Rio de
Janeiro, COPPE

1. Previsão de Carga de Curto Prazo
2. Seleção de Variáveis Relevantes
3. Dinâmica não Linear
4. Caos
5. Informação Mútua
6. Falsos Vizinhos
7. Dimensão da Correlação
8. Teoria de Wavelets
9. Redes Neurais
10. Séries temporais

I. COPPE/UFRJ II. Título (série)

DEDICATÓRIA

Dedico esta tese à minha esposa, Adriana pelo seu amor e apoio nestes anos, pela paciência da espera, pela renúncia a inúmeros momentos de lazer, pela compreensão a minhas freqüentes ausências e principalmente pelas incontáveis correções ao texto. Obrigado!

AGRADECIMENTOS

Esta tese tem a particularidade de ter sido desenvolvida em três cidades diferentes. Assim, neste longo período, foram muitas as pessoas que conheci e que, de uma ou outra forma, contribuíram para sua gestação.

No Rio de Janeiro: agradeço a meu orientador Alexandre pelas sempre pertinentes observações e precisas sugestões, sem as quais, não teria conseguido finalizar este trabalho. Ao professor Djalma, que desde os tempos do mestrado, sempre acreditou no meu potencial e ajudou a desenvolvê-lo. Aos colegas do CEPEL-Centro de Pesquisas de Energia Elétrica que foram partícipes dos primórdios desta tese.

Em Salvador: agradeço a meus queridos colegas baianos da COELBA-Companhia de Eletricidade do Estado da Bahia, lugar em que trabalhei entre os anos 2003-2005 e principalmente a Edmilson Santana, que sempre facilitou minhas eventuais ausências para cumprir com minhas atividades acadêmicas “paralelas”.

No Recife: agradeço a meus novos colegas pernambucanos da CELPE-Companhia Energética de Pernambuco, particularmente a Ary Pinto Ribeiro, que em várias oportunidades mostrou interesse pelo meu trabalho, permitindo ausências e viagens para o Rio de Janeiro.

A todos eles, **MUITO OBRIGADO!**

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

SELEÇÃO DE VARIÁVEIS DE ENTRADA PARA PREVISORES NEURAIIS DE CARGA

Roberto Miguel Gutiérrez Velásquez

Novembro/2006

Orientadores: Alexandre P. Alves da Silva

Djalma Mosqueira Falcão.

Programa: Engenharia Elétrica

A qualidade da previsão de carga fornecida por qualquer modelo está fortemente associada à correta escolha das variáveis explicativas. Se tais variáveis não são suficientemente representativas do problema a ser modelado, a capacidade de ajuste e previsão do modelo fica comprometida. Nesta tese é apresentado um método automático de seleção de variáveis relevantes para modelos de previsão. A metodologia de seleção está baseada na aplicação prática do Teorema de Reconstrução de Espaço Fase de Takens em séries de carga a partir da definição prévia de duas variáveis: Dimensão da Imersão e Passo de Reconstrução. Para o cálculo da Dimensão da Imersão são utilizadas duas técnicas originadas no âmbito da dinâmica não linear e da teoria do caos: Dimensão da Correlação e Falsos Vizinhos Mais Próximos. Já para a determinação do Passo de Reconstrução, a metodologia utilizada é conhecida como Método da Informação Mútua, originária da Teoria da Informação. Visando aproximar as hipóteses dos métodos para assim realizar uma correta aplicação do Teorema de Takens, é aplicada a Teoria de *Wavelets* nas séries de carga antes do processo de reconstrução. Com o mesmo objetivo são reduzidas as componentes de tendência e sazonalidade das séries antes do processo de reconstrução. A qualidade do processo de reconstrução de Espaço Fase é avaliada usando estatísticas de ajuste e previsão em modelos de redes neurais.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

INPUT SELECTION OF VARIABLES FOR LOAD NEURAL FORECASTERS

Roberto Miguel Gutiérrez Velásquez

November/2006

Advisors:: Alexandre P. Alves da Silva
Djalma Mosqueira Falcão.

Department: Electrical Engineering

The quality of load forecasting, delivered by any model, is strongly related to the choice of the input variables. If those variables are not a truly representation of the problem that is going to be modelled, the capacity of model's adjustment and forecast can be jeopardized. This thesis presents an automatic method of selection of the most relevant variables for forecast models. The selection methodology is based on the practical application of the Takens' Embedding Space Reconstruction Theorem, in load series from the previous definition of two variables; Embedding Dimension and Time Delay. In order to calculate the Embedding Dimension we can use two techniques that come from the non-linear dynamic area, and from the Chaos Theory; Dimension of Correlation and False Nearest Neighbours. The calculation of Time Delay is known as Mutual Information Method, which comes from the Information Theory. In order to obtain a better relation among the aforementioned methods, and to get the right application of the Takens' Theorem, one use the Wavelets Theory, in the load series the series of load before the reconstruction process. Besides one discard, we reduce the components of trend and seasonality of the series before reconstruction process. The quality of the reconstruction process of the space of phase is evaluated using some statistics of adjustment and forecast via neural networks.

INTRODUÇÃO.....	1
1.1 Previsão de Carga	2
1.2 Redes Neurais	4
1.3 Previsão de Carga de Curto Prazo Usando Redes Neurais.....	5
1.4 Extração de Variáveis Relevantes	5
1.4.1 Abordagem Estatística	6
1.4.2 Abordagem Via Generalização do Erro	7
1.4.3 Outras Abordagens	8
1.5 Extração de Variáveis a partir de uma Série Histórica.....	9
1.6 Método Físico de Extração de Variáveis a partir de uma Série Histórica.....	11
1.7 Motivação	12
1.8 Objetivos da Tese	13
1.9 Contribuições da tese.....	14
1.10 Etapas do trabalho	15
CAPÍTULO 2	17
CONCEITOS BÁSICOS DE SISTEMAS DINÂMICOS E CAOS	17
2.1 Introdução.....	17
2.2 Aspectos Históricos	18
2.3 Definição de Caos.....	19
2.4 Bifurcações	20
2.5 Atrator.....	21
2.6 Atrator: Definição Formal	23
2.7 Expoente de Lyapunov	25
2.8 Dimensões de um Atrator	29
2.9 Conclusões.....	34
CAPÍTULO 3	36
RECONSTRUÇÃO DO ESPAÇO DE IMERSÃO	36
3.1 Introdução.....	36
3.2 Conceito de Reconstrução	36
3.3 Teorema de Reconstrução de Takens	38
3.4 Reconstrução de Takens Autoregressiva.....	39
3.5 Calculando a Dimensão de Correlação.....	41
3.6 Influência do Passo de Reconstrução τ na Qualidade da Reconstrução	43
3.6.1 Sistema de Van der Pol	44
3.6.2 Reconstrução do Sistema de Van der Pol.....	45
3.6.3 Exemplo 4. Sistema de Hénon	46
3.6.4 Reconstrução do Sistema de Hénon.....	47
3.7 Determinação do Passo de Reconstrução τ	48
3.7.1 Informação Mútua.....	48
3.8 Definição da Dimensão da Imersão m	49
3.8.1 Falsos Vizinhos Mais Próximos	50
3.9 Conclusões.....	52
CAPÍTULO 4	53
APLICAÇÕES DO TEOREMA DE TAKENS	53
4.1 Introdução.....	53
4.2 Série Simulada através de Processo Auto-regressivo.....	55
4.2.1 Ajuste de uma Série de Carga Usando um Modelo AR(2)	56
4.2.2 Reconstrução do Espaço de Imersão Série AR(2)	56
4.2.3 Análise dos Resultados da Reconstrução Série AR(2)	58
4.2.4 Reconstrução do Espaço de Imersão da Série AR(2) Estacionada. .	59

4.2.5	Análise dos Resultados da Reconstrução Série AR(2) Estacionada.	60
4.3	Sistema de Lorenz	61
4.3.1	Reconstrução do Espaço de Imersão do Sistema de Lorenz.....	62
4.3.2	Análise dos Resultados da Reconstrução do Sistema de Lorenz.....	63
4.4	Conclusões.....	64
CAPÍTULO 5		66
APLICAÇÕES DA TEORIA DE <i>WAVELETS</i>		66
5.1	Introdução.....	66
5.2	Premissas Utilizadas na Aplicação de Wavelets	66
5.3	Exemplo 1: Análise de um Sinal Elétrico.....	67
5.4	Exemplo 2: Extração de Ruído de um Sinal.....	75
5.5	Conclusões.....	80
CAPÍTULO 6		81
APLICAÇÕES e RESULTADOS.....		81
6.1	Introdução.....	81
6.2	Características das Redes Utilizadas	81
6.3	Estatísticas Utilizadas	83
6.4	Diminuição das Componentes de Tendência e Sazonalidade	84
6.5	Minimização do Ruído Usando Wavelets	85
6.6	Procedimentos Utilizados para Reconstrução do Espaço de Imersão	85
6.7	Exemplo utilizando a série de carga EUNITE.....	88
6.7.1	Série de carga EUNITE.....	88
6.7.2	Análise dos dados.....	89
6.7.3	Reconstrução do Espaço Fase Série EUNITE.....	90
6.7.4	Previsão de Carga usando RNA.....	93
6.7.5	Reconstrução Espaço de Imersão Série EUNITE 1-2952.....	93
6.7.6	Reconstrução por Segmentos da Série EUNITE sem Processamento	95
6.7.7	Reconstrução de Segmentos da Série EUNITE com Processamento.	95
6.7.8	Qualidade da Reconstrução para Série EUNITE.....	96
6.8	Exemplo Usando Série Teórica com Resolução de 1 minuto	99
6.8.1	Reconstrução do Espaço de Imersão	100
6.8.2	Avaliação da Reconstrução via Previsão	101
6.9	Exemplo Usando Série PUGET	103
6.9.1	PUGET: Exemplo 1.....	104
6.9.2	PUGET: Exemplo 2.....	105
6.9.3	PUGET: Exemplo 3.....	105
6.9.4	PUGET: Exemplo 4 - Influência da Temperatura.....	106
6.10	Exemplo Usando Série de Carga Nacional com Resolução Horária.....	107
6.10.1	Influência do Processamento da Série na Reconstrução.....	108
6.10.2	Influência da Resolução da Série na Reconstrução do Espaço de Imersão	111
6.10.3	Influência da Aplicação de <i>Wavelets</i> na Reconstrução do Espaço de Imersão	112
6.10.4	Influência do Processamento de T_t e S_t na série + <i>Wavelets</i> na Reconstrução do Espaço de Imersão	113
6.11	Comparação Resultados Competição EUNITE 2001.....	115
6.11.1	Resultados Utilizando Amostra de Inverno (W).....	115
6.11.1.1	Série Amostra de Inverno (W) sem processamento T e S.....	116

6.11.1.2	Série Amostra de Inverno (W) com processamento T e S	117
6.11.2	Resultados Utilizando Amostra de Janeiro-Fevereiro (JF)	120
6.11.3	Comparação de Resultados	122
6.12	Conclusões.....	123
CAPÍTULO 7	125
CONCLUSÕES	125
7.1	Trabalhos futuros.....	128
ANEXO A	129
GLOSSÁRIO DE TERMOS	129
ANEXO B	133
MAPA LOGÍSTICO	133
B.1	Introdução	133
B.2	Para valores de $r < 1$	133
B.3	Comportamento para $1 < r < 3$	134
B.4	Quando r é maior que 3.....	134
B.5	Atrator de N pontos	135
B.6	Diagrama de Bifurcação	135
B.7	Sensibilidade as condições Iniciais	136
ANEXO C	140
TENDÊNCIA e SAZONALIDADE: Conceitos Básicos.....		140
C.1	Introdução.....	140
C.2	Tendências	141
C.2.1	Tendência polinomial	141
C.2.2	Suavização	142
C.2.2.1	Médias Móveis	142
C.3	Sazonalidade.....	144
C.3.1	Sazonalidade Determinística – Método de Regressão	147
C.3.2	Sazonalidade Estocástica – Método de Médias Móveis.....	149
INTRODUÇÃO À TEORIA DE WAVELETS	151
D.1	Introdução.....	151
D.2	Análise de Fourier	151
D.2.1	Análise de Fourier de Curta Duração	151
D.3	Análise de Wavelets	152
D.3.1	O que é possível realizar com <i>wavelets</i> ?.....	153
D.3.2	Definição de <i>wavelets</i>	154
D.4	A Transformada Contínua de Wavelets (CWT).....	155
D.4.1	Escalamento	156
D.4.2	Defasagem	157
D.4.3	Procedimento para Transformação de <i>Wavelets</i> Contínua	157
D.4.4	Escala e Freqüência	159
D.4.5	O que é contínuo na Transformada Contínua de <i>Wavelet</i> CWT ?	159
D.5	A Transformada Discreta de Wavelet - DWT.....	160
D.5.1	Filtragem de um Estágio: Aproximação e Detalhes	160
D.5.2	Decomposição de vários níveis	163
D.5.3	Reconstrução de <i>wavelets</i>	163
D.5.4	Reconstrução de Filtros	164
D.5.5	Reconstrução de Aproximação e Detalhes	165
BIBLIOGRAFIA	168

CAPÍTULO 1

INTRODUÇÃO

Uma informação fundamental para a operação e planejamento de um sistema elétrico é a previsão de carga em seus diferentes horizontes. Particularmente as previsões de carga de curto prazo aliadas a informações de disponibilidade de unidades geradoras, cenários de armazenamento e afluências hidrológicas, orientam o operador a definir o despacho ótimo de forma a atender a carga do sistema.

A literatura técnica apresenta uma grande variedade de metodologias que podem ser empregadas na previsão de curto prazo de consumo de energia elétrica. Esta variedade de técnicas reflete a busca por métodos capazes de gerar previsões mais precisas, o que propicia um melhor uso da capacidade instalada do sistema elétrico, entre outras ações. O impacto que a atividade de previsão de carga causa, não somente na operação, mas também no planejamento e na comercialização de energia, tem incrementado estudos no sentido de aperfeiçoar as ferramentas teóricas disponíveis [1].

Além da escolha da metodologia de previsão, a qualidade dos resultados na projeção de carga está fortemente associada à correta escolha das variáveis de entrada. Se tais variáveis não apresentam dependência com a variável de carga, a capacidade de ajuste e previsão do modelo fica comprometida.

Neste trabalho é modelado o comportamento futuro de um sistema baseado no seu desempenho passado, adicionando, eventualmente, alguns sinais exógenos de controle com aplicação em séries temporais. O desempenho da previsão depende, principalmente, da utilização de um vetor de regressão adequado, que contenha as variáveis explicativas mais relevantes. Com entradas insuficientes o modelo não tem informação necessária, sendo inabilitado para capturar os perímetros de trabalho do sistema, resultando assim, em altos erros de aproximação. Por outro lado, um modelo com entradas ou variáveis explicativas irrelevantes tende a ficar sobre-parametrizado, o que usualmente provoca um desempenho fraco do ponto de vista de previsão, além de aumentar o custo computacional.

Neste trabalho é apresentada uma metodologia de extração de variáveis relevantes para a previsão de carga a partir de informações contidas nas séries temporais. Em particular deseja-se observar “o que os dados têm a dizer” sobre a dinâmica geradora da série, sendo o objetivo final do trabalho a utilização das variáveis

extraídas em previsores neurais de carga. A técnica de extração apresentada está baseada em métodos originados na física particularmente na Teoria de Sistemas não Lineares e Teoria do Caos. Além disso é utilizada uma técnica oriunda da Teoria da Informação. É investigado ainda o impacto do pré-processamento das séries temporais no sucesso das técnicas de extração de variáveis relevantes, filtrando-se as componentes de tendência e sazonalidade das séries de carga com a finalidade de tornar “mais estacionárias” as séries de carga. Também é aplicada A Teoria de *Wavelets* com o objetivo de minimizar as componentes de ruído para posteriormente alimentar técnicas de extração de variáveis relevantes.

1.1 Previsão de Carga

A previsão de carga tem sido, recentemente, uma das áreas com maior pesquisa em Engenharia Elétrica. Neste esforço a maioria das técnicas têm sido baseadas na estatística tradicional ou nas chamadas “técnicas inteligentes” [1]. As ferramentas estatísticas mais usadas são: modelos de regressão linear univariados ou multivariados, modelos de alisamento exponencial, modelos baseados em filtros de Kalman, modelos Box Jenkins e suas variações (AR, MA, ARMA, ARIMA, SARIMA), etc [2]. No grupo das técnicas inteligentes, especial destaque merecem as Redes Neurais Artificiais (RNA), tema que surge em centenas de artigos nos últimos anos. Outras técnicas, como sistemas *fuzzy* e sistemas especialistas, também têm mantido o interesse dos pesquisadores. Entretanto, quando o assunto é previsão, estas outras técnicas geralmente aparecem como complemento das RNA[1].

A previsão de carga é comumente classificada em três horizontes de previsão: curto, médio e longo prazo. Cada uma destas previsões tem importância para as empresas elétricas: a previsão de curto prazo é usada para determinar o despacho de carga; a previsão de médio prazo é necessária para programar a manutenção e diversos intercâmbios de carga e a previsão de carga de longo prazo é útil para embasar análises financeiras e de planejamento do sistema elétrico. A previsão de carga é uma atividade complexa que tem um grau notório de incerteza, sendo que o trabalho do analista de previsão consiste em selecionar a melhor ferramenta de previsão, coletar os dados requeridos, processar os dados, construir um modelo matemático e finalmente gerar e validar a previsão[3].

A demanda de energia elétrica varia no tempo e é afetada por diversos fatores, quais sejam sociais, econômicos, meteorológicos, entre outros. Como na previsão de

curto prazo o horizonte é relativamente pequeno, as variáveis sociais e econômicas não têm influência, adquirindo o fator clima uma maior importância. É importante ressaltar que o padrão de carga dos dias úteis difere dos dias do final de semana, como também o padrão de carga dos feriados é diferente dos dias normais. O pico de carga normalmente coincide com os valores extremos de temperatura, dependendo se a energia é usada com o objetivo de esfriar ou aquecer o ar. Em países com estações do ano bem definidas o consumo mensal de energia tende a seguir o perfil da temperatura. Na Figura 1.1 podem ser observados os dados de temperatura e carga média (mensal) de uma distribuidora do hemisfério Norte. Como se trata de uma região fria com uma média anual de temperatura de aproximadamente 8 graus celsius, percebe-se claramente que a carga tem seus maiores valores nos períodos mais frios, provavelmente devido a aparelhos de calefação elétricos.

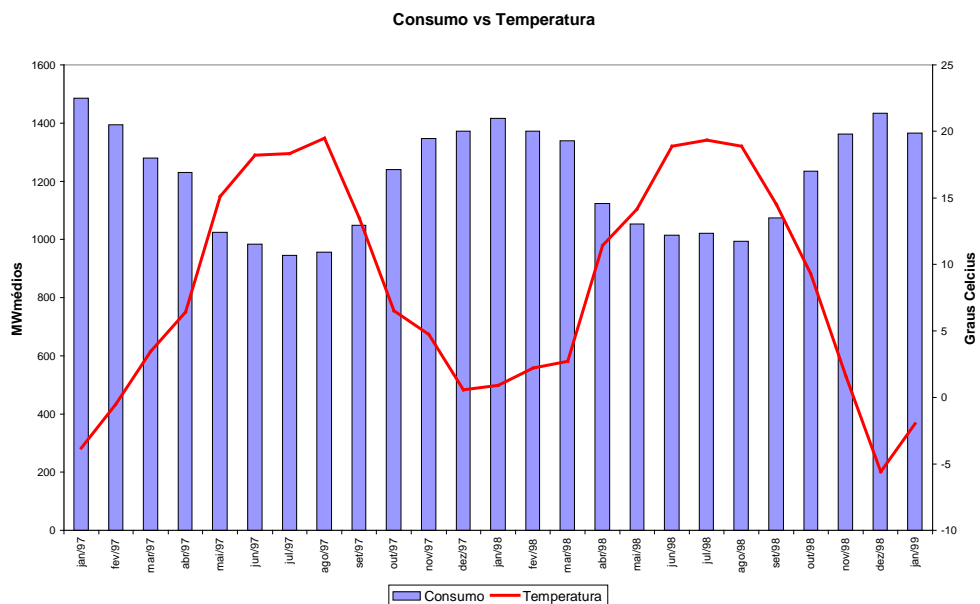


Figura 1.1 Evolução das variáveis carga e temperatura média em base mensal jan/97-jan/99.

Como já mencionado antes, a previsão de médio prazo é necessária para o gerenciamento de combustíveis, manutenção de unidades e diversos intercâmbios de carga. O horizonte de previsão de médio prazo geralmente é definido entre um e cinco anos, onde são previstas as variáveis demanda e energia. A previsão do pico de carga mensal é necessária para a manutenção de unidades e para intercâmbios, já a previsão de consumo é útil para gerenciamento de combustíveis para as empresas que possuem unidades de geração ou para as distribuidoras que precisam definir seus *portfolios* de contratos de fornecimento com os geradores.

A previsão de longo prazo, também conhecida como previsão anual de demanda e consumo, é de extrema importância para o planejamento da expansão da transmissão e geração, sendo utilizada para estudos de interconexão, requerimentos de combustíveis de longo prazo e planejamento tarifário. A definição do horizonte de previsão de longo prazo geralmente varia entre 5 e 25 anos[3].

Neste trabalho, o foco de interesse está na previsão de carga de curto prazo, particularmente o horizonte de previsão a ser abordado é de 1-24 horas à frente.

1.2 Redes Neurais

Como anteriormente citado, a popularidade de métodos alternativos à modelagem estatística tradicional continua aumentando. Alguns modelos não paramétricos, isto é, que não utilizam premissas sobre as formas paramétricas das relações funcionais entre as variáveis, têm atualmente fácil aplicação devido aos avanços na ciência da computação[4]. Uma classe de modelos conhecido como modelos de formas funcionais flexíveis oferece uma alternativa, mantendo também indefinida a forma das relações funcionais. Embora estes modelos contenham parâmetros, um grande número deles não são globalmente identificados, ou usando a terminologia comum, estimáveis. A identificação, quando alcançada no melhor dos casos, é local. Além disso, os parâmetros não são interpretáveis como freqüentemente acontece nos modelos paramétricos. As RNAs são um proeminente exemplo de tais formas funcionais flexíveis, com aplicações em inúmeros campos [4]. As RNA são sistemas paralelos distribuídos, compostos por unidades de processamento simples (neurônios) dispostas em uma ou mais camadas que são interligadas por um grande número de conexões geralmente uni-direcionais e com pesos para ponderar a entrada recebida por cada neurônio.

O uso de RNA é geralmente motivado pela premissa de que, sob determinadas condições, uma RNA simples pode aproximar qualquer função não linear com um determinado nível de precisão. Tal aproximador contém um número finito de parâmetros. Como especificar tal modelo, o seu significado e, como encontrar a combinação ideal de parâmetros e variáveis constitui um dos tópicos centrais na literatura de RNA [4].

1.3 Previsão de Carga de Curto Prazo Usando Redes Neurais

Existe uma ampla literatura sobre previsão de carga de curto prazo usando RNA. Em [1] é realizada uma abrangente análise sobre o estado da arte do assunto em questão. São avaliados mais de uma centena de artigos e comparados entre outros tópicos: arquitetura das redes usadas (número de neurônios, funções de ativação, regras de parada), entradas utilizadas, número de parâmetros, tamanho das amostras (treinamento e testes), algoritmos de treinamento, entre outros.

Os trabalhos avaliados utilizam a maioria dos modelos existentes com o objetivo de realizar previsões de carga a curto prazo. As técnicas estatísticas tradicionais, sejam elas modelos de regressão lineares, modelos dinâmicos lineares e não lineares, modelos B&J, Holt Winters. etc, são testadas e comparadas com o desempenho das RNA, que apresentam geralmente melhor desempenho. As redes avaliadas, utilizam em sua grande maioria variáveis de temperatura, umidade do ar, velocidade do vento, carga histórica, além de indicadores de sazonalidade como variáveis de entrada (dia da semana, mês do ano, etc) [1].

Embora exista uma grande quantidade de artigos escritos sobre previsão de demanda a curto prazo usando RNA, é fácil verificar que muitos são repetitivos, aplicando as mesmas técnicas, e em geral, relatando bons resultados principalmente quando comparados com técnicas tradicionais [1]. As principais dificuldades, quando são usadas RNA, passam pela escolha ótima da arquitetura do modelo, pela seleção das variáveis relevantes, pela filtragem dos dados de entrada, pelos problemas de *overfitting* e falta de generalização (problemas de aprendizado).

Nesta tese, o principal interesse encontra-se na determinação adequada das variáveis de entrada do modelo neural, de forma a aumentar a qualidade de previsão do modelo. Esta escolha não é trivial e pode ser abordada com algumas ferramentas disponíveis na literatura.

1.4 Extração de Variáveis Relevantes

A extração das variáveis de entradas mais relevantes é um importante pré-requisito para aplicação bem sucedida de uma RNA ou de qualquer modelo que tenha como objetivo realizar previsão de carga. Sua importância fica evidente quando se tem em mente que uma RNA é treinada para identificar relações não lineares entre pares de dados de entrada e saída. Se estes não são suficientemente representativos do problema a ser modelado a capacidade de generalização da RNA fica comprometida. Além disso,

se variáveis não representativas do problema forem consideradas, os tempos de treinamento podem ser prolongados e resultados de qualidade inferior podem ser obtidos.

Considere um problema de modelagem de séries temporais padrão: uma seqüência de medições $y(t)$ com $1 \leq t \leq T$ é coletada. Deseja-se prever $y(t)$ utilizando um conjunto de valores passados $y(t-d)$, com $d > 0$. A extração das variáveis relevantes consiste em encontrar um conjunto de m atrasos $(d_1, d_2, d_3, \dots, d_m)$ tal que o vetor de entradas $x(t) = [y(t-d_1) y(t-d_2) \dots y(t-d_m)]$ produza a melhor previsão de $y(t)$.

1.4.1 Abordagem Estatística

Do ponto de vista estatístico, a extração dos atrasos relevantes é um caso especial da seleção de características, a que por sua vez, é uma parte de um problema maior de análise da estrutura dos dados [6]. O método estatístico baseia-se na especificação de modelos f (paramétricos ou não paramétricos) com os quais se tenta realizar o mapeamento de entrada e saída $\hat{y}(t) = f(x(t))$. Na seleção de características convencional, uma importante premissa é a disponibilidade de todas as variáveis necessárias. Contanto que os dados sejam amostrados corretamente, esta premissa é usualmente satisfeita no caso das séries temporais. Os atrasos são inteiros positivos $d_j > 0$, $1 \leq j \leq m$, variáveis explicativas ou características são valores passados da série temporal $y(t-d_j)$ e entradas são os vetores $x(t)$ contendo esses valores passados. Encontrar as variáveis relevantes equivale a encontrar as entradas que produzem o desempenho de previsão ótimo. A seleção de características convencionais baseia-se em três diferentes componentes [7]:

1. um método de seleção de busca através do subconjunto de possíveis variáveis;
2. um critério de avaliação para calcular a qualidade de cada subconjunto;
3. uma condição de parada, que decide se foi obtido um subconjunto satisfatório.

Sabe-se que o único método ótimo para selecionar as melhores características entre um determinado conjunto F é a realização de uma busca exaustiva através de $2^F - 1$ possíveis subconjunto, de variáveis. Entretanto, essa metodologia ótima já torna-

se inaplicável para valores relativamente pequenos de F . Além disso, na modelagem temporal, o máximo atraso, e conseqüentemente o número total de variáveis F , não é conhecido de antemão, o que pode levar a experimentar valores de atraso demasiadamente elevados. Como critério de avaliação monótono, o algoritmo *branch and bound* proporciona uma alternativa eficiente, desafortunadamente, a habilidade de previsão não é um critério monótono [8]. Alternativas sub-ótimas realizam uma busca incrementando ou diminuindo regularmente o número de características selecionadas [6]. Alguns exemplos metodológicos são citados em [7]; destacando-se os abaixo elencados, a saber:

- Seleção para frente: começa com um conjunto vazio (sem variáveis), adiciona variáveis de acordo com o critério de avaliação até que o critério de parada seja alcançado.
- Eliminação para trás: começa com conjunto cheio (todas as possíveis variáveis), elimina uma variável de acordo com critério de avaliação até que o critério de parada seja alcançado.
- *Stepwise regression*: alternativa entre ambos métodos, por exemplo, realiza uma eliminação após cada inclusão ou escolhe entre adicionar ou eliminar variáveis de acordo com a evolução dos critérios de avaliação.

Observe que nestes casos o foco está no uso de redes neurais, quase que inteiramente através de diversas metodologias de poda [9]. Todos esses métodos sub-ótimos estão fortemente baseados em critérios de avaliação. Opções típicas incluem por exemplo, a estatística F [6], extensões a modelos não lineares, ou informação mútua [10].

1.4.2 Abordagem Via Generalização do Erro

No contexto de modelos não paramétricos, o objetivo é obter a melhor previsão. Assim, o critério de avaliação usado precisa ser uma medida que quantifique a habilidade de realizar uma previsão [7]. Para um determinado modelo f , que mapeia um vetor de entrada $x(t) = [y(t-d_1)y(t-d_2).....y(t-d_m)]$ e uma saída $y(t)$, o risco $\ell(y, f(x))$ mede o custo associado com saída a y . Assumindo um ruído gaussiano na saída usualmente opta-se por uma função de risco quadrático, $\ell(y, f(x)) = (y - f(x))^2$. A generalização do erro (ou risco esperado) associado com o modelo f é definido como

o valor esperado do risco sobre a desconhecida, embora fixa, distribuição de entrada-saída.

$$G(f) = \int (y - f(x))^2 p(x, y) dy dx \quad (1.1)$$

Este é conhecido também como erro de previsão ou erro quadrático integrado. Idealmente, o critério de avaliação para um dado subconjunto é a generalização do erro do modelo usando o dito subconjunto de dados como entrada [11]. No entanto, não pode ser usado diretamente devido a distribuição conjunta de entrada-saída ser desconhecida. Podendo ser estimada a partir dos dados disponíveis $(x(t), y(t))$, com $t \in D = \{T_{\min}, \dots, T_{\max}\}$.

Em [11] o estimador amostral SS (*split-sample*) é obtido substituindo a distribuição conjunta $p(x, y)$ pelo estimador empírico através do conjunto de validação $V \subset D$

$$\hat{G}_{SS} = \frac{1}{|V|} \sum_{t \in V} (y(t) - f^{-V}(x(t)))^2 \quad (1.2)$$

onde $|V|$ é a cardinalidade de V , e f^{-V} é o modelo estimado no conjunto de treinamento $T = D/V$. Note que T e V devem ser não conjunta visando que \hat{G}_{SS} seja uma estimativa não tendenciosa de $G(f^{-V})$.

Em [7] é proposto um método que se relaciona com técnicas iterativas de seleção de características, usados na estatística tradicional [6]. O método é construído acima de especificidades do processamento temporal de forma a proporcionar uma maneira original de seleção de variáveis de entrada. A relevância do atraso candidato é calculada diretamente pelo efeito que causa no desempenho da previsão, o qual é medido em função de estimadores da generalização do erro. O método em questão foi batizado como ERD (*Extraction of the Relevant Delays*) e trata-se de uma evolução na estimativa do erro generalizado acima apresentado.

1.4.3 Outras Abordagens

Diferentes métodos para identificação de variáveis de entrada de RNA têm sido utilizados, sendo que a maioria deles conjuga usualmente técnicas de análise de correlação com heurísticas e experiência do operador [12]. A autocorrelação linear é o mais comum entre os métodos de extração de variáveis, permitindo a captura de informação de segunda ordem entre duas variáveis. A autocorrelação pode ser definida

como a correlação entre membros de séries de observações ordenadas no tempo (como em uma série temporal) ou no espaço (dados de corte) [13]. Uma outra técnica utilizada, baseada numa transformação linear, é conhecida como método das componentes principais, cujo objetivo básico é a obtenção de um pequeno número de combinações lineares (componentes principais) de um conjunto de variáveis, que retenham o máximo possível da informação contida nas variáveis originais. Assim, um pequeno número de componentes é usado em lugar das variáveis originais nas análises de regressões, análises de agrupamentos etc [14].

Em [15] é realizado um estudo comparativo do desempenho entre duas técnicas usadas para seleção de variáveis de entrada de redes neurais. A primeira delas utiliza a função de autocorrelação na identificação das variáveis de entrada mais relevantes. Já a segunda vale-se de uma metodologia não linear baseada no método das coordenadas em atraso. Os resultados apresentados pelos autores indicam uma superioridade da técnica de extração baseada em autocorrelação linear.

Em [12] é apresentado um resumo das diferentes combinações de variáveis de entradas presentes na literatura e relatada a considerável variedade de opções que já foram utilizadas por diversos autores. No mesmo artigo conclui-se que não existe uma metodologia consistente para a seleção das variáveis relevantes de entrada. Também em [12] é proposto um método baseado no cálculo da deformação local integral de trajetórias no Espaço Fase para descobrir relações não lineares entre as variáveis candidatas, com o intuito de selecionar as variáveis de entrada mais relevantes.

Em [16] é apresentada uma técnica de seleção de variáveis para modelos não lineares, que está baseada em uma aproximação polinomial de modelos não lineares. Realizar a seleção implica em repetir a aplicação de mínimos quadrados de modelos que têm parâmetros lineares.

1.5 Extração de Variáveis a partir de uma Série Histórica

Em termos gerais, pode-se dizer que a previsão de fenômenos futuros está presente na maioria das atividades acadêmicas e empresariais. O principal objetivo é obter capacidade de antecipação que forneça embasamento para a adoção de estratégias corretas, seja no âmbito comercial ou no âmbito operacional. Obviamente, a capacidade de previsão está associada à complexidade do fenômeno que se deseja estudar.

No mundo real, a maioria dos fenômenos apresenta características dinâmicas, estocásticas, não lineares e principalmente dependem de N variáveis, que interagem entre si, provocando diversas consequências no fenômeno de interesse.

Em geral não é possível obter medições de todas as variáveis independentes de um processo. De fato, na maioria das situações reais, a única informação direta que se possui sobre o funcionamento do sistema estudado é sua resposta, isto é, uma série temporal em um determinado período de tempo.

Nesse contexto, torna-se importante a análise de sistemas dinâmicos sem que se conheçam a “física”, ou os fundamentos de sua dinâmica, inexistindo portanto um modelo matemático. Basicamente, deseja-se ver “o que os dados têm a dizer sobre a dinâmica geradora da série”.

Nos sistemas elétricos, o fenômeno que define a evolução da carga é dinâmico, não linear e multivariado. Em termos gerais, algumas das variáveis que influenciam o comportamento da carga elétrica são: as climatológicas, como temperatura, precipitação, umidade do ar, velocidade do vento e as variáveis econômicas, como renda, PIB, taxa de juros, planos econômicos e sociais. etc. Cada uma dessas variáveis, tem algum grau de responsabilidade no resultado final do fenômeno. Dessa forma, procura-se modelar a relação entre as diferentes variáveis que, no caso da carga, são escolhidas em função do horizonte de previsão que se deseja atingir. No entanto, como em qualquer sistema dinâmico multivariado, a única informação direta que se tem é uma série temporal de consumo ou demanda de energia elétrica.

A partir do estudo de sistemas dinâmicos multidimensionais com múltiplas entradas e somente uma saída, similares aos sistemas que controlam o consumo de energia elétrica, pode ser levantada a seguinte questão: É possível inferir as características do sistema original através do histórico das saídas para posteriormente prever seu comportamento futuro?

Takens [18] afirmou que é possível recriar um retrato equivalente da topologia do sistema multidimensional usando os valores observados de uma única variável de saída. A idéia básica é que o efeito de todas as outras variáveis não observáveis se refletem na variável observável de saída. As regras que governam o comportamento do sistema original podem ser reconstruídas a partir da saída, isto é, vetores em um novo espaço, o Espaço de Reconstrução ou Espaço de Imersão (*embedding space*), são formados através dos valores em atraso da série temporal observada. Trata-se de um enfoque físico da extração de variáveis relevantes de um sistema. No Anexo A é

apresentado um glossário de termos tendo como foco principal conceitos de dinâmica não linear e caos.

1.6 Método Físico de Extração de Variáveis a partir de uma Série Histórica

O método físico depende da definição da Dimensão da Imersão m das séries temporais [19]. Isto essencialmente equivale a encontrar um conjunto “primário” de atrasos, isto é, atrasos com uma influência explícita dos valores observados. Por exemplo, para séries temporais geradas por $x(t) = g(x(t-1))$, 1 é o único atraso “primário”.

Vários métodos tem sido propostos para estimar a Dimensão da Imersão m e o Espaço de Imersão (*embedding space*) de uma série temporal. Pi & Peterson [20] introduziram o “ δ -test” na literatura de redes neurais. He & Asada [21] propuseram o uso dos “Coeficientes de Lipschitz” para identificar a ordem de sistemas de entradas e saídas não lineares. Um método geométrico independente, no entanto similar ao anterior, foi aplicado na literatura de processamento de sinais [22]. Todos estes métodos partem da premissa que g é contínua e razoavelmente amortecida. A existência de um mapeamento contínuo entre $x(t)$ e $y(t)$ significa que entradas próximas $x(u)$ e $x(v)$ terão o mapeamento de suas saídas $y(u)$ e $y(v)$ próximas uma da outra. Por outro lado, para um espaço de entrada insuficiente (isto é, perda de algumas variáveis de entrada ou atrasos), entradas próximas podem trazer como consequência saídas arbitrariamente distantes. A quantificação desta não proximidade pode ser calculada considerando-se a probabilidade que duas saídas estejam próximas dado que suas entradas são próximas (δ -test) ou calculando a taxa entre as distâncias de entrada e saída (Coeficientes de Lipschitz). Uma importante característica das metodologias acima mencionadas reside no fato de que requerem alto esforço computacional [7]. Estas técnicas são não paramétricas, e estão baseadas somente nos dados, sendo desnecessário a existência de um modelo, isto é, a análise é realizada exclusivamente em cima de uma série temporal.

Como mencionado antes, a extração de variáveis do ponto de vista físico, tem como objetivo a reconstrução do espaço de estado ou Espaço de Imersão do sistema gerador do sinal temporal $x(t)$ que é definido por;

$$y(i) = \{ x(i), x(i-\tau), x(i-2\tau), \dots, x(i-(m-1)\tau) \} \quad (1.3)$$

Na prática significa a determinação de duas variáveis; Dimensão da Imersão- m e o Passo de reconstrução- τ (*time delay*). Para determinar estas variáveis, existem na literatura diversos esforços metodológicos mencionados a seguir e que são abordados mais profundamente nos capítulos seguintes.

A Informação Mútua defasada no tempo foi sugerida por Fraser e Swinney em [23] como ferramenta para determinar um atraso adequado τ . Diferentemente da função de autocorrelação, a Informação Mútua considera também correlações não lineares. A Informação Mútua captura informações de momentos superiores diferentemente da correlação linear, que somente captura informações de segunda ordem. O conceito batizado como Falsos Vizinhos Mais Próximos foi introduzido por Kennel, Brown e Abarbanel [24] e visa determinar o valor da Dimensão da Imersão- m . A idéia fundamental é procurar pontos no conjunto de dados que sejam vizinhos num determinado espaço de imersão, mas que poderiam não ser vizinhos em espaços de dimensões superiores, o que poderia acarretar uma evolução temporal futura muito diferente. Em outras palavras o que se procura são “falsos vizinhos”.

Outra metodologia para o cálculo da Dimensão da Imersão- m foi apresentada em [25] e é conhecida como Dimensão de Correlação, constituindo-se em um dos primeiros parâmetros utilizados para descrever atratores. Basicamente, fornece uma medida da complexidade do sistema em relação ao número de graus de liberdade do mesmo. Em [25] Grassberg e Procaccia propuseram um método definir m de maneira experimental.

1.7 Motivação

O motor do desenvolvimento deste trabalho está na procura de uma metodologia que reconstrua a dimensão de um sistema desconhecido, gerador de uma série de carga de energia elétrica. A obtenção da ordem deste sistema é feita utilizando informações contidas na própria série temporal de carga. A vantagem desta abordagem é que não depende da construção de um modelo matemático para a realização de previsões de carga de energia elétrica. Outra importante motivação está na utilização de ferramentas oriundas da teoria de sistemas não lineares, isto é, tratar do problema de identificação de variáveis de entrada “mais” relevantes (que é um problema com características não lineares) usando efetivamente ferramentas não lineares.

Embora as ferramentas aplicadas nesta tese já foram aplicadas pela primeira vez em [12] e posteriormente em [15], ditos trabalhos não esgotaram as possibilidades de análise, desta forma pode-se considerar esta tese como um aprofundamento dos

conceitos apresentados em ditos trabalhos. Uma motivação mais simples está vinculada com as dificuldades que o autor teve durante o desenvolvimento desta tese, particularmente em relação à oferta de literatura sobre os assuntos abordados. Como trata-se de uma tese multidisciplinar, os diversos assuntos não foram encontrados vinculados entre si, sendo assim o autor tenta neste documento unificar diversas metodologias da forma mais clara possível, fornecendo um documento que facilite o acesso principalmente aos leitores com formação em engenharia.

1.8 Objetivos da Tese

O objetivo da tese é investigar processos de extração de variáveis relevantes visando aperfeiçoar a previsão de carga de curto prazo em sistemas elétricos de distribuição de energia elétrica. O foco é concentrado na utilização de métodos físicos que permitam a extração de variáveis relevantes para seu uso em previsão de carga. Para isto são utilizadas exclusivamente, as informações contidas nas séries temporais de carga analisadas. São estudados, avaliados e comparados diversos métodos de reconstrução do Espaço Fase ou Espaço Estado fundamentados no Teorema de Takens (método da Dimensão de Correlação, método dos Falsos Vizinhos mais Próximos) além de métodos baseados na Teoria da Informação (método da Informação Mútua). Os objetivos complementares são citados a seguir:

- analisar se estes métodos propostos na literatura, vários dos quais tiveram seu origem na física (particularmente na teoria do caos), funcionam para reconstrução de sistemas não lineares geradores de séries de carga que possuem natureza estocástica não linear e multivariada. Em resumo, faz-se necessário avaliar a qualidade das previsões quando utilizadas as entradas fornecidas pelo métodos físicos;
- investigação da sensibilidade dos métodos físicos de extração de variáveis quando aplicadas variações na quantidade de dados amostrais (comprimento da série).
- investigação da sensibilidade em função da qualidade dos dados processados (exemplo: presença de ruído);
- investigação da sensibilidade dos métodos físicos de extração de variáveis quando são diminuídas ou eliminadas as características não estacionárias da série de carga,

- investigação da sensibilidade dos métodos físicos de extração de variáveis (do ponto de vista de previsão) quando se incrementa o grau de resolução dos dados amostrais da série (exemplo: dados horários, 30 minutos, 15 minutos, 5 minutos),
- análise da influência da aplicação da Teoria de *Wavelets* na aplicação dos métodos de extração de variáveis relevantes. Neste caso as séries são desagregadas nas suas componentes de aproximação e detalhes, para posterior aplicação dos métodos de extração para cada série desagregada. São avaliados os impactos em função da qualidade da previsão realizada;
- comparação dos métodos utilizados nesta tese, para calcular a Dimensão da Imersão (m), de forma a identificar qual fornece melhores resultados e em que condições;
- análise de sensibilidade do método de Informação Mútua em função da quantidade de dados usados e em função do grau de detalhes utilizado para realizar o cálculo;
- avaliação do impacto da variável temperatura quando incluída no processo de previsão,

1.9 Contribuições da tese

A seguir as principais contribuições desta tese:

- apresentação dos resultados da investigação, sobre a influência que a não estacionariedade das séries de carga, causam nos métodos de extração de variáveis utilizados;
- apresentação das conseqüências que o nível de resolução das séries de carga, causam no desempenho dos métodos de extração utilizados;
- apresentação dos impactos que a metodologia de *Wavelets* causa nos resultados obtidos pelos métodos de reconstrução aplicados nesta tese;
- apresentação de resultados sobre a sensibilidade dos métodos de extração de variáveis mais relevantes, quando da variação do tamanho amostral das séries de carga;

1.10 Etapas do trabalho

Esta tese começa com uma introdução aos conceitos de sistemas dinâmicos e sistemas caóticos, a qual é apresentada no capítulo 2. No mesmo capítulo são apresentados conceitos como caos, trajetórias, atratores, dimensão de um atrator, cálculo da dimensão de um atrator, expoente de Lyapunov, entre outros. Além disso é dada a definição do conceito de Dimensão da Correlação. O objetivo do capítulo 2 é fornecer subsídios para uma melhor compreensão dos conceitos de reconstrução levantados por Takens e outros autores, muitos deles físicos e matemáticos. Este trabalho continua no capítulo 3 com a definição do conceito de reconstrução de Espaço Estado, Espaço Fase ou Espaço de Imersão. Cabe ressaltar que esses conceitos são sinônimos, no entanto neste trabalho será adotado quase que exclusivamente o termo Espaço de Imersão. Neste capítulo é apresentado em detalhes o Teorema de Reconstrução de Espaço Fase de Takens e são abordados de forma profunda os métodos de extração de variáveis relevantes baseados nas informações contidas nas séries temporais; método dos Falsos Vizinhos mais Próximos, método da Dimensão da Correlação e método da Informação Mútua. Também são apresentados alguns exemplos de sistemas dinâmicos conhecidos visando ressaltar o conceito de reconstrução de forma didática e o impacto da correta escolha do Passo de Reconstrução. No capítulo 4 são apresentadas algumas aplicações do Teorema de Takens em sistemas cujos atratores tem dimensões previamente conhecidas. O objetivo é avaliar a confiabilidade da aplicação teorema e a coerência dos resultados obtidos.

Esta tese continua no capítulo 5 com aplicações da Teoria de *Wavelets*. Os aspectos de maior interesse que este capítulo proporciona estão relacionados com a possibilidade de diminuição ou eliminação do ruído presente na maioria das séries de consumo de energia elétrica. No capítulo 6 são apresentados os resultados da unificação e aplicação das ferramentas estudadas nos capítulos anteriores. São avaliados os métodos de reconstrução de Espaço e Imersão e a aplicação do Teorema de Takens em função da qualidade da previsão obtida quando utilizadas as entradas sugeridas pelos critérios de seleção de variáveis de entradas. Os testes de previsão são realizados utilizando como critério a qualidade na previsão avaliada em função de uma série de estatísticas como MAPE, Erro Máximo, RMSE, etc. No capítulo 7 são apresentadas as conclusões obtidas no desenvolvimento desta tese além de uma análise dos potenciais caminhos a serem seguidos para o aprofundamento dos conceitos estudados.

Nesta tese foram inseridos os seguintes anexos visando o esclarecimento ao leitor de conceitos não aprofundados no desenvolvimento do documento; por não ser o foco principal do trabalho ou não estimar adequada a incorporação no texto principal. Os anexos em questão são:

- Anexo A, com um glossário de termos associados principalmente a sistemas não lineares.
- Anexo B, onde são apresentadas as simulações realizadas com o mapa logístico sistema muito usado em diversas áreas principalmente em biologia, e embora matematicamente simples possui características que ajudam a ilustrar os conceitos sobre dinâmica apresentados no trabalho.
- Anexo C, onde são apresentadas as metodologias existentes para a eliminação ou diminuição das componentes de tendência e sazonalidade presentes em series temporais, algumas destas utilizadas no desenvolvimento deste trabalho,
- Anexo D, é apresentada uma introdução à Teoria de *Wavelets*, ferramenta utilizada no Capítulo 5. Neste Anexo são apresentados os conceitos de escalamento, defasamento, Transformada Contínua de *Wavelets*, Transformada Discreta de *Wavelets* e principalmente o conceito de decomposição em aproximação e detalhes.

Finalmente é apresentada uma completa bibliografia citando a literatura utilizada no desenvolvimento deste trabalho.

CAPÍTULO 2

CONCEITOS BÁSICOS DE SISTEMAS DINÂMICOS E CAOS

2.1 Introdução

Um sistema pode ser definido como um conjunto de objetos agrupados por alguma interação ou dependência, de modo que existam relações de causa e efeito nos fenômenos que ocorrem com os elementos desse conjunto. São exemplos de sistemas: o circuito elétrico de um telescópio, Júpiter e seus satélites, o sistema nervoso de um ser humano, o ecossistema de um mangue, o sistema que produz uma série de consumo de energia elétrica, etc. Um sistema é dinâmico quando algumas grandezas que caracterizam seus objetos constituintes variam no tempo, nos exemplos anteriores variam: a tensão entre as placas de um capacitor, a posição de um dos satélites, a atividade dos neurônios do córtex visual, o número de caranguejos fêmeas, a temperatura [26]. Determinar, teoricamente, a evolução temporal das grandezas que caracterizam um sistema pode ser importante nas situações citadas a seguir:

- sistemas que ainda não existem;
- explicação do comportamento de sistemas existentes;
- previsão do comportamento de sistemas existentes;
- impossibilidade de realizar testes experimentais.

O objetivo de cada um dos estudos teóricos citados acima é “prever o futuro” a partir de um ponto de vista teórico. Para tanto, é necessário conhecer como as coisas são e compreender as regras que governaram as mudanças que ocorreram no passado. Os sistemas dinâmicos não-lineares podem ser classificados em estocásticos e determinísticos, sendo os estocásticos aqueles que apresentam uma saída estocástica quando aplicada uma entrada estocástica. No entanto, os sistemas determinísticos também são passíveis de apresentar respostas caóticas. Modernamente, o caos é entendido como o comportamento estocástico de sistemas determinísticos.

Este capítulo apresenta uma revisão sobre as principais características de sistemas dinâmicos e caóticos: sensibilidade às condições iniciais, dimensão fractal, invariantes geométricos, que servem para caracterizar sistemas dinâmicos, etc. No entanto o conceito mais importante a ser apresentado neste capítulo será o atrator, e particularmente, a dimensão de um atrator e sua estimativa. A dimensão de um sistema

dinâmico está diretamente relacionado com o objetivo desta tese, que visa a partir de informações contidas exclusivamente numa série temporal de demanda de energia elétrica, reconstruir o Espaço de Imersão ou Espaço Fase do sistema que gerou dita série. O objetivo deste capítulo é fornecer alguns subsídios teóricos para os tópicos que serão levantados no capítulo 3. No Anexo A é incluído um glossário de termos que pode auxiliar no entendimento dos conceitos aqui apresentados.

2.2 Aspectos Históricos

Com o estudo de modelos lineares foram criados paradigmas que se enraizaram na tradição histórica. O determinismo estrito é um bom exemplo disto e pode ser bem representado pelo pensamento de Laplace: *“Deve-se ver o estado presente do universo como o efeito do seu estado anterior, e como a causa daquele que virá. Uma inteligência que, em qualquer instante dado, conhecesse todas as forças pela qual o mundo natural se move e a posição de cada uma de suas partes componentes, e que tivesse também a capacidade de submeter todos estes dados à análise matemática, poderia encompassar na mesma fórmula os movimentos dos maiores objetos do universo e aqueles dos menores átomos; nada seria incerto para ele, e o futuro, assim como o passado, estaria presente diante de seus olhos”* [17].

Em fins do século XIX Poincaré, enquanto estudava o problema da dinâmica de três corpos, chegou à conclusão que o acaso deveria se contrapor ao determinismo estrito de Laplace: *“Uma causa muito diminuta, que nos escapa, determina um efeito considerável, que não podemos deixar de ver, e então dizemos que este efeito é devido ao acaso. Se pudéssemos conhecer exatamente as leis da natureza e a situação do universo no instante inicial, seríamos capazes de prever exatamente a situação deste mesmo universo no instante seguinte. Mas mesmo quando as leis naturais já não tivessem mais segredo para nós, só poderíamos conhecer a situação inicial aproximadamente. Se isto nos permite antecipar a situação futura com o mesmo grau de aproximação, ficamos satisfeitos, dizemos que o fenômeno foi previsto, que é governado por leis. Mas nem sempre isto ocorre; pode acontecer que diferenças mínimas nas condições iniciais produzam diferenças muito grandes no fenômeno final; um erro mínimo nas condições iniciais produziria um erro enorme no resultado final do fenômeno. A previsão torna-se impossível e temos o fenômeno do acaso”* [17].

Apesar da clara visão de Poincaré, foi só em 1963, quando Lorenz [27] desenvolvia estudos sobre problemas atmosféricos, que se retomou a idéia do acaso na análise de sistemas dinâmicos. Contando com o auxílio de um computador, Lorenz tratava o modelo de Rayleigh-Bernard para convecção de fluidos e observou que uma pequena variação nas condições iniciais poderia acarretar grandes diferenças na evolução do sistema. Este fenômeno ficou conhecido como “efeito borboleta”, como uma alusão de que se uma borboleta batesse suas asas em algum lugar do planeta, poderia alterar a resposta de um sistema dinâmico do outro lado da Terra. Tratava-se de um sistema totalmente determinístico cujos resultados poderiam ser aleatórios.

Gleick [28] coloca que o trabalho de Lorenz trouxe a assustadora compreensão de que equações matemáticas simples podiam servir de modelo para sistemas muito violentos. Iniciava-se aí o moderno estudo do caos, cujas idéias básicas haviam sido lançadas por Poincaré. O caos é uma das inúmeras possibilidades de comportamento de um sistema não-linear, podendo-se dizer que, de alguma forma, se trata da dinâmica libertada da previsibilidade.

A idéia de que a natureza é bem comportada, que domina o senso comum, é uma conseqüência da exagerada utilização de modelos lineares, o que limitou o espectro de análise por muitos anos, eliminando uma incrível variedade de comportamentos.

2.3 Definição de Caos

Caos é definido como o comportamento aparentemente estocástico de alguns sistemas determinísticos. Cabe estabelecer a distinção entre fenômenos aleatórios e caóticos. Os fenômenos aleatórios dizem respeito a sistemas não-determinísticos. Ao reverso os fenômenos caóticos são determinísticos. Assim, num sistema caótico, para uma entrada totalmente conhecida e determinada surge uma resposta aparentemente aleatória.

Após Lorenz, inúmeros pesquisadores passaram a se debruçar sobre o caos, analisando diferentes sistemas dinâmicos associados a uma série de situações físicas. May [29], por exemplo, tratou um sistema dinâmico relacionado com o crescimento populacional de espécies de insetos. Este trabalho ficou conhecido como Mapa Logístico e avalia a população em um ano, a partir do ano anterior. Sem dúvida, trata-se de um sistema simples do ponto de vista matemático, mas que possui uma dinâmica muito rica. O Mapa Logístico é abordado com maiores detalhes no Anexo B.

Em sistemas dissipativos as trajetórias podem convergir para uma região limitada do Espaço Fase ou espaço de estado, chamada de atrator. Um sistema dinâmico, que apresenta comportamento caótico, possui um atrator estranho no seu Espaço Fase. Ele é estranho porque apresenta detalhes em escalas infinitesimalmente pequenas. Tal figura geométrica é chamada de fractal.

2.4 Bifurcações

Bifurcação refere-se à mudança qualitativa do retrato de fases de um sistema dinâmico quando algum parâmetro do sistema passa por um valor crítico. A idéia de bifurcação está intimamente ligada ao conceito de estabilidade estrutural. Um sistema dinâmico é estruturalmente estável se ele é orbital e topologicamente equivalente a uma versão perturbada dele mesmo. Se, no entanto, ao se variar o valor do parâmetro de controle em torno de um valor crítico ocorre uma mudança qualitativa no seu retrato de fases, então o sistema é estruturalmente instável para aquele valor crítico do parâmetro. Esta mudança na topologia do retrato de fases é chamada de bifurcação. A Figura 2.1 mostra o diagrama de órbitas do mapa logístico. Cabe ressaltar que no diagrama de órbitas de um sistema dinâmico são representadas somente as soluções atratoras (assintoticamente estáveis), o que lhe diferencia de um diagrama de bifurcação, no qual aparecem também as soluções instáveis.

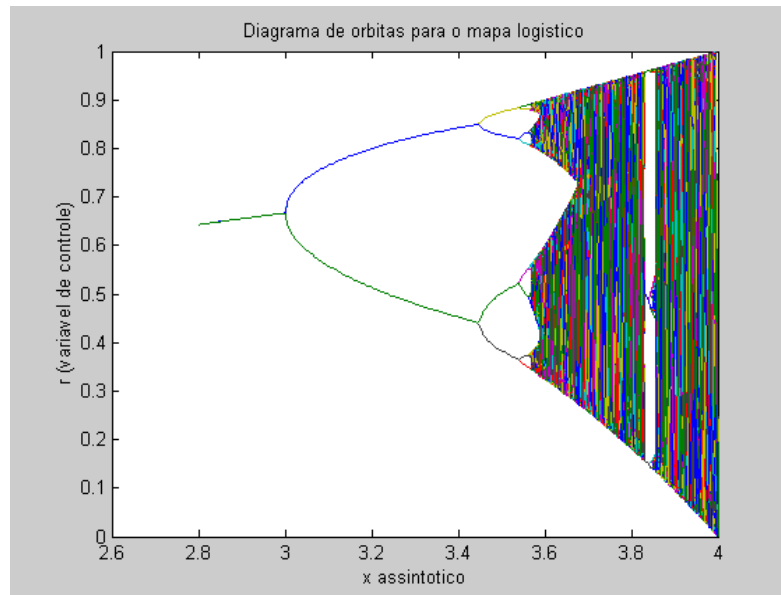


Figura 2.1. Diagrama de órbitas do mapa logístico

Na tabela 2.1, são apresentados os valores da variável de controle μ para os quais nascem as órbitas de período 2^k

Tabela 2.1. Valores de μ para os quais nascem as órbitas de período 2^k

k	$\mu(k)$	Nasce o período 2^k
1	3	2
2	3,449490	4
3	3,544090	8
4	3,564407	16
5	3,568759	32
6	3,569692	64
7	3,569891	128
8	3,569934	256
⋮	⋮	⋮
⋮	⋮	⋮
∞	3,569946	2^∞

2.5 Atrator

Simplificadamente, atrator pode ser definido como uma figura geométrica que representa o comportamento assintótico de um sistema dinâmico, num espaço formado pelas variáveis que caracterizam a evolução desse sistema. Um atrator possui uma bacia de atração que é definida pelo maior conjunto de pontos nesse espaço tais que, partindo-se de um desses pontos, tende-se assintoticamente para o atrator [26].

Um sistema dinâmico que apresenta comportamento caótico possui um atrator estranho no seu Espaço Fase, o que significa que o sistema possui alta sensibilidade às condições iniciais. A resolução de um modelo meteorológico bastante simplificado, por parte de Lorenz [27], realizada através de simulações num computador resultaram na descoberta do primeiro atrator estranho. Na Figura 2.2, é possível observar o atrator para o sistema de Lorenz, que representa a trajetória das órbitas do sistema dinâmico representado a través do seguinte sistema de equações diferenciais:

$$\begin{aligned}
 \frac{dx_1}{dt} &= -\sigma x_1 + \sigma x_2 \\
 \frac{dx_2}{dt} &= -x_1 x_3 + r x_1 - x_2 \\
 \frac{dx_3}{dt} &= x_1 x_2 - b x_3
 \end{aligned}
 \tag{2.1}$$

$(\sigma = 10, b = 8/3, r = 28)$

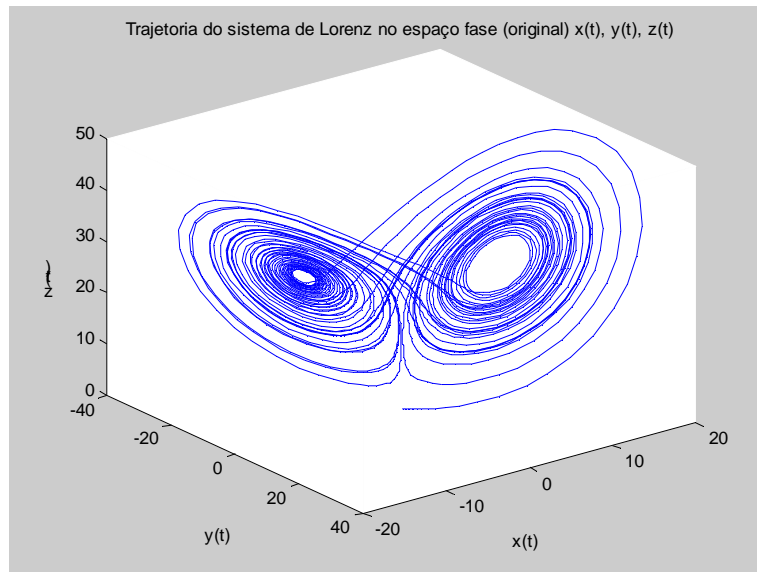


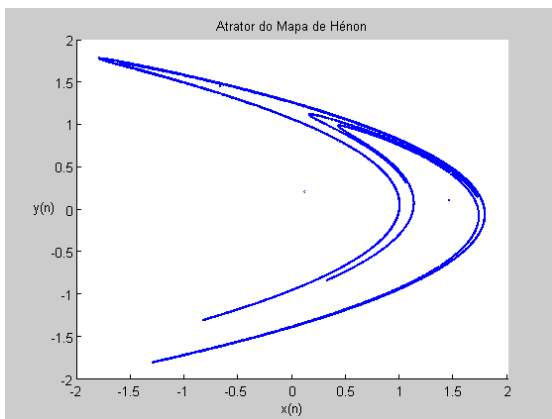
Figura 2.2. Atrator do Sistema de Lorenz.

Já na Figura 2.3 (a) é possível observar o atrator de um conhecido sistema dinâmico. Trata-se do mapa de Hénon, definido pelo seguinte sistema de equações:

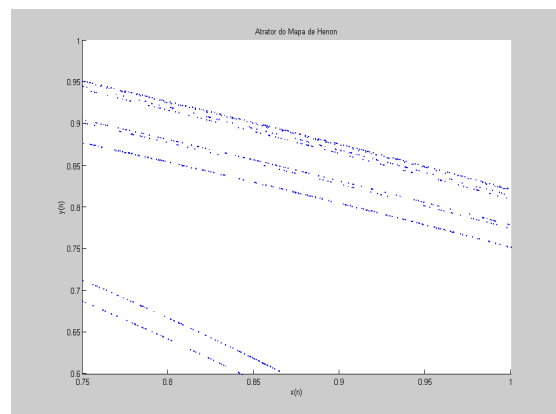
$$x(i+1) = 1 - a \cdot x^2(i) + y(i) \tag{2.2}$$

$$y(i+1) = b \cdot x(i)$$

$$a = 1,4 \text{ e } b = 0,3$$



(a)



(b)

Figura 2.3. Atrator do mapa de Hénon (a) e Atrator de Hénon amplificado (b)

2.6 Atrator: Definição Formal

Um conjunto de pontos A , no Espaço Fase de um sistema dinâmico, é definido como atrator se:

- A é um conjunto invariante; ou seja, qualquer trajetória $\bar{x}(t)$ que começa em A , permanece em A por todo o tempo;
- A atrai um conjunto aberto de condições iniciais; isto é, há um hiper-volume esférico B , que contém A , tal que para qualquer condição inicial $\bar{x}(0)$ pertencente a B , então a distância entre a trajetória $\bar{x}(t)$ e A tende a zero, quando $t \rightarrow \infty$. O maior conjunto de condições iniciais que satisfaz essa propriedade é chamado bacia de atração de A ;
- A é mínimo; ou seja, não há sub-conjunto de A que satisfaça as duas condições anteriores.

Num sistema dinâmico de tempo contínuo, autônomo e bidimensional, o teorema de Poincaré-Bendixson garante que a figura atratora pode ser de dois tipos, a saber:

- o ponto de equilíbrio, que corresponde a uma solução cujo comportamento dinâmico é independente do tempo;
- o ciclo limite, que descreve um comportamento periódico no tempo, com amplitude e período determinados pela forma das equações e pelos valores dos seus parâmetros.

Num sistema tridimensional, existem mais duas figuras atradoras, que são:

- a superfície toroidal, que representa um regime periódico ou quase-periódico, com 2 frequências fundamentais independentes;
- o atrator estranho, que apresenta dependência sensível às condições iniciais.

Um sistema dinâmico determinístico, cuja evolução temporal leva assintoticamente para um atrator estranho, apresenta dinâmica caótica. O nome atrator estranho foi introduzido por D. Ruelle e F. Takens, em 1971, num trabalho sobre o processo de transição de fluxo laminar para fluxo turbulento em fluidos [30].

Ruelle e Takens propuseram que um fluido com velocidade constante pode sofrer uma bifurcação e passar a ter seu movimento caracterizado por uma órbita periódica, de modo que a velocidade do fluido oscila com frequência bem definida.

Segundo esses autores, um atrator estranho pode surgir quando, com essa órbita, acontecem (algumas) sucessivas bifurcações, conforme se varia a magnitude de algum parâmetro. Em experimentos com fluidos, este parâmetro costuma a ser a diferença de temperatura ou diferença de velocidade entre as placas que o confinam. A rota para o caos sugerida por Ruelle e Takens ocorre via quase-periodicidade. Nessa rota, a órbita periódica sofre uma primeira bifurcação, sendo substituída por um toro atrator caracterizado por duas frequências, cuja razão é um número irracional. Frequências com esta propriedade são dita incomensuráveis. As trajetórias correspondentes repousam sobre uma superfície toroidal e jamais se fecham, descrevendo um movimento que é chamado de quase-periódico. Uma segunda bifurcação produz um novo toro atrator com três frequências incomensuráveis; e uma terceira leva a um atrator estranho. Para Ruelle e Takens, o início do comportamento turbulento coincide com a bifurcação que gera o atrator estranho.

Para que possa existir comportamento caótico em sistemas contínuos dissipativos, é necessário que ele seja não-linear e pelo menos tridimensional, como o sistema de Lorenz, apresentado na seção anterior. Em mapas dissipativos, há caos num sistema unidimensional, desde que ele seja não-invertível, como o mapa logístico, também apresentado em detalhes no Anexo B.

Caos advém de uma dependência sensível às condições iniciais. Diz-se que um mapa $x_{j+1} = F(x_j)$ depende sensivelmente das condições iniciais se há um número $\varepsilon > 0$, tal que para qualquer condição inicial x_0 e para “qualquer” número $\delta > 0$, existe pelo menos um ponto x'_0 com $|x'_0 - x_0| < \delta$, tal que $|F^{(k)}(x'_0) - F^{(k)}(x_0)| \geq \varepsilon$. Assim, não importa o valor de x_0 e nem a vizinhança de δ , pode-se sempre encontrar um ponto x'_0 nessa vizinhança cuja órbita se separa da órbita de x_0 de pelo menos ε . Além disso, essa distância ε independe de x_0 .

Note que essa definição de sensibilidade às condições iniciais não exige que a órbita de x'_0 afaste-se da órbita de x_0 em todas as iterações. Ela exige somente que o k -ésimo ponto da órbita de x_0 esteja afastado do k -ésimo ponto da órbita de x'_0 por uma distância ε .

Sobre a sensibilidade às condições iniciais Lorenz afirmou que o *pinball* está para o caos assim como o dado está para a probabilidade [27]. Diferenças mínimas nas condições iniciais do lançamento da bola podem acarretar diferenças muito grandes

depois de um certo tempo, inviabilizando certas jogadas. De maneira análoga, uma pequena variação no impulso de um esquiador no início de uma pista pode acarretar grandes diferenças em um determinado ponto da pista. A Figura 2.4 mostra a sensibilidade do mapa logístico (com $\mu = 3,99$) às condições iniciais para uma pequena diferença ($x_1=0.3$ comparada com $x_1=0.301$). Note que, para o mesmo sistema determinístico definido pela equação (2.3), uma variação de 0,001 nas condições iniciais (x_1) provoca trajetórias completamente diferentes. Verifica-se que as trajetórias começam a se deslocar aproximadamente, a partir do ciclo 10.

$$x(n+1) = \mu \cdot x(n) \cdot [1 - x(n)] \quad (2.3)$$

onde μ é a variável de controle do sistema.

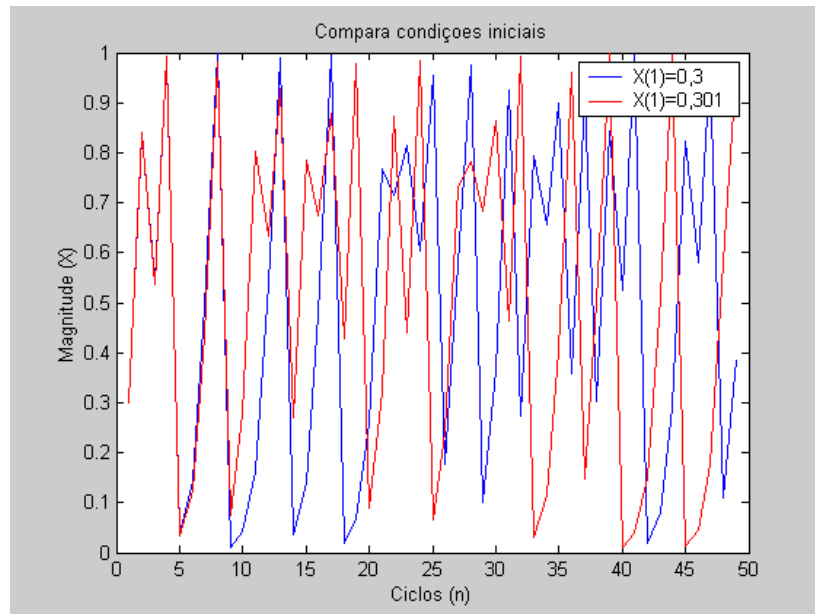


Figura 2.4. Duas series temporais para $\mu=3.99$, $x_1=0.3$ comparada com $x_1=0.301$

O estudo do caos e da dinâmica não-linear envolve uma série de ferramentas próprias. A definição dos invariantes geométricos é fundamental no diagnóstico de uma resposta caótica. A sensibilidade às condições iniciais, por exemplo é avaliada através do cálculo de um destes invariantes geométricos: os expoentes de Lyapunov.

2.7 Expoente de Lyapunov

Os expoentes de Lyapunov constituem-se em uma ferramenta que permite avaliar a sensibilidade às condições iniciais de um sistema de dinâmica não linear. Os expoentes de Lyapunov verificam a divergência exponencial no tempo de trajetórias vizinhas e representam um dos critérios mais importantes utilizados para definir o caos

em sistemas dinâmicos. O comportamento caótico é caracterizado pela existência de, pelo menos, um dos expoentes de Lyapunov positivo.

Os expoentes de Lyapunov descrevem como as trajetórias de um atrator se movem sob a evolução da dinâmica e é conhecida como uma característica invariante. O termo invariante significa que os expoentes de Lyapunov de um processo caótico permanecem inalterados sob modificações não lineares suaves do sistema de coordenadas do processo. O maior expoente de Lyapunov também define o horizonte de previsibilidade de um processo caótico. Especificamente, a previsibilidade de curto prazo de um processo caótico é aproximadamente igual ao inverso do maior expoente de Lyapunov [32].

A partir da determinação desses expoentes é possível avaliar outros invariantes do sistema, o que mostra a importância de se calcular precisamente os expoentes de Lyapunov. Quando o sistema dinâmico possui um modelo matemático estabelecido que permite a sua linearização em torno de uma determinada trajetória, os expoentes de Lyapunov podem ser calculados com precisão a partir do algoritmo proposto em [33] por Wolf et al. A determinação desses expoentes partindo de um sinal experimental é bem mais complicada e, basicamente, existem dois métodos disponíveis: o método das trajetórias ou método direto ou do espaço real e o método das perturbações ou dos espaços tangentes ou método da matriz Jacobiana.

Para introduzir a idéia geral da determinação dos expoentes de Lyapunov, considere uma trajetória do sistema dinâmico que evolui a partir de uma condição inicial e uma vizinhança desta trajetória, que pode ser definida por uma esfera de diâmetro d_0 . De cada ponto desta esfera parte uma nova trajetória e o conjunto de todas as trajetórias possíveis forma, em cada tempo t , esferas deformadas. Deseja-se avaliar como duas trajetórias, suficientemente próximas, divergem uma da outra na medida em que o sistema evolui. Em outras palavras, deseja-se avaliar como a esfera inicial se comporta na medida em que o tempo evolui, vide Figura 2.5.

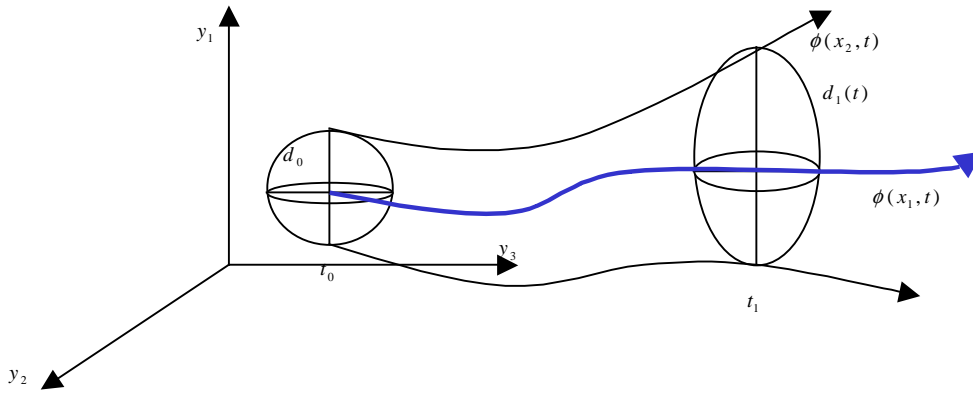


Figura 2.5. Representação gráfica da avaliação dos expoentes de Lyapunov.

A variação do diâmetro desta esfera pode ser expressa a partir da seguinte expressão: $d(t) = d_0 \cdot b^{\lambda t}$, onde b é uma base de referência. Se o expoente λ for negativo ou nulo as trajetórias não divergem. Se por outro lado, λ for positivo, indica que as trajetórias divergem, caracterizando o caos.

Algoritmos próprios devem ser considerados para avaliar os expoentes de Lyapunov [33], [34]. A divergência de uma trajetória caótica é localmente exponencial. Todavia, uma distância d desta trajetória não deve ir para infinito, uma vez que ela representa um sistema físico. Desta forma, para avaliar a média desta divergência das trajetórias, deve-se tomar uma média do crescimento exponencial em vários pontos sobre a trajetória. Assim, quando a distância $d(t)$ torna-se muito grande, define-se um novo $d_0(t)$ para reavaliar a divergência.

Com isso é possível identificar uma média capaz de medir a divergência, sendo os expoentes de Lyapunov definidos da seguinte forma:

$$\lambda = \frac{1}{t_n - t_0} \sum_{k=1}^n \log_b \left(\frac{d(t_k)}{d_0(t_{k-1})} \right) \quad (2.4)$$

A partir do espectro de Lyapunov é possível formular um procedimento que avalie a dimensão do atrator, considerando o espectro de Lyapunov ordenado de forma decrescente. A conjectura de Kaplan-Yorke estabelece que a dimensão de Lyapunov é dada por [33]:

$$D = j + \frac{\sum_{i=1}^j \lambda_i}{|\lambda_{j+1}|} \quad (2.5)$$

onde j é definido a partir das seguintes condições:

$$\sum_{i=1}^j \lambda_i > 0 \quad \text{e} \quad \sum_{i=1}^{j+1} \lambda_i < 0 \quad (2.6)$$

Como exemplo considere a Figura 2.6 onde é possível observar como varia o expoente de Lyapunov para o mapa logístico em função do parâmetro μ . Para esse mapa, $\ln|dF(x)/dx| = \ln|\mu - 2\mu x|$. Note que na região $\mu < \mu_\infty \cong 3,57$, na qual a solução atratora é ou um ponto fixo ou uma órbita de período 2^n , o expoente λ permanece negativo, aproximando-se de zero nos pontos de bifurcação. Comportamento caótico ocorre para $\mu < \mu_\infty$, pois nessa região o expoente λ apresenta valores positivos. Para $\mu < \mu_\infty \leq 4$, os valores negativos são causados pela presença de janelas de comportamento periódico.

Ressalte-se que o expoente λ depende da condição inicial x_0 escolhida. Esta dependência é minimizada tomando-se $N \rightarrow \infty$. Como, na prática, trabalha-se com N finito, calcula-se o expoente λ a partir de várias condições iniciais e para valores de N suficientemente grandes, de modo a tornar o valor do expoente λ praticamente independente da condição inicial. Em geral, costuma-se expressar λ como um valor médio, com um desvio padrão associado.

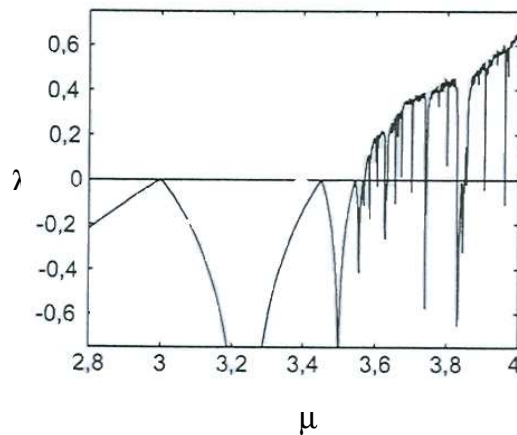


Figura 2.6: Expoentes de Lyapunov, para o mapa logístico, em função de μ .

2.8 Dimensões de um Atrator

A dimensão de um conjunto de pontos é o número mínimo de coordenadas necessárias para localizar cada ponto do conjunto. Por exemplo, uma curva é unidimensional, porque cada ponto é determinado por um único número, que é o comprimento do arco medido a partir de algum ponto de referencia fixo nessa curva.

Seja um conjunto de pontos A num Espaço Fase de dimensão n, este conjunto de pontos pode ser coberto por hiper-cubos iguais de lado r. Em [35], Kolmogorov definiu Dimensão de Contagem de Caixas (“Box-counting dimension”) D_0 como:

$$D_0 = \lim_{r \rightarrow 0} \frac{\log N(r)}{\log(1/r)} \quad (2.7)$$

$N(r)$ é o número mínimo de hiper-cubos idênticos de lado r necessário para cobrir todo o conjunto de pontos. Outro nome para D_0 é Dimensão de Capacidade.

A modo de exemplo pode-se usar a fórmula (2.7) para calcular a dimensão de um segmento de reta de comprimento L. Esse segmento de reta é coberto por uma “caixa” de comprimento L, ou por duas “caixas” de comprimento L/2, ou por quatro “caixas” de comprimento L/4, ou generalizando por 2^n “caixas” de comprimento L/2ⁿ, como mostra na Figura 2.7. Considerar $r \rightarrow 0$ equivale a tomar $n \rightarrow \infty$, de modo que:

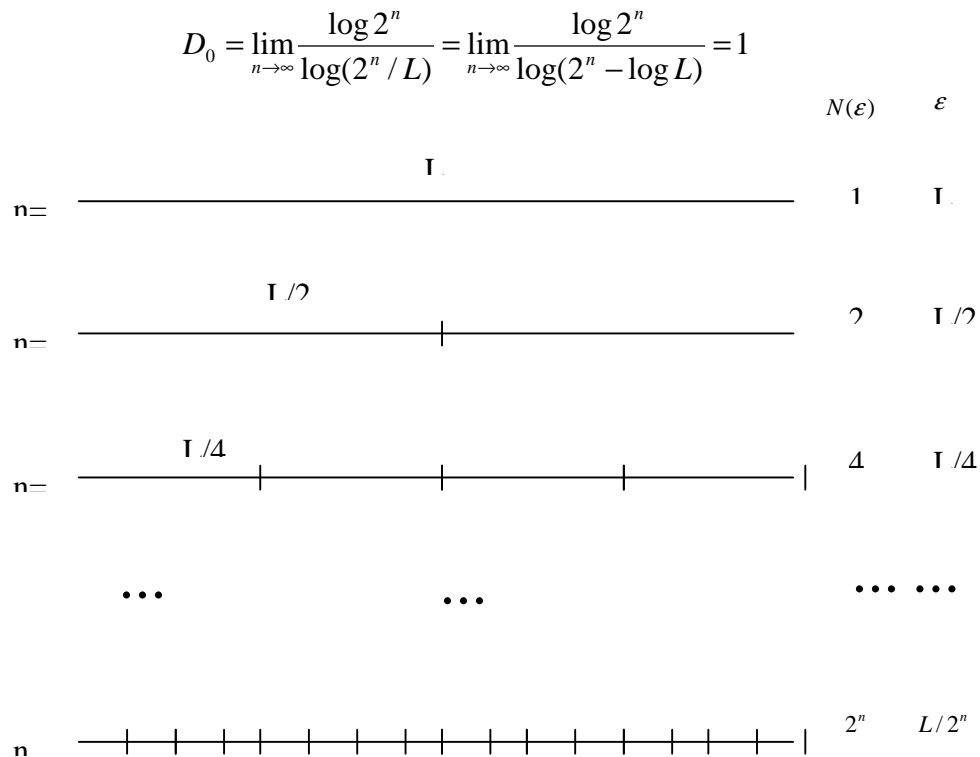


Figura 2.7. Cálculo de D_0 para um segmento de reta.

É claro que se obtém o mesmo resultado quando se considera que o segmento de reta pode ser coberto por uma “caixa” de comprimento L , ou por q “caixas” de comprimento L/q , ou por q^2 “caixas” de comprimento L/q^2 ou, generalizando, por q^n “caixas” de comprimento L/q^n . Agora:

$$D_0 = \lim_{n \rightarrow \infty} \frac{\log q^n}{\log(q^n / L)} = 1$$

Um segundo exemplo pode ser a aplicação da fórmula (2.7) para calcular a dimensão de um quadrado de área L^2 . Esse quadrado pode ser coberto por uma “caixa” de um comprimento L , ou por q^2 “caixas” de comprimento L/q , ou por q^4 “caixas” de comprimento L/q^2 , ou generalizando, por q^{2n} “caixas” de comprimento L/q^n , ou generalizando, por q^{2n} “caixas” de comprimento L/q^n . Nesse caso:

$$D_0 = \lim_{n \rightarrow \infty} \frac{\log q^{2n}}{\log(q^n / L)} = 2$$

Um terceiro exemplo é a utilização da fórmula (2.7) para calcular a dimensão do conjunto de Cantor. O conjunto imaginado, em 1833, por G. Cantor é construído da maneira descrita aqui. Tome o intervalo $S_0 = [0,1]$ e remova seu terço central. Isso produz o par de intervalos $[0,1/3]$ e $[2/3,1]$, como mostrado na Figura 2.8. Remova agora os terços centrais de S_1 , produzindo S_2 . Repita esse procedimento infinitas vezes. O conjunto limite S_∞ obtido é chamado conjunto de Cantor, sendo formado por infinitos pedaços infinitesimais separados por infinitos intervalos de vários tamanhos. Sua estrutura é auto-similar, pois contém cópia de si próprio em todas as escalas, conforme sugere a ilustração 2.8:

$$D_0 = \lim_{n \rightarrow \infty} \frac{\log q^{2n}}{\log(3^n / L)} = \frac{\log 2}{\log 3} \cong 0,63$$

Assim, o conjunto de Cantor é caracterizado por um valor de D_0 fracionário, ou seja, não-inteiro. Esse valor foi obtido, pela primeira vez, por F. Hausdorff em 1918.

Outro exemplo clássico é o triângulo de Sierpinski, que é construído a partir de um triângulo equilátero, do qual se remove o triângulo com vértices no ponto médio entre os lados. Repetindo esse processo indefinidamente, cria-se a figura imaginada por W. Sierpinski, que possui $D_0 = (\log 3)/(\log 2) \cong 1,58$.

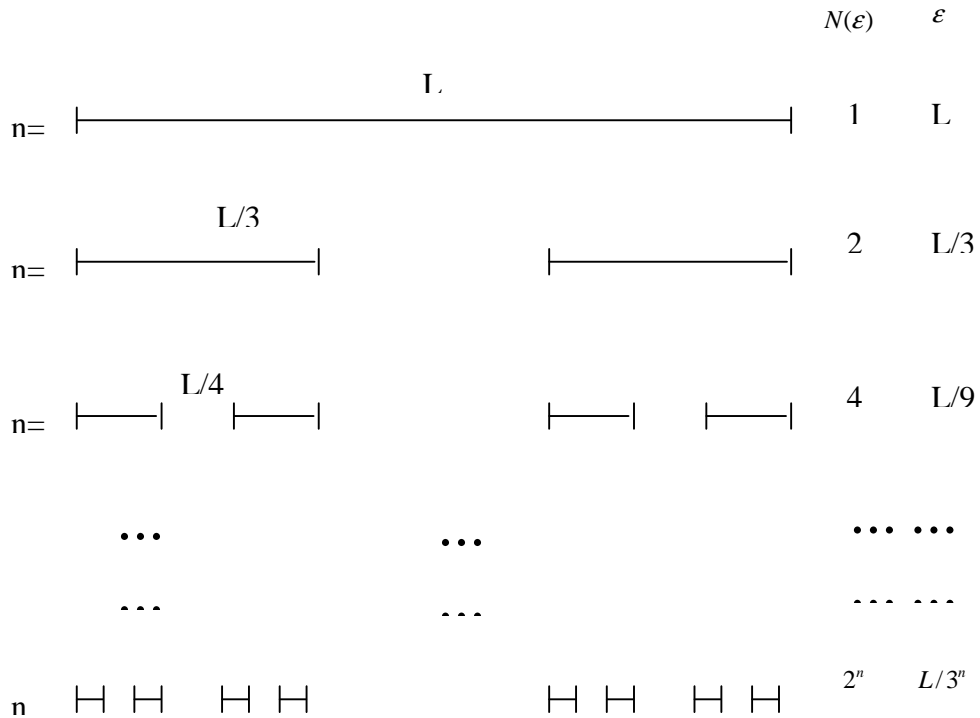


Figura 2.8: Construção do conjunto de Cantor.

Causou espanto a descoberta de que fractais podem ser gerados pelo comportamento assintótico de um sistema dinâmico, representado no Espaço Fase correspondente, ou seja, fractais podem ser atratores de sistema dinâmicos. E estes atratores podem ser caracterizados, por exemplo, pelo valor de D_0 .

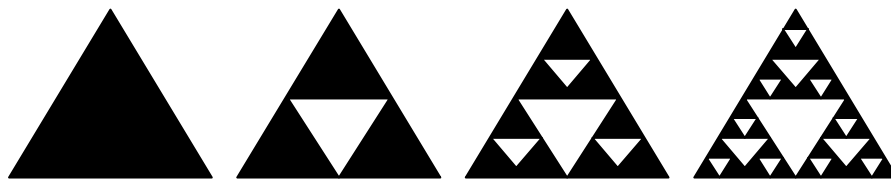


Figura 2.9: Primeiros passos para se construir o triângulo de Sierpinski.

De fato, o cálculo de D_0 é útil na caracterização de atratores. A dimensão dos atratores de processos não-caóticos, em sistemas de tempo contínuo, é inteira. Por exemplo, para um sistema tridimensional, tem-se como atratores: o ponto de equilíbrio, que possui $D_0 = 0$; o ciclo-limite, com $D_0 = 1$; o toro, com $D_0 = 2$. Para processos caóticos, os atratores possuem dimensão fracionária. Por exemplo, o sistema de Lorenz, para $\sigma = 10$, $b = 8/3$ e $r = 28$, possui $D_0 \cong 2,06$, assim o atrator que parece uma asa de borboleta é “uma superfície com uma ‘ligeira’ espessura”. O valor de D_0 revela o

número de variáveis necessárias para se descrever o comportamento assintótico do sistema.

Os algoritmos numéricos usados para o cálculo de D_0 são chamados algoritmos de contagem de caixas. Eles são elaborados do seguinte modo:

- itera-se (no caso de um mapa) ou integra-se (no caso de equações diferenciais) o sistema “muitas” vezes, despreza-se o comportamento transiente e assinalam-se, no Espaço Fase, os pontos que pertencem ao atrator;
- divide-se a região do Espaço Fase ocupada pelo atrator em $N(r)$ caixas de dimensão r ;
- contam-se quantas caixas $N(r)$ contém pelo menos um ponto do atrator;
- refazem-se os cálculos para outros valores de r ;
- D_0 corresponde à inclinação média de $\log N(r)$ em função de $\log(1/r)$.

A dimensão de contagem de caixas é um caso especial da dimensão proposta por Hausdorff. Para o cálculo da dimensão da Hausdorff, D_H , é necessário encontrar o conjunto de hiper-cubos de lados r_j , com $t_j \leq r$, que cubra o hiper-volume ocupado pelo atrator, de modo que o número de hiper-cubos utilizados seja mínimo. É praticamente impossível determinar o valor de D_H para um atrator estranho, pois, de todas as maneiras possíveis para se cobrir tal atrator, deve-se encontrar a solução ótima. Claramente, $D_H \leq D_0$. Tanto D_0 , quanto D_H , costumam ser chamados de dimensões fractais.

Muitos atratores estranhos não são homogêneos, o que implica que algumas de suas regiões são mais visitadas do que outras. Por isso, para se caracterizar um atrator, calculam-se outras dimensões que levam em conta suas não-homogeneidades. Essas dimensões foram propostas, por J. Belatoni e A. Rényi [36]. Tais dimensões se baseiam na frequência relativa f_j com que cada caixa j é visitada. A frequência f_j é definida simplesmente como:

$$f_j = \lim_{N \rightarrow \infty} \frac{N_j}{N} \quad (2.8)$$

sendo N o número total de pontos que pertencem ao atrator e N_j o número de pontos na caixa j .

Dentre as dimensões generalizadas de Rényi, as mais relevantes (além de D_0) são a dimensão de informação D_1 :

$$D_1 = -\lim_{r \rightarrow 0} \sum_{j=1}^{N(r)} \frac{f_j \log f_j}{\log(1/r)} \quad (2.9)$$

e a dimensão de correlação D_2 :

$$D_2 = \lim_{r \rightarrow 0} \sum_{j=1}^{N(r)} \frac{(f_j)^2}{\log r} \quad (2.10)$$

O numerador de D_1 é conhecido como a entropia de Shannon.

Para um atrator não-homogêneo, as frequências relativas são desiguais, o que implica:

$$D_1 < D_0 \quad \text{pois} \quad \sum_j f_j \log(1/f_j) < \log N(r)$$

Na verdade, pode-se mostrar que:

$$D_2 \leq D_1 < D_0 \quad (2.11)$$

sendo que a igualdade só ocorre para atratores homogêneos.

O termo $(f_j)^2$ que aparece na fórmula (2.10) representa a frequência relativa com que dois pontos de um atrator caem na j -ésima caixa de tamanho r . Em [25] Grassberger e Procaccia sugeriram aproximar $(f_j)^2$ pela frequência relativa com que dois pontos estão separados por uma distância menor ou igual a r . Seja $q(r)$ a fração de dois pontos do atrator que está dentro de uma hiper-esfera de raio r , centrada em \bar{x}_j .

Essa fração é expressa pela relação:

$$q(r) = \frac{1}{N} \sum_{k=1}^N H(\tau - \|\bar{x}_j - \bar{x}_k\|) \quad (2.12)$$

sendo N o número de pontos no atrator. A função degrau $H(y)$ é tal que $H=1$ se $y \geq 0$, e $H=0$ se $y < 0$.

Define-se a função de correlação $C(r)$ como o valor médio de $q(r)$, calculado sobre todos os pontos \bar{x}_j . Assim:

$$C(r) = \frac{1}{N} \sum_{j=1(j \neq k)}^N q(r) \quad (2.13)$$

na prática, elege-se, aleatoriamente, N' pontos \bar{x}_j ($N' < N$) e denomina-se o número de pontos que estão contidos na hiper-esfera de raio r . Desse modo:

$$C(r) \cong \frac{1}{N'} \sum_{j=1(j \neq k)}^{N'} q(r) \quad (2.14)$$

O valor de C é proporcional a r se os pontos do atrator estão dispostos sobre uma reta; proporcional a r^2 , se tais pontos estão uniformemente espalhados num plano. No caso de estarem sobre uma estrutura fractal, Grassberger e Procaccia alegaram que C é proporcional a r^{D_2} , ou seja:

$$D_2 \cong \lim_{N \rightarrow \infty} \left(\lim_{\tau \rightarrow 0} \frac{\log C(r)}{\log r} \right) \quad (2.15)$$

Normalmente, calcula-se D_2 para vários r e estima-se seu valor a partir da inclinação da parte linear do gráfico $\log C(r)$ em função do $\log r$.

Mas qual é a utilidade prática de estimar D_2 ? A resposta para esta questão tem duas partes. A primeira refere-se à dificuldade computacional de se calcular D_0 . Numericamente, D_0 é calculado usando-se um algoritmo de contagem de caixas. Tal algoritmo, entretanto, exige um esforço computacional (memória e tempo de processamento) muito maior do que o algoritmo de Grassberger e Procaccia, principalmente quando $D_0 > 2$. E, ao se estimar D_2 , obtém-se um limite inferior para o valor de D_0 , pois a relação (2.11) é verdadeira. Obter D_0 é importante para a caracterização do atrator, já que esse número revela quantas variáveis são necessárias para descrever o comportamento assintótico do sistema. Contudo, a principal razão para se estimar D_2 aparece quando se trabalha com séries temporais reais. No capítulo 3 será definida com maiores detalhes a maneira de calcular a Dimensão da Correlação.

2.9 Conclusões

Neste capítulo foram apresentados uma série de tópicos relacionados com sistemas dinâmicos e teoria do caos. Foram definidos conceitos como; caos, bifurcações, invariantes geométricos, expoente de Lyapunov, Dimensão da Correlação, sensibilidade às condições iniciais etc. No entanto o conceito mais importante apresentado neste capítulo foi o de atrator, e particularmente, o da dimensão de um atrator e sua estimativa. A dimensão de um determinado sistema dinâmico está diretamente relacionado com o objetivo desta tese, que visa reconstruir o Espaço de Imersão ou Espaço Fase do sistema que gerou dita série, a partir de informações contidas exclusivamente numa série temporal de demanda de energia elétrica.

No seguinte capítulo é apresentado o conceito de reconstrução de Espaço de Imersão, que junto com alguns exemplos simples, mostram como é possível “operacionalizar” a reconstrução de um Espaço de Imersão desconhecido. No Capítulo 3 a seguir, são também apresentados os conceitos de Falsos Vizinhos Mais Próximos e Método da Informação Mútua, que junto com o método da Dimensão da Correlação são a base do processo de reconstrução utilizado nesta tese.

CAPÍTULO 3

RECONSTRUÇÃO DO ESPAÇO DE IMERSÃO

3.1 Introdução

Neste capítulo será apresentado o conceito de reconstrução de Espaço Fase ou Espaço de Imersão. Particularmente serão descritas as características do Teorema de Takens que é a base de sustentação desta tese, que visa a re-criação de um retrato de fases a partir da análise de série temporais de consumo ou demanda de energia elétrica. Serão apresentados alguns exemplos de como representar um vetor de reconstrução a partir de uma série de carga. Serão descritos em detalhes os passos para o cálculo da Dimensão da Correlação apresentada no capítulo 2 e a utilidade desta metodologia como alternativa para a definição da Dimensão da Imersão. Também com o objetivo de calcular a Dimensão da Imersão m , será apresentado o método dos Falsos Vizinhos mais Próximos. Por outro lado e visando o cálculo do Passo de Reconstrução τ , será apresentado o método de Informação Mútua que é pré-requisito para o cálculo da Dimensão da Imersão..

3.2 Conceito de Reconstrução

Se as equações diferenciais que governam um sistema dinâmico são conhecidas, pode-se caracterizar o atrator deste sistema calculando as dimensões D_0 , D_1 , D_2 além dos expoentes de Lyapunov (cabe ressaltar que as dimensões D_0 , D_1 , D_2 junto com os expoentes de Lyapunov são conhecidas como invariantes geométricos do sistema). Entretanto, em problemas reais, normalmente tem-se uma série temporal de dados experimentais. Não se sabe, de antemão, quais são as variáveis de estado deste sistema e nem a forma das equações diferenciais que governam a evolução dessas variáveis.

Considere o sistema de Rössler que é composto pelas seguintes equações diferenciais :

$$\begin{aligned} \frac{dx}{dt} &= -y - z, \\ \frac{dy}{dt} &= x + ay, \end{aligned} \tag{3.1}$$

$$\frac{dz}{dt} = b + (x - c)z,$$

sendo a, b, c constantes ($a = 0,2$; $b = 0,4$; $c = 5,7$).

Existe um atrator estranho no Espaço Fase deste sistema dinâmico (atrator estranho tem dimensão fracionária e sensibilidade às condições iniciais). O espaço é, obviamente, formado pelo eixo- x , pelo eixo- y e pelo eixo- z . Em 1980, N. H. Packard, J.P. Crutchfield, J.D. Farmer e R.S. Shaw [37] analisaram o comportamento deste sistema no espaço formado pelo eixo- x , pelo eixo- dx/dt e pelo eixo- d^2x/dt^2 . Foi mostrado que, neste novo espaço, a figura geométrica que caracteriza o comportamento assintótico do sistema é topologicamente equivalente ao atrator estranho original. Essa figura é chamada de atrator reconstruído. Assim, o atrator original (no espaço x, y, z) e o atrator reconstruído (no espaço $x, dx/dt, d^2x/dt^2$) são caracterizados pelos mesmos valores de dimensões e expoente de Lyapunov. Portanto, a partir da evolução temporal de uma única variável, $x(t)$ no caso, pode-se determinar as características do atrator.

Como exemplo considere um sistema dinâmico que possui um ciclo-limite circular, assintoticamente estável no seu Espaço Fase, descrito pelas expressões $x(t) = \sin t$ e $y(t) = \cos t$. Suponha que seja conhecida apenas a variação temporal de $x(t)$. Para reconstruir o atrator, a idéia é traçar a evolução temporal do sistema no Espaço Fase formado pelo eixo- x e pelo eixo- dx/dt , sendo dx/dt aproximado por $[x(t + \tau) - x(t)]/\tau$. A Figura (3.7) ilustra o atrator reconstruído, tomando-se alguns valores diferentes de τ . Nos três casos, o atrator reconstruído é topologicamente equivalente ao atrator original. Repare que $\tau \rightarrow 0$, obtem-se o atrator original, pois, para ele, $dx/dt = y$.

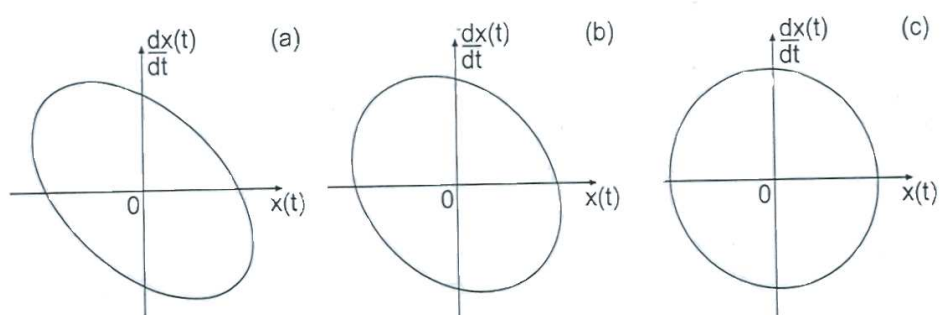


Figura 3.1. Atrator reconstruído para $\tau = 1,0$ (a), $\tau = 0,5$ (b) e $\tau = 0,1$ (c).

Este exemplo é bastante artificial já que em geral, tem-se que $dx/dt = f(x, y)$ e $dy/dt = g(x, y)$.

A pergunta que surge é: como se sabe de antemão, qual a dimensão do Espaço de Imersão no qual se fará a reconstrução do atrator? Como se sabe, no exemplo anterior, que esse espaço deve ser formado apenas pelo eixo- x e pelo eixo- dx/dt ? Por que não considerar também o eixo- d^2x/dt^2 ? Ou seja, como se pode determinar a dimensão que é adequada para se fazer a reconstrução do atrator? Além disso, como descobrir qual o valor do Passo de Reconstrução τ que é o mais conveniente?

Na verdade, o método de reconstrução do atrator proposto por Packard e colaboradores é ligeiramente diferente daquele que foi descrito nos parágrafos anteriores. Sabe-se que dx/dt , d^2x/dt^2 ou derivadas de ordem superior podem ser representadas por equações de diferenças com passo infinitesimal. Assim:

$$\frac{dx}{dt} \cong \frac{x(t+\tau) - x(t)}{\tau} \quad \tau \rightarrow 0 \quad (3.2)$$

$$\frac{d^2x}{dt^2} \cong \frac{x(t+2\tau) - 2x(t+\tau) + x(t)}{\tau^2} \quad \tau \rightarrow 0 \quad (3.3)$$

Uma série temporal real consiste de um conjunto $x_j = x(t_j)$ de valores obtidos nos instantes de tempo t_j , para $j=1, \dots, N$. A determinação numérica de derivadas a partir de um conjunto discreto de pontos é bastante sensível a ruído, o que torna seu cálculo muito impreciso. Além disso, o passo τ possui um valor finito, de maneira que a qualidade da aproximação de uma derivada por uma equação de diferença diminui com o aumento da ordem da derivada.

Em vez de usar $x(t)$, $dx(t)/dt$, $d^2x(t)/dt^2$, como as variáveis que formarão o espaço no qual o atrator será reconstruído, Packard e colaboradores sugeriram usar, diretamente, $x(t)$, $x(t+\tau)$, $x(t+2\tau)$... etc.

3.3 Teorema de Reconstrução de Takens

Em 1981, F. Takens [18] provou que, no Espaço Fase formado pelo eixo- $x(t)$, pelo eixo- $x(t+\tau)$, pelo eixo- $x(t+2\tau)$... e pelo eixo- $x(t+(m-1)\tau)$, o atrator reconstruído é topologicamente equivalente ao atrator “real”, sobre o qual conhece-se apenas a evolução em tempo discreto da variável de estado x . Na sua prova, Takens assumiu que a série é formada por infinitos pontos x_j e que não há ruído. Se essas condições são satisfeitas, as propriedades topológicas do atrator reconstruído são preservadas, tomando-se :

$$m \geq 2D_0 + 1 \quad (3.4)$$

sendo D_0 a dimensão de contagem de caixas do atrator “real”.

Chama-se Espaço de Imersão o espaço no qual realiza-se a reconstrução. Denomina-se m de Dimensão de Imersão e τ de Passo da Reconstrução. Note que τ deve ser múltiplo do tempo de amostragem do sinal contínuo $x(t)$.

Considerando-se como exemplo o caso do Sistema de Lorenz, com $\sigma = 10$, $b = 8/3$ e $r = 28$, onde a dimensão de contagem de caixas do atrator “real” vale $D_0 \cong 2,06$ [26]. Assim, segundo Takens, pode-se reconstruir o atrator a partir de $x(t)$, tomando-se $m \geq 2D_0 + 1$, nesse caso, $m = 6$ parece ser suficiente. Entretanto, num problema real, não se conhece o atrator original, nem seu valor D_0 , podendo ser determinado conforme mostrado a seguir.

Pelo método dos atrasos temporais de Takens, a cada instante t_j , assinala-se o ponto de coordenadas

$$X(j) = \{ x(t_j), x(t_j + \tau), x(t_j + 2\tau) \dots x(t_j + (m-1)\tau) \} \quad (3.5)$$

onde X é o vetor reconstruído num espaço multidimensional ou espaço de imersão. Variando-se j de 1 até N , obtém-se a trajetória reconstruída.

Assumindo-se que $X(\alpha)$ represente a posição do ponto no espaço de imersão, no instante t_α , tem-se que a trajetória reconstruída é formada pela seqüência:

$$\begin{aligned} X(1) &= (x(t_1), x(t_1 + \tau), \dots, x(t_1 + (m-1)\tau)) \\ X(2) &= (x(t_2), x(t_2 + \tau), \dots, x(t_2 + (m-1)\tau)) \\ &\vdots \\ X(\alpha) &= (x(t_\alpha), x(t_\alpha + \tau), \dots, x(t_\alpha + (m-1)\tau)) \end{aligned} \quad (3.6)$$

com $\alpha = 1, \dots, M$. As constantes m , τ , N , M relacionam-se por $N = M + (m-1)\tau$.

3.4 Reconstrução de Takens Autoregressiva

Dado que o objetivo desta tese é a reconstrução do espaço de imersão a partir de valores passados de uma única série temporal faz-se necessária uma pequena adaptação da equação 3.5 para:

$$X(j) = \{ x(t_j), x(t_j - \tau), x(t_j - 2\tau) \dots x(t_j - (m-1)\tau) \} \quad (3.7)$$

As séries temporais são expandidas em espaços dimensionais superiores, nos quais toma lugar a dinâmica geradora do processo. Isto é realizado utilizando-se a técnica de coordenadas em atraso já descrita acima, que recria um retrato da dinâmica do sistema em análise, a partir da série temporal de valores observados.

Como exemplo, considere uma pequena amostra de dados de demanda horária de uma distribuidora de energia elétrica:

Tabela 3.1. Amostra de dados de uma série de carga

tempo t_j	Valor $x(t_j)$	tempo t_j	Valor $x(t_j)$
t_1	8029	t_6	8592
t_2	7785	t_7	8975
t_3	7900	t_8	8500
t_4	8417	t_9	8340
t_5	8452	t_{10}	8623

Como premissa considere $\tau=1$ e $m=3$, sabe-se também que $N=10$. Desta forma, é possível gerar 8 ($M = 8 = N - (m - 1)\tau$) coordenadas ou vetores reconstruídos, num espaço multidimensional $m=3$:

$$X(1) = \{7900, 7785, 8029\}$$

$$X(2) = \{8417, 7900, 7785\}$$

$$X(3) = \{8452, 8417, 7900\}$$

$$X(4) = \{8592, 8452, 8417\}$$

$$X(5) = \{8975, 8592, 8452\}$$

$$X(6) = \{8500, 8975, 8592\}$$

$$X(7) = \{8340, 8500, 8975\}$$

$$X(8) = \{8623, 8340, 8500\}$$

Observe que a partir de vetores reconstruídos em espaços superiores pode-se estimar a Dimensão de Correlação D_2 , calculando-se $C(r)$ segundo definido no capítulo 2. Mas em um problema real não se sabe a priori que Dimensão de Imersão m é conveniente para a reconstrução do atrator. Para se encontrar o valor de m mais adequado, calcula-se $C(r)$ para vários valores de τ , variando m , por exemplo entre 2 e 7; e representam-se estes resultados num gráfico $\log C(r)$ em função de $\log(r)$. A

inclinação da parte linear dos gráficos fornece uma estimativa para o valor de D_2 para cada m . O valor de m satura a partir de um valor crítico. Essa é a Dimensão de Imersão que permite “desdobrar” o atrator adequadamente, de modo que suas trajetórias não se cruzem no Espaço de Imersão.

É interessante apontar que se $x(t_j)$ é um sinal aleatório, D_2 não satura para nenhum valor de m . Lembre-se que o valor de saturação de D_2 é uma estimativa para a dimensão D_0 do atrator “real”.

O trabalho de Takens, no qual se considera uma série infinita e sem ruídos, não impõe restrição sobre o Passo de Reconstrução τ . Entretanto, ao se trabalhar com séries finitas e ruidosas, a escolha de τ torna-se crítica. Considere, por exemplo um gráfico com $m=2$, $x(t)$ vs $x(t-\tau)$. Se τ é muito “pequeno”, então $x(t)$ e $x(t-\tau)$ são praticamente idênticos ficando o atrator comprimido ao longo da diagonal principal do Espaço de Imersão; se τ é muito “grande”, então $x(t)$ e $x(t-\tau)$ não apresentam correlação alguma.

3.5 Calculando a Dimensão de Correlação

A Dimensão de Correlação é uma medida estatística para avaliar a auto-similaridade da geometria de um conjunto de pontos no Espaço Fase. Esta quantidade define o número de variáveis independentes que seriam necessárias para descobrir a dimensão de um sistema, i.e, os graus de liberdade do mesmo.

A dimensão de correlação m é um dos primeiros parâmetros utilizados para descrever atratores. Basicamente, fornece uma medida da complexidade do sistema em relação ao número de graus de liberdade do mesmo. Considerando-se que m converge a um valor finito no caso de sistemas determinísticos e que não converge no caso de sistemas estocásticos, pode ser útil para avaliar a natureza determinística ou estocástica de um sistema. Como já foi citado Grassberg e Procaccia propuseram em [25] um método para definir m de maneira experimental, variações deste método são abordados também em [28], [29], [38].

Usando uma série temporal $(y_i, i=1, \dots, N)$ são construídos M pontos num Espaço Fase de m dimensões (Dimensão da Imersão) de acordo com o teorema de Takens [18]

No Espaço Fase a integral de correlação $C_m(r)$ está definida por:

$$C_m(r) = \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \Theta \left(r - \left\| \vec{x}_i - \vec{x}_j \right\| \right) \quad (3.8)$$

$\Theta(z)$ é a função Heaviside

$$\Theta(z) = \begin{cases} 0, & z < 0 \\ 1, & z > 0 \end{cases} \quad (3.9)$$

$$\left\| \vec{x}_i - \vec{x}_j \right\| = \sqrt{\sum_{k=1}^d (x_{i,k} - x_{j,k})^2} \quad (3.10)$$

é a norma euclideana e d é a dimensão do Espaço Fase do atrator.

Grassberg e Procaccia demonstraram que D_2 pode ser obtida através da seguinte equação:

$$D_2 = \lim_{r \rightarrow 0} \left[\frac{\log C_m(r)}{\log(r)} \right]. \quad (3.11)$$

Plotando-se o gráfico de $\log C_m(r)$ vs $\log(r)$, o coeficiente angular resultante do ajuste de uma regressão linear (para valores baixos de r) produz a dimensão de correlação. $C_m(r)$ é calculada para cada uma das sucessivas dimensões de m , determinando-se assim, os coeficientes angulares para cada valor de m , definido como $d(m)$. Conforme o valor de m aumenta, $d(m)$ tende a um valor constante, isto é, à saturação. O valor da Dimensão da Imersão que provoca saturação da função $d(m)$ é conhecido como Dimensão de Correlação. Na Figura 3.2 pode ser observado o resultado do cálculo da Dimensão da Correlação para diferentes valores de imersão do Espaço Fase m para o mapa logístico. Observe que o valor da imersão é $m=2$.

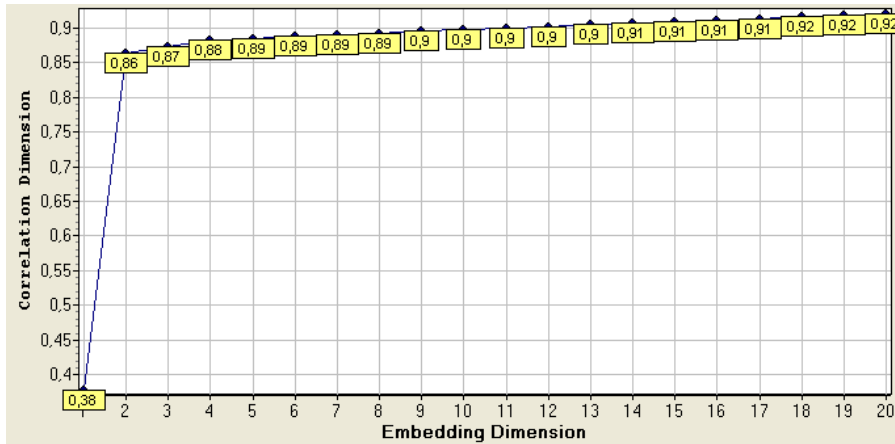


Figura 3.2. Cálculo da Dimensão de Correlação para o mapa logístico ($m=2$)

A seguir é apresentado o procedimento para cálculo de $C_m(r)$ ou $C(r)$:

- construir a matriz $(N-1 \times N-1)$ das distâncias $\|\bar{x}_i - \bar{x}_j\|$;
- para um valor predefinido de r , construir a matriz para a função Heaviside $Z(r - \|\bar{x}_i - \bar{x}_j\|)$, onde $Z=0$, para valores de $r - \|\bar{x}_i - \bar{x}_j\| \leq 0$, e $Z=1$, para $r - \|\bar{x}_i - \bar{x}_j\| > 0$;
- definir um valor de imersão m inicial;
- calcular o valor de $C(r)$, para diversos valores de r ;
- calcular $\log(C(r))$ e $\log(r)$;
- plotar um gráfico de dispersão;
- ajustar a uma reta. O coeficiente angular define o valor da Dimensão de Correlação (D_2), para um determinado m ;
- repetir para diferentes valores de imersão m ;
- plotar um gráfico entre a Dimensão da Imersão m vs a Dimensão de Correlação D_2 ;
- o valor de imersão apropriado é aquele que produz a saturação da Dimensão de Correlação D_2 .

3.6 Influência do Passo de Reconstrução τ na Qualidade da Reconstrução

Nesta seção são apresentados dois exemplos de reconstrução do Espaço de Imersão em sistemas dinâmicos não lineares de até três variáveis de estado. Os sistemas são amplamente citados na literatura e sua utilização neste trabalho contribui para o

aumento do grau de compreensão do conceito de reconstrução. Como neste caso o objetivo é pedagógico os sistemas utilizados são conhecidos, o que significa que as dimensões dos atratores são também conhecidas, cabendo apenas avaliar o impacto de uma correta escolha do Passo de Reconstrução τ , para o qual é utilizado o “bom e velho” método de tentativa e erro.

3.6.1 Sistema de Van der Pol

Considere a clássica equação de Van der Pol que descreve um oscilador:

$$\frac{d^2x}{dt^2} - u \cdot (1 - x^2) \frac{dx}{dt} + x = 0 \quad (3.11)$$

Como se sabe, as equações diferenciais de ordem superior podem ser reescritas em termos de um conjunto equivalente de equações diferenciais de primeira ordem. Para a equação (3.11), isso é obtido pela definição de duas novas variáveis:

Sejam $y_1 = x$ e $y_2 = \frac{dx}{dt}$

então

$$\frac{dy_1}{dt} = y_2 \quad (3.12)$$

$$\frac{dy_2}{dt} = u \cdot (1 - y_1^2) - y_2$$

Usando um método numérico para resolver (3.12) obtém-se os resultados apresentados na Figura 3.3, onde é possível observar o comportamento das variáveis x e y no tempo.

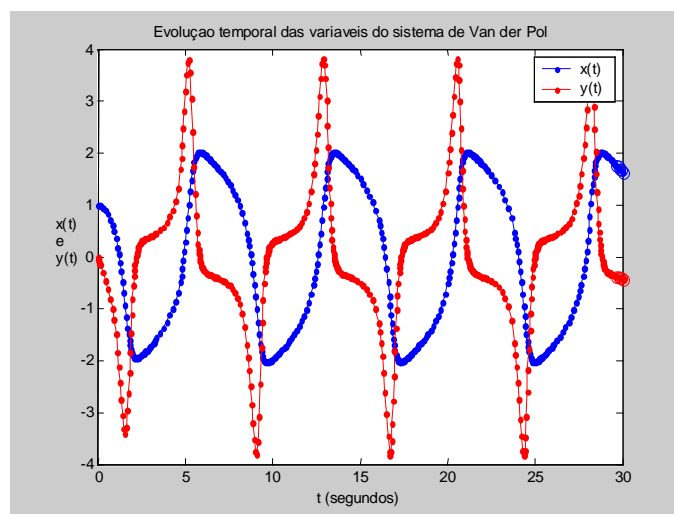


Figura 3.3. Solução para o sistema de Van der Pol

Na Figura 3.4 pode-se observar a dinâmica da trajetória do sistema no Espaço Fase original, isto é, com suas duas variáveis de estado $x(t)$ e $y(t)$. Como já mencionado, a Figura 3.4 também é conhecida como atrator do sistema.

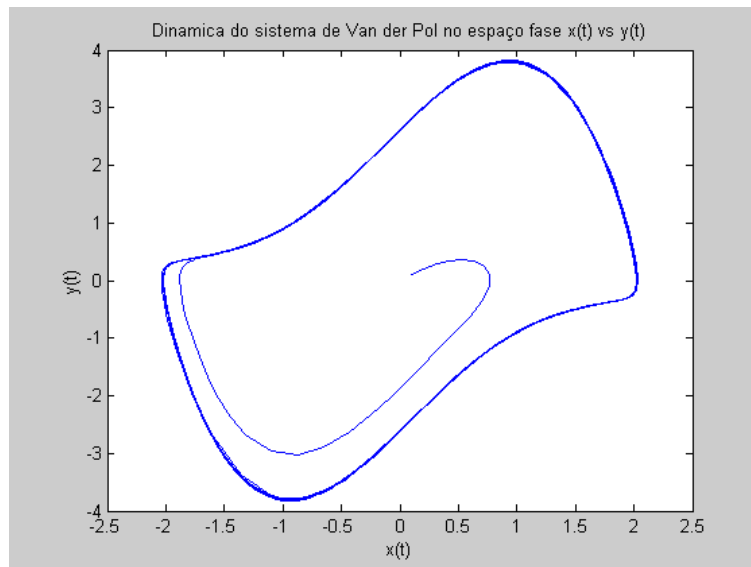


Figura 3.4. Atrator do sistema de Van der Pol no Espaço Fase original ($x(t)$ vs $y(t)$)

3.6.2 Reconstrução do Sistema de Van der Pol

A premissa adotada a seguir considera a impossibilidade de observação de todas as variáveis de estado do sistema. De fato neste caso, somente está disponível o sinal $x(t)$, a partir do qual se deseja reconstruir o “desconhecido” sistema gerador do sinal. O sinal $x(t)$ é mostrado na Figura 3.5.

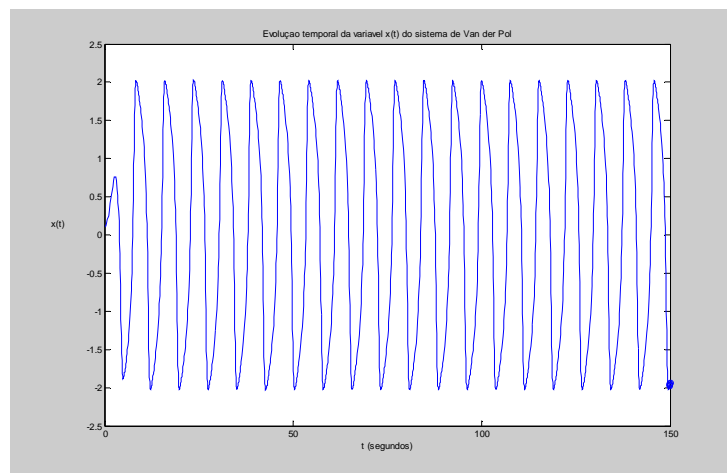


Figura 3.5. Variável $x(t)$ do sistema de Van der Pol

Na Figura 3.6 são apresentadas duas alternativas de reconstrução do retrato de fases: na primeira (a) usando $x(t)$ vs $x(t-1)$; na segunda (b) é fácil observar que as

variáveis utilizadas ($x(t)$ vs $x(t-10)$) representam melhor a trajetória do atrator original da Figura 3.4.

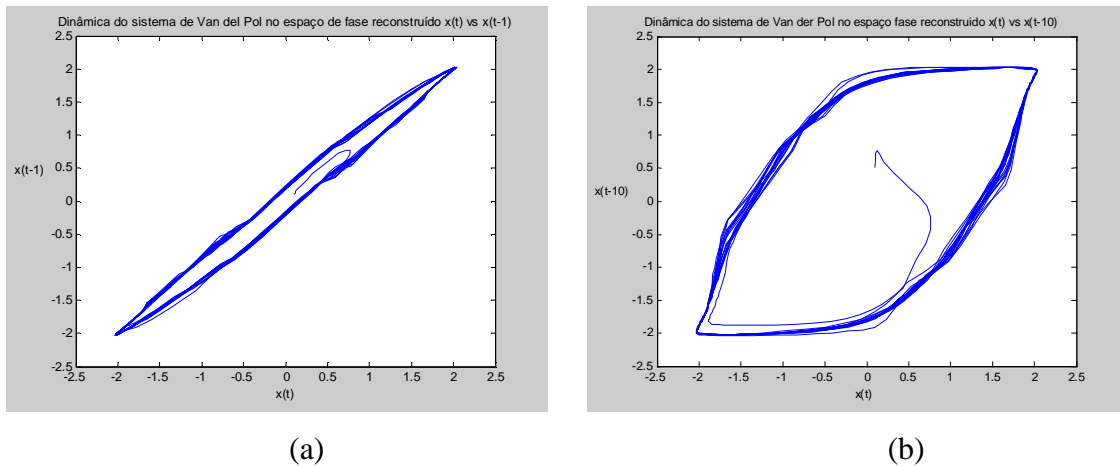


Figura 3.6. Trajetória reconstruída a partir da variável $x(t)$ observada.

3.6.3 Exemplo 4. Sistema de Hénon

O último exemplo deste capítulo refere-se ao sistema ou mapa de Hénon, que possui duas variáveis de estado e também comumente citado na literatura técnica. O mapa é definido por um sistema não linear de equações algébricas como definido a seguir:

$$x(i+1) = 1 - a \cdot x^2(i) + y(i) \tag{3.13}$$

$$y(i+1) = b \cdot x(i)$$

O atrator do mapa de Hénon do Espaço Fase original é representado na Figura 3.7

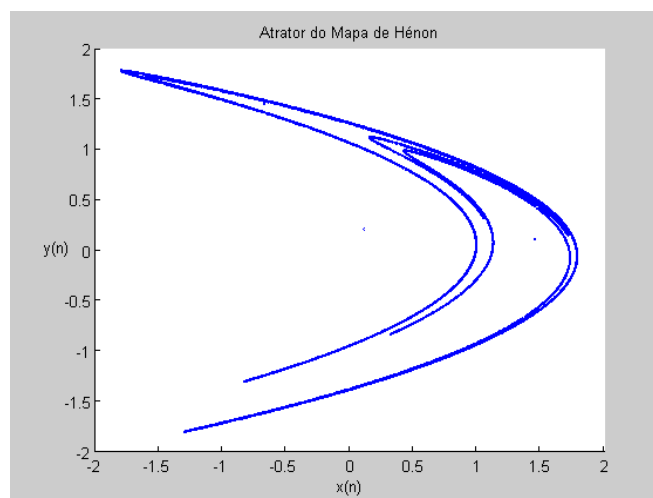


Figura 3.7. Atrator do mapa de Hénon no Espaço Fase original ($x(t)$ vs $y(t)$)

3.6.4 Reconstrução do Sistema de Hénon

Nesta seção pretende-se reconstruir o Espaço Fase do sistema de Hénon a partir da observação de somente uma de suas variáveis, sendo disponibilizada a série temporal da variável $x(t)$. Na Figura 3.8 observa-se o comportamento da variável $x(t)$. Nas Figuras 3.9(a) e 3.9(b) são apresentadas duas tentativas de reconstrução do Espaço Fase do sistema de Hénon: na primeira (a) é usado o Espaço Fase $\{x(t), x(t-24)\}$ e na segunda (b) é utilizado o Espaço Fase $\{x(t), x(t-2)\}$ com o qual se observa uma leve melhoria na representação, pelo menos quando comparada com a Figura 3.10(a). Finalmente, na Figura 38 é apresentada a terceira tentativa de reconstrução do Espaço Fase do mapa de Hénon. Nesta última versão as variáveis utilizadas foram $\{x(t), x(t-1)\}$, e observa-se claramente ser esta mais adequada para a representação do retrato de fases do sistema em questão.

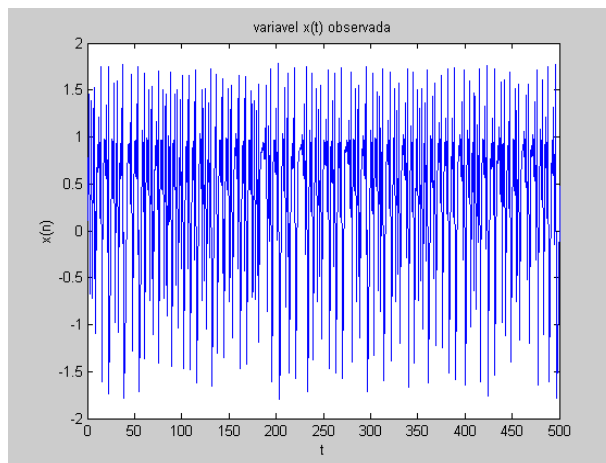


Figura 3.8. Variável $x(t)$ do mapa de Hénon

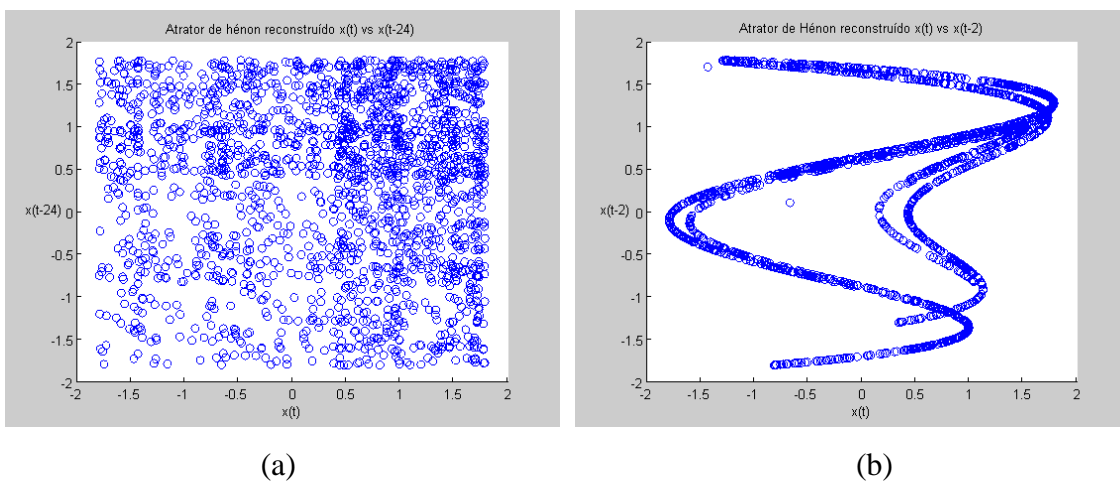


Figura 3.9. Trajetórias reconstruídas a partir da variável $x(t)$ observada

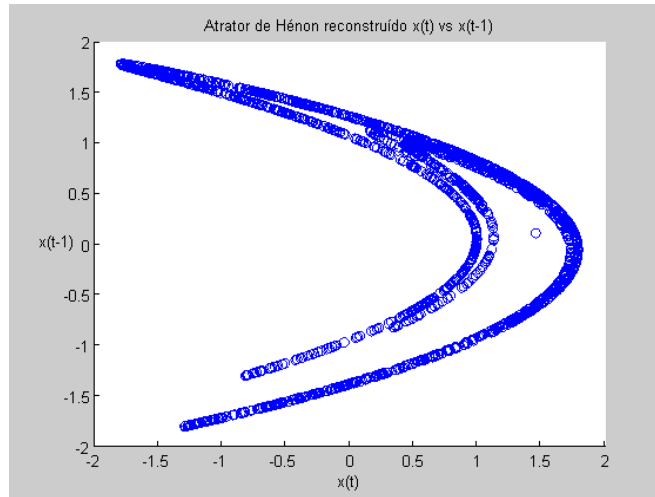


Figura 3.10. Trajetórias reconstruídas a partir da variável $x(t)$ observada

3.7 Determinação do Passo de Reconstrução τ

Na seção anterior foram apresentados alguns exemplos teóricos onde se evidencia a importância de uma escolha correta para a variável Passo de Reconstrução τ , nesta etapa do trabalho é detalhada a metodologia disponível para realizar o cálculo desta variável.

3.7.1 Informação Mútua

A informação mútua defasada no tempo foi sugerida por Fraser e Swinney em [40] como ferramenta para determinar um atraso adequado. Diferentemente da função de autocorrelação, a informação mútua considera também interdependência não lineares. A informação mútua captura informações de momentos superiores a diferença da correlação linear que somente captura informações de segundo ordem. Pode-se defini-la da seguinte forma:

$$S = - \sum_{ij} p_{ij}(\tau) \cdot \ln \left(\frac{p_{ij}(\tau)}{p_i p_j} \right) \quad (3.14)$$

onde para uma partição de números reais, p_i é a probabilidade de encontrar um valor na série de tempo no i -ésimo intervalo, e $p_{ij}(\tau)$ é a probabilidade condicional que uma observação falhe (não ocorra) no i -ésimo intervalo e que a observação falhe τ tempo mais tarde no j -ésimo intervalo. Em teoria esta expressão não é sistematicamente dependente do tamanho da partição e pode ser facilmente calculada. Existem bons argumentos que dizem que se a informação mútua defasada exibe um mínimo tendendo

a um certo valor de τ , então este valor é um bom candidato a ser o atraso adequado [39], [10], [41]. No entanto, estes argumentos são modificados quando a imersão do espaço é maior que dois. Além disso, não são todas as aplicações que trabalham otimamente com o mesmo atraso [29].

Como exemplo, na Figura 3.11, é apresentado gráfico de informação mútua para o mapa logístico (vide Anexo B). Particularmente foi utilizado um valor de 3,544090 para a variável de controle r . Os cálculos foram realizados utilizando o aplicativo VRA4.7 versão demo [42].

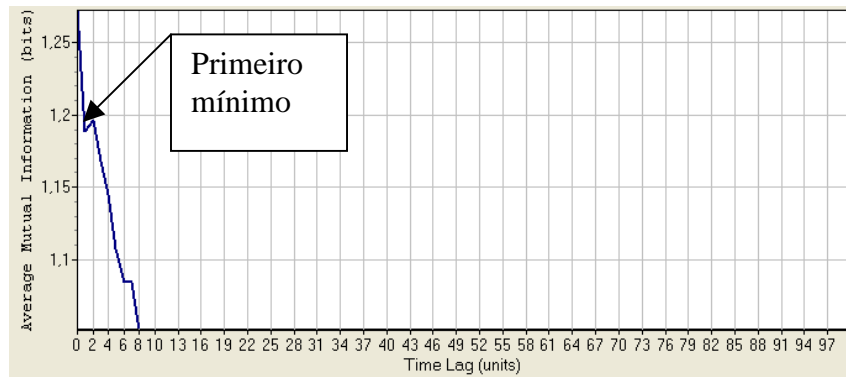


Figura 3.11. Cálculo da Informação Mútua para o mapa logístico ($\tau = 1$).

3.8 Definição da Dimensão da Imersão m

Como dito anteriormente, ao se iniciar a análise de uma série temporal, desconhece-se a dimensão D_0 do seu sistema gerador, informação que é formalmente imprescindível para definir m ($m \geq 2D_0 + 1$), que juntamente com o Passo de Reconstrução τ constituem-se nos parâmetros necessários para a reconstrução do Espaço de Imersão. O procedimento para a obtenção destes parâmetros depende muito da dinâmica implícita nos dados e do tipo de análise. O mais importante é que os teoremas de imersão “garantem” que, para dados ideais, isto é, livres de ruído, existe uma dimensão m tal, que os vetores reconstruídos são equivalentes aos vetores do Espaço Fase original.

É desejável ter um conhecimento preciso de m caso se pretenda explorar o determinismo com mínimo esforço computacional. Se a dimensão m produz uma representação fidedigna do espaço estado, cada reconstrução m' com $m' > m$ também o fará. A escolha de valores muito grandes de m para dados caóticos é uma redundância e somente prejudica a performance de muitos algoritmos, tais como os usados para calcular expoentes de Lyapunov. Dada a instabilidade dos movimentos caóticos, os

primeiros e últimos elementos de um vetor defasado estão menos relacionados devido a sua grande diferença temporal. Por isto adotar um valor grande de m não ajuda muito, pelo contrário, poderia inclusive confundir os algoritmos de cálculo.

3.8.1 Falsos Vizinhos Mais Próximos

Assumindo-se que a dinâmica no Espaço Fase é representada por um vetor de fluxo amortecido (dinâmica suave), é razoável assumir que os estados vizinhos estarão sujeitos à mesma evolução temporal. Assim, após um curto intervalo de tempo futuro, as duas trajetórias emergentes ainda serão vizinhas próximas mesmo que o caos introduza uma divergência exponencial entre as duas. Faz-se referência a esta propriedade com o objetivo de discutir estatísticas que algumas vezes ajudam a identificar se determinadas dimensões são suficientes para a reconstrução do Espaço de Imersão.

Como já foi citado o conceito batizado como Falsos Vizinhos Mais Próximos foi introduzido por Kennel, Brown e Abarbanel em [24]. Em [43] e [44] é apresentado o método com pequenas modificações, as quais evitam alguns resultados espúrios com ruído.

A idéia fundamental é procurar pontos no conjunto de dados que sejam vizinhos num determinado espaço de imersão, mas que poderiam não ser vizinhos em espaços de dimensões superiores, o que poderia acarretar uma evolução temporal futura muito diferente.

De um ponto de vista geométrico, primeiramente adota-se como premissa que a imersão correta para um conjunto de dados é m_o e posteriormente analisa-se o mesmo conjunto de dados numa imersão dimensional mais baixa $m < m_o$. A transição desde m_o a m é uma projeção, que ocorre eliminando alguns eixos no sistema de coordenadas. Os pontos cujas coordenadas foram eliminadas por diferir fortemente na sua projeção, podem ser considerados “falsos vizinhos” no espaço m dimensional.

A estatística para avaliar os falsos vizinhos pode ser definida da seguinte forma: para cada ponto da série temporal é identificado o seu vizinho mais próximo num espaço m dimensional e calculada a taxa das distâncias entre esses dois pontos no espaço dimensional $m+1$ e o espaço m dimensional. Se esta taxa é maior que uma tolerância R , o vizinho é falso. Esta tolerância deve ser grande o suficiente de forma a permitir divergência exponencial na presença de caos determinístico [40]. Se o desvio

padrão dos dados é definido como σ e é usada a norma máxima, a estatística pode ser calculada da seguinte forma [44]:

$$X_{fmm}(R) = \frac{\sum_{n=1}^{N-m-1} \Theta\left(\frac{|s_n^{m+1} - s_{k(n)}^{m+1}|}{|s_n^m - s_{k(n)}^m|} - R\right) \Theta\left(\frac{\sigma}{R} - |s_n^m - s_{k(n)}^m|\right)}{\sum_{n=1}^{N-m-1} \Theta\left(\frac{\sigma}{R} - |s_n^m - s_{k(n)}^m|\right)}. \quad (3.15)$$

onde s_n é o vetor reconstruído e $s_{k(n)}^m$ é o vizinho mais próximo de s_n em m dimensões, isto é, $k(n)$ é o índice do elemento k da série temporal diferente de n para o qual $|s_n - s_k| = \min$. A primeira função no numerador é unitária quando o vizinho mais próximo é falso, i.e, se a distância aumenta num valor maior que R quando a dimensão de imersão aumenta uma unidade. Além disso a segunda função suprime todos aqueles pares cuja distância inicial seja maior que σ/R . Os pares cujas distâncias são maiores que σ/R , por definição, não podem ser considerados falsos vizinhos, dado que em média não existe espaço suficiente para afastar de σ , sendo portanto considerados candidatos inválidos para o método, excluindo-se sua contabilização, o que também reflete na normalização.

Isto significa que num espaço de atraso de m_o dimensões, o atrator reconstruído é uma imagem do atrator no Espaço Fase original. Particularmente, as propriedades topológicas são preservadas. Na Figura 3.12 pode ser observado o resultado do cálculo do percentual de falsos vizinhos para diferentes valores de imersão do Espaço Fase m para o mapa logístico. Observe que o valor da imersão é $m=2$.

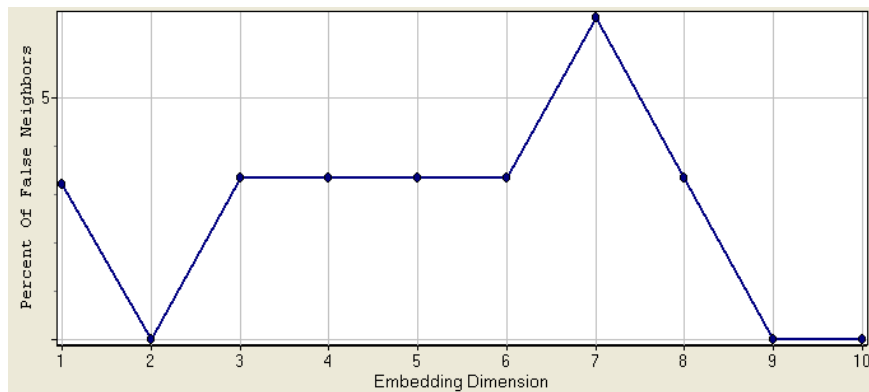


Figura 3.12. Cálculo dos Falsos Vizinhos para o mapa logístico ($m=2$)

Outras variações do método dos falsos vizinhos podem ser encontradas em [45], [46], [47], [48], [49].

3.9 Conclusões

Neste capítulo foi apresentado em detalhes o conceito de reconstrução de Espaço Fase ou Espaço de Imersão. Particularmente foram descritas as características do Teorema de Takens que é a base de sustentação desta tese, que tem como objetivo a recriação de um retrato de fases a partir da análise de informações contidas numa única série temporal de consumo ou demanda de energia elétrica. Foi dado um exemplo de formação de um vetor de reconstrução a partir de uma série de carga. Foi realizada uma descrição passo a passo do cálculo da Dimensão da Correlação e a utilidade desta metodologia como alternativa para a definição da Dimensão da Imersão m , que é uma das variáveis necessária para a aplicação do Teorema de Takens. Também com o objetivo de calcular a Dimensão da Imersão m , foi apresentado o método dos Falsos Vizinhos mais Próximos. Visando o cálculo do Passo de Reconstrução τ , foi apresentado o método de Informação Mútua que é pré-requisito para o cálculo da Dimensão da Imersão. Para ressaltar a importância do cálculo correto da variável τ , foram apresentados dois exemplos teóricos de reconstrução cuja Dimensão da Imersão m , era conhecida previamente.

No Capítulo 4 a seguir, serão apresentados dois exemplos que usam séries de tempo originadas por atratores com dimensões D_0 conhecidas. O objetivo é testar os resultados do processo de reconstrução, aplicando os métodos apresentados nas seções anteriores desta tese.

CAPÍTULO 4

APLICAÇÕES DO TEOREMA DE TAKENS

4.1 Introdução

Como mencionado nos capítulos anteriores, o Teorema de Takens permite a reconstrução do Espaço de Imersão a partir da definição de duas variáveis: m denominada Dimensão da Imersão e τ conhecida como Passo de Reconstrução, sendo que a análise realizada utiliza exclusivamente dados oriundos de uma série temporal gerada por um sistema desconhecido ou atrator. A aplicação destas duas variáveis na equação 3.7 permite a reconstrução do dito Espaço de Imersão.

Na Figura 4.1 é apresentado um diagrama esquemático do processo de reconstrução utilizado nesta tese. Pode ser observado que a partir de uma série temporal é definido o Passo de Reconstrução através do método da Informação Mútua. Posteriormente é utilizada a mesma série temporal junto com a variável τ (previamente definida) para calcular a Dimensão da Imersão usando os dois critérios apresentados nos capítulos anteriores: método dos Falsos Vizinhos mais Próximos e método da Dimensão da Correlação. Finalmente são aplicadas as duas variáveis m e τ na equação 3.7 visando a reconstrução do Espaço de Imersão. Observe que, para o cálculo de m , sempre existirão duas alternativas visto que se trabalha com dois métodos. A escolha de uma ou outra opção será função de alguns critérios empíricos comentados no desenvolvimento das próximas seções deste trabalho.

Não obstante o processo de reconstrução parecer simples, existem alguns problemas de origem prática na hora da sua aplicação, problemas que serão descritos a seguir:

1. segundo o capítulo 3, para que a Dimensão da Imersão m seja válida a mesma tem que cumprir a seguinte restrição $m \geq 2D_0 + 1$ (equação 3.4), onde D_0 é a dimensão do atrator que gerou a série em análise. Ocorre que a única informação disponível encontra-se na série temporal, isto é, não existe nenhum conhecimento sobre o sistema que gerou a série ou sobre as variáveis explicativas do fenômeno e, menos ainda, a dimensão do atrator D_0 responsável pela geração dos dados;

2. o trabalho de Takens considera uma série infinita e sem ruídos. Sabemos que no mundo real não existem séries infinitas e que, particularmente as séries de energia elétrica, são também constituídas de ruído;

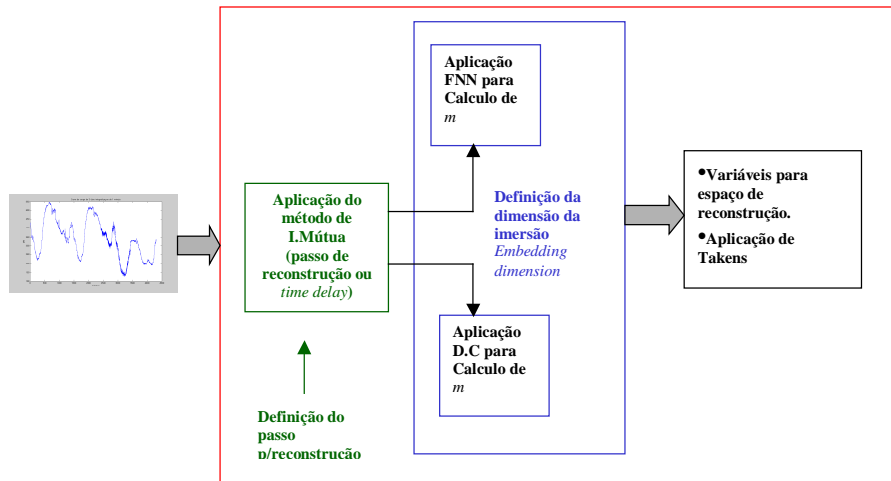


Figura 4.1. Processo de reconstrução do Espaço de Imersão utilizado

3. como descrito no capítulo 3, o estudo de Takens não impõe restrição sobre o Passo de Reconstrução τ , entretanto, ao se trabalhar com séries finitas e ruidosas, a escolha de τ torna-se crítica.

Neste capítulo é testada a aplicação do Teorema de Takens em alguns sistemas cuja dimensão do atrator D_0 é previamente conhecida. São feitas simulações originadas de informações disponíveis nas séries geradas por estes sistemas e em seguida extraídas variáveis necessárias para aplicação do Teorema de Takens. De forma resumida, pode-se afirmar que o objetivo deste capítulo é avaliar a confiabilidade e viabilidade da aplicação do teorema de reconstrução.

4.2 Série Simulada através de Processo Auto-regressivo.

Para explicar melhor o objetivo deste exemplo tem que ser considerados os seguintes aspectos: sabe-se que o objetivo da reconstrução do Espaço de Imersão não é mais que a obtenção da dimensão D_0 do atrator que gerou a série. Como nas séries de carga de energia elétrica não existe nenhuma informação sobre dita dimensão, nunca se terá certeza se o Espaço de Imersão apontado pelo processo de reconstrução é correto. Sendo assim, fica impossível avaliar a qualidade dos resultados após aplicação de Takens. Para contornar esta situação é necessário utilizar sistemas com dimensão D_0 conhecida, por esta razão utilizamos um modelo ajustado de ordem 2, o qual posteriormente será utilizado para gerar via simulação uma série de carga. A partir da série simulada pelo sistema (agora conhecido) de ordem 2, é aplicado o processo de reconstrução do Espaço de Imersão.

A seguir, na Figura 4.2, são apresentados 2952 dados da conhecida série de carga EUNITE. Nesta seção pretende-se utilizar a mencionada série para ajustar um modelo auto-regressivo puro de ordem 2, AR(2) o qual posteriormente será utilizado para simular dados de uma série de carga. Ditos dados serão utilizados para realizar a reconstrução do espaço de imersão do sistema gerador. A vantagem de realizar este tipo de análise é o conhecimento a priori da dimensão do atrator D_0 que gerou a série, neste caso igual a 2.

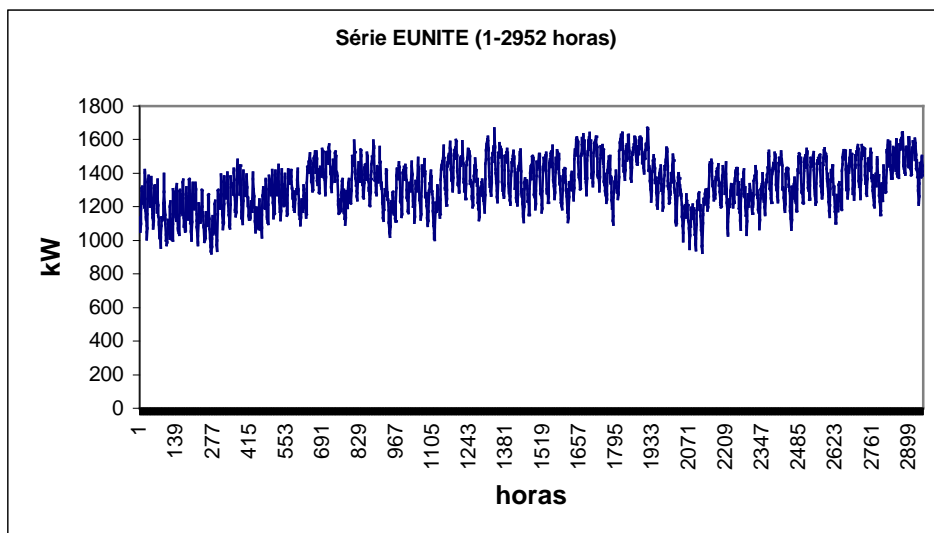


Figura 4.2. Série de carga EUNITE

4.2.1 Ajuste de uma Série de Carga Usando um Modelo AR(2)

O passo seguinte é ajustar os dados da Figura 4.2 a um modelo AR(2) para o qual será utilizado o programa Forecast Pro for Windows (FPW 3.0D). O resultado do ajuste do modelo pode ser observado na Figura 4.3 onde são apresentadas as estatísticas de saída junto com os parâmetros do modelo. A equação do modelo é apresentada a seguir:

$$Y(t) = 102,84 + 1,19 \cdot Y(t-1) - 0,27 \cdot Y(t-2) \quad (4.1)$$

A seguir, na Figura 4.4 é apresentada uma simulação de 744 dados utilizando o modelo definido na equação 4.1.

Forecast Model for CurvasdecargaEunite				
ARIMA(2,0,0)				
Term	Coefficient	Std. Error	t-Statistic	Significance

a[1]	1.1997	0.0177	67.9445	1.0000
a[2]	-0.2769	0.0177	-15.6855	1.0000
_CONST	102.8396			
Within-Sample Statistics				

Sample size	2952	Number of		
parameters	2	Standard deviation		
Mean	1332			
142.7				

Figura 4.3. Saída do programa FPW para modelo AR(2)

4.2.2 Reconstrução do Espaço de Imersão Série AR(2)

Na tabela 4.1 são apresentados os resultados da aplicação dos métodos não lineares visando obter as variáveis Passo de Reconstrução τ e Dimensão da Imersão m . A tabela 4.1 tem as seguintes características:

- a primeira coluna representa o intervalo de dados da série simulada utilizado para o cálculo das variáveis de reconstrução,
- a segunda coluna representa o cálculo da variável Passo de Reconstrução τ , usando para isto, o método da Informação Mútua. Observe que para cada intervalo existem três diferentes valores de τ , isto acontece porque foi pesquisada a sensibilidade

apresentada no cálculo de τ quando modificado o grau de detalhes, ou em outras palavras quando modificado o tamanho da partição utilizada na equação 3.14. Para realizar este cálculo foi utilizado o programa Visual Recurrence Analysis (VRA 4.7).

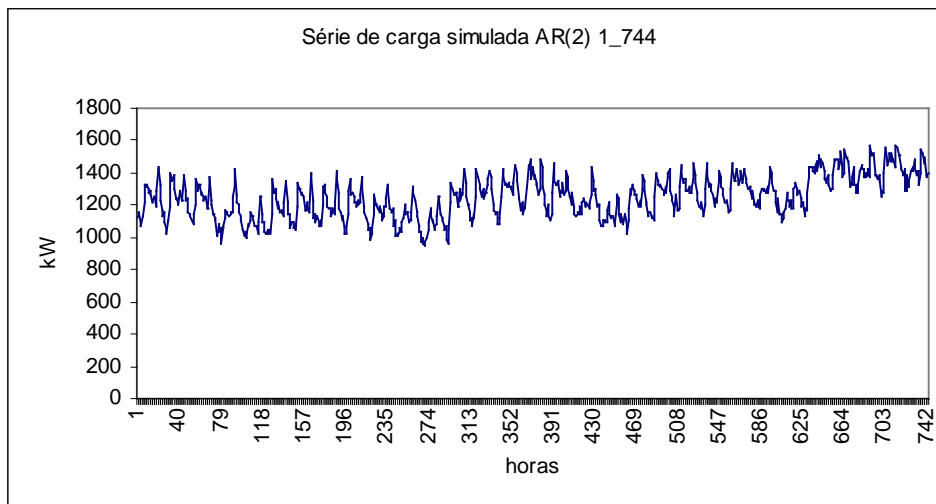


Figura 4.4. Simulação modelo AR(2)

- da terceira a quinta coluna é apresentado o resultado da aplicação do método dos Falsos Vizinhos Mais Próximos (FNN). Observe que este cálculo é realizado para cada valor de τ calculado na coluna 2. Como observado na seção 3.8.1, o cálculo dos falsos vizinhos depende de uma determinada tolerância R . Desta forma entre a segunda e quinta coluna são apresentados os resultados quando R varia de menor (maior precisão) a maior (menor precisão). Para o cálculo de FNN também foi utilizado o programa VRA 4.7;
- na sexta e última coluna são apresentados os resultados da utilização do método da Dimensão da Correlação (D_2) definido nos capítulos 2 e 3. Note que o cálculo m usando D_2 também foi realizado para cada um dos valores de τ . Para o cálculo de D_2 também foi utilizado o programa VRA 4.7.

Tabela 4.1. Variáveis para reconstrução série AR(2)

Dados da série	Informação mútua τ	FNN (precisão Max) m	FNN (precisão Med) m	FNN (precisão Min) m	D ₂ m
1-744	(τ) min= 7	29	13	13	5
	(τ) med= 6	37	36	31	8
	(τ) Max= 6	34	31	27	8
1-1440	(τ) min= 6	26	29	16	10
	(τ) med= 8	30	28	23	11
	(τ) Max= 9	30	30	16	13
1-2160	(τ) min= 10	29	25	24	10
	(τ) med= 12	28	26	18	11
	(τ) Max= 11	30	20	17	11
1-2952	(τ) min= 12	23	23	26	12
	(τ) med= 8	31	34	8	11
	(τ) Max= 12	23	23	26	12

4.2.3 Análise dos Resultados da Reconstrução Série AR(2)

Ao observar a tabela 4.1 é possível realizar os seguintes comentários:

- a Dimensão da Imersão calculada pelo FNN apresenta, na maioria dos casos, valores superiores ao cálculo realizado para essa mesma variável, pelo método D₂;
- em geral, o incremento no grau de precisão para o cálculo de FNN provoca, um incremento nos valores da Dimensão da Imersão m . Isto pode ser explicado devido ao fato de quando incrementa-se o grau de precisão, a variável R diminui, o que provoca na equação 3.23 uma tendência de aumento na quantidade de falsos vizinhos,
- o cálculo do Passo de Reconstrução τ , via Informação Mútua, não apresenta grandes variações em função do grau de detalhes (min, méd, max) para um mesmo intervalo de dados;
- observa-se que, em geral, quanto maior o intervalo de dados, maior o valor do passo de reconstrução τ .

Os resultados apresentados não deixam em evidência quais os valores τ e m devem ser escolhidos para realizar a reconstrução do espaço de imersão. No entanto, podem-se utilizar alguns os critérios para realizar a escolha:

- Takens apontou que a série deve ser infinita, sendo isto impossível, deve-se utilizar a maior quantidade de dados possíveis, isto é, o intervalo de 1-2952;

- para escolher o valor de τ , entre os três possíveis, faz sentido utilizar o máximo de detalhes. Assim desta forma o valor escolhido seria $\tau = 12$;
- a escolha da Dimensão da Imersão m é ainda mais complicada, já que temos disponíveis quatro alternativas três fornecidas pelo método FNN (23, 23 e 26) e uma conseqüência do cálculo de D_2 (12). Observando-se os resultados globais, isto é, todos os intervalos de dados, pode-se aplicar um critério de parcimônia, escolhendo-se para m os fornecidos pelo método D_2 . Aplicando-se a equação 3.4 ($m \geq 2D_0 + 1$), com D_0 conhecido e igual a 2, tem-se $m \geq 5$. Como o resultado apontado foi $m = 12$, verifica-se ser ele coerente com a definição dada por Takens.

4.2.4 Reconstrução do Espaço de Imersão da Série AR(2) Estacionada.

Nesta seção à semelhança da seção 4.2.2, será realizada a aplicação de métodos não lineares visando obter as variáveis Passo de Reconstrução τ e Dimensão da Imersão m . No entanto, a série produzida pelo modelo da equação 4.1 sofrerá um pré-processamento que consiste na eliminação, ou melhor dito, na diminuição das variáveis de tendência e sazonalidade claramente presentes na Figura 4.2. O objetivo deste experimento é avaliar o desempenho dos métodos quando aplicado este tipo de tratamento na série. A metodologia aplicada neste caso é bastante conhecida e apresentada com detalhes no Anexo C. Particularmente foram utilizados os critérios presentes na seção C.2 para tendência e C.3 para sazonalidade. A Figura 4.5 apresenta os resultados para um intervalo de 744 dados. Observe que na aplicação dos critérios foram perdidos 25 dados (1 na tentativa de retirar a tendência e 24 na tentativa de retirar a componente de sazonalidade)

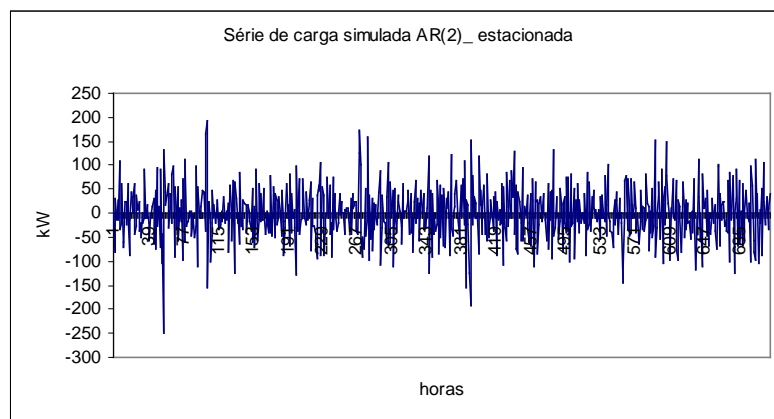


Figura 4.5. Simulação modelo AR(2) estacionada

4.2.5 Análise dos Resultados da Reconstrução Série AR(2) Estacionada.

Na tabela 4.2 são apresentados os resultados do cálculo das variáveis de reconstrução para a série apresentada na Figura 4.5. Sobre estes resultados é possível realizar os seguintes comentários:

- diferentemente da tabela 4.1, os resultados da dimensão da imersão calculada pelo FNN apresentam na grande maioria dos casos, valores inferiores ao cálculo realizado para essa mesma variável, pelo método D_2 ;
- da mesma forma que na tabela 4.1, o cálculo do passo de reconstrução τ , via Informação Mútua, não apresenta grandes variações em função do grau de detalhes (min, méd, max) para um mesmo intervalo de dados. Não obstante os valores de τ foram inferiores em todos os casos, quando comparados com os resultados da tabela 4.1;
- observa-se uma uniformidade nos valores para o passo de reconstrução τ , variando entre 2 e 4 em todos os intervalos analisados;
- os resultados apresentados pelo cálculo da Dimensão da Correlação D_2 não apresentam grandes variações quando comparados com a tabela 4.1.

Tabela 4.2. Variáveis para reconstrução série AR(2) estacionada.

Dados da série	Informação mútua τ	FNN (precisão Max) m	FNN (precisão Med) m	FNN (precisão Min) m	D_2 m
1-719	(τ) min= 2	1	4	1	8
	(τ) med= 3	5	7	3	5
	(τ) Max= 4	1	7	1	8
1-1415	(τ) min= 2	5	6	4	11
	(τ) med= 2	5	6	4	11
	(τ) Max= 2	5	6	4	11
1-2135	(τ) min= 2	5	6	1	6
	(τ) med= 2	5	6	1	6
	(τ) Max= 3	3	6	3	9
1-2927	(τ) min= 2	6	6	4	8
	(τ) med= 3	6	2	3	9
	(τ) Max= 4	6	7	10	10

Para realizar a escolha das variáveis adequadas para a reconstrução do espaço de imersão são utilizados critérios similares aos usados na seção 4.2.3:

- Takens apontou que a série tem que ser infinita, mas sendo isto impossível temos que utilizar a maior quantidade de dados possíveis, isto é, o intervalo de 1-2927,
- para escolher o valor de τ entre os três possíveis faz sentido utilizar o máximo de detalhes. Desta forma, o valor escolhido seria $\tau=4$,
- embora a Dimensão da Imersão m apresenta quatro alternativas disponíveis três fornecidas pelo método FNN (6, 7 e 10) e uma consequência do cálculo de D_2 (10), as mesmas são muito mais próximas quando comparadas com os resultados da tabela 4.1. Aplicando um critério de parcimônia o valor escolhido deve ser o fornecido pelo método FNN com $m=6$,

Da mesma forma que na seção 4.2.3 é utilizada a equação 3.4 ($m \geq 2D_0 + 1$), onde neste caso D_0 é conhecido e igual a 2. A restrição a cumprir é $m \geq 5$ e como o resultado apontado foi $m = 6$, é ele coerente com a definição dada por Takens. Este resultado é importante porque evidencia principalmente uma melhoria na capacidade de capturar a Dimensão da Imersão por parte do método FNN quando diminuída a não estacionariedade da série de carga.

4.3 Sistema de Lorenz

O segundo exemplo utilizado neste capítulo é o sistema de Lorenz que é um dos mais conhecidos sistemas dinâmicos determinísticos com características caóticas. O sistema de Lorenz pode ser definido usando o seguinte sistema de equações diferenciais:

$$\begin{aligned} \frac{dx_1}{dt} &= -\sigma x_1 + \sigma x_2 \\ \frac{dx_2}{dt} &= -x_1 x_3 + r x_1 - x_2 \\ \frac{dx_3}{dt} &= x_1 x_2 - b x_3 \end{aligned} \tag{4.2}$$

($\sigma = 10, b = 8/3, r = 28$)

Na Figura 4.6 é exibida a trajetória dinâmica do sistema de Lorenz no Espaço Fase original (que lembra as asas de uma borboleta) e na Figura 4.7 é apresentada uma das variáveis (x_1) que faz parte do Espaço Fase original do sistema.

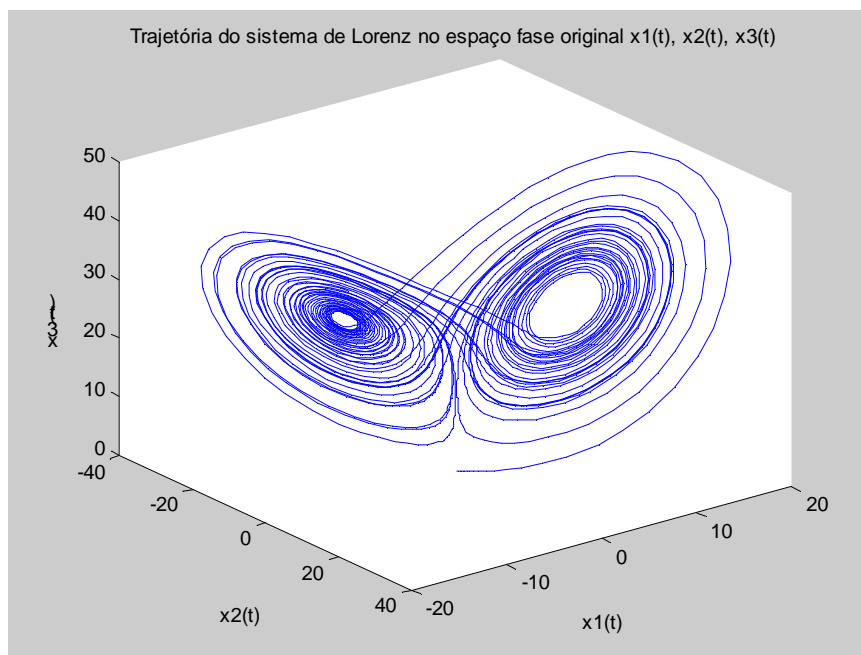


Figura 4.6. Atrator do sistema oscilador harmônico no Espaço Fase $x(t)$, $y(t)$, $z(t)$

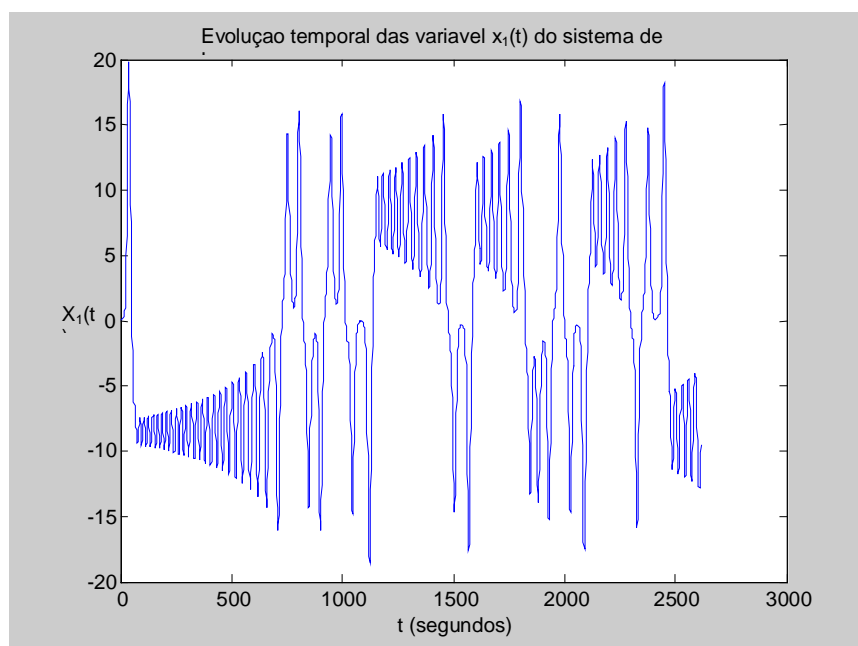


Figura 4.7. Variável $x(t)$ do sistema de Lorenz

4.3.1 Reconstrução do Espaço de Imersão do Sistema de Lorenz

Os resultados da aplicação das ferramentas para reconstrução do Espaço de Imersão são apresentados a seguir na tabela 4.3.

Tabela 4.3. Variáveis para reconstrução sistema de Lorenz.

Dados da série	Informação mútua τ	FNN (precisão Máx) m	FNN (precisão Med) m	FNN (precisão Mín) m	D_2 m
1-1000	(τ) min= 29	9	10	10	12
	(τ) med= 8	2	7	7	18
	(τ) Max= 8	2	7	7	18
1000-2000	(τ) min= 15	14	13	14	8
	(τ) med= 11	2	2	2	8
	(τ) Max= 11	2	2	2	8
1-2000	(τ) min= 14	2	10	10	12
	(τ) med= 10	2	3	3	12
	(τ) Max= 9	2	3	3	12
1-2613	(τ) min= 14	2	2	2	12
	(τ) med= 10	2	2	2	12
	(τ) Max= 9	2	2	3	10

4.3.2 Análise dos Resultados da Reconstrução do Sistema de Lorenz.

Para realizar a escolha das variáveis adequadas para a reconstrução do espaço de imersão são utilizados critérios similares aos usados na seção 4.2.5:

- Takens apontou que a série tem que ser infinita. Não obstante ser isto possível para o sistema de Lorenz, neste caso serão utilizados somente dados referentes ao intervalo de 1-2613;
- para escolher o valor de τ entre os três possíveis valores disponíveis, é coerente utilizar o máximo de detalhes, por este motivo o valor escolhido seria $\tau = 9$;
- a Dimensão da Imersão m apresenta quatro alternativas disponíveis três fornecidas pelo método FNN (2, 2 e 3) e uma consequência do cálculo de D_2 (10);
- da mesma forma que na seção 4.2.5, é utilizada a equação 3.4 ($m \geq 2D_0 + 1$), onde neste caso o valor D_0 para o sistema de Lorenz é conhecido e igual a 2,06 [26]. A restrição definida pela equação 3.4 inviabiliza a escolha dos valores calculados via método FNN, já que os valores de m tem que ser $\geq 5,12$, isto é no mínimo $m=6$,
- assim sendo, a única possibilidade de escolha de m é fornecida pelo método da Dimensão da Correlação que sugere $m=10$,

4.4 Conclusões

Como mencionado nos capítulos anteriores o Teorema de Takens permite a reconstrução do Espaço de Imersão a partir da definição de duas variáveis: m denominada Dimensão da Imersão e τ conhecida como Passo de Reconstrução, sendo que a análise realizada utiliza exclusivamente dados oriundos de uma série temporal gerada por um desconhecido sistema ou atrator. A aplicação destas duas variáveis na equação 3.7 permite a reconstrução do dito Espaço de Imersão.

Neste capítulo foi apresentado o processo de reconstrução utilizado nesta tese. A partir de uma série temporal foi definido o Passo de Reconstrução usando o método da Informação Mútua. Posteriormente foram realizados cálculos visando obter a Dimensão da Imersão com os dois critérios apresentados nos capítulos anteriores: método dos Falsos Vizinhos mais Próximos e método da Dimensão da Correlação. Também neste capítulo foram levantados alguns problemas de origem prática na hora da aplicação do Teorema de Takens.

A seguir algumas conclusões gerais:

1. para o processo de reconstrução, quanto maior a quantidade de dados utilizados melhor, já que o Teorema de Takens está baseado em séries infinitas;
2. para o cálculo da Informação Mútua, em todos os casos apresentados, a influência do grau de detalhes foi mínimo, isto é, observou-se uma estabilidade ou coerência nos resultados independente do grau de detalhes. Apesar disto, recomenda-se que no cálculo da Informação Mútua o critério a ser utilizado seja considerando o maior grau de detalhes;
3. nos exemplos apresentados, a reconstrução do Espaço de Imersão apresentou resultados coerentes com as restrições propostas por Takens ($m \geq 2D_0 + 1$);
4. quando utilizada a simulação da série de carga EUNITE previamente ajustada a um modelo AR(2), foi verificado que a dimensão da imersão calculada pelo método FNN apresenta, na maioria dos casos, valores superiores ao cálculo realizado para essa mesma variável pelo método D_2 . Entretanto, para a série com pré-processamento de tendência e sazonalidade a situação foi inversa, isto é, os valores apresentados pelo método FNN foram menores que os apresentados pelo método D_2 . Para o caso AR(2) com a série pré-processada, observa-se que a variável Dimensão da Imersão $m=6$, calculada pelo método FNN, está muito mais próxima do valor teórico de m , segundo Takens ($m \geq 5,12$). Sem processamento, FNN apontava a um valor de $m=23$, o que leva à conclusão de

que o processo de pré-processamento da série de carga é benéfico do ponto de vista de reconstrução, pois permite uma melhor captura do Espaço Fase gerador do fenômeno;

5. o resultado anterior é importante porque evidencia, principalmente uma melhoria na capacidade de capturar a Dimensão da Imersão por parte do método FNN quando diminuída a não estacionariedade da série de carga;
6. a diferença nos resultados de cada método para a variável Dimensão da Imersão, deve-se à especificidade da concepção de cada um deles;
7. os resultados apresentados pelo método D_2 não mudam consideravelmente com a aplicação de pré-processamento nos dados;
8. na hora de escolher o valor de m recomenda-se aplicar o critério de parcimônia;
9. para o caso do sistema AR(2) com a série pré-processada, embora a Dimensão da Imersão m apresenta quatro alternativas disponíveis três fornecidas pelo método FNN (6, 7 e 10) e uma consequência do cálculo de D_2 (10) as mesmas são muito mais próximas entre si, quando comparadas com os resultados da série AR(2) sem pré-processamento;
10. no caso do sistema de Lorenz, a Dimensão da Imersão m apresenta quatro alternativas disponíveis: três fornecidas pelo método FNN (2, 2 e 3) e uma consequência do cálculo de D_2 (10). Segundo Takens a equação 3.4 inviabiliza a escolha dos valores calculados via método FNN, já que os valores de m tem que ser ≥ 5.12 , isto é, no mínimo $m=6$. Assim, a única sendo a única possibilidade de escolha de m é fornecida pelo método da Dimensão da Correlação que sugere $m=10$,

No capítulo 5 a seguir, são apresentados alguns exemplos com aplicações da Teoria de *Wavelets* em sinais de tempo visando expor as utilidades que dita teoria tem para os objetivos deste trabalho uma vez que os sinais de energia elétrica são frequentemente compostos de características ruidosas, e o Teorema de Takens tem como uma de suas principais premissas, a ausência de ruído no sinal para sua correta aplicação.

CAPÍTULO 5

APLICAÇÕES DA TEORIA DE WAVELETS

5.1 Introdução

Neste capítulo são aplicados os conceitos da Teoria de *Wavelets* que estão detalhados no ANEXO D desta tese. Neste capítulo são ressaltadas as potencialidades e vantagens que esta teoria possui quando comparada a outras técnicas de análise de sinais disponíveis na literatura. Merecem destaque, a capacidade de detecção de anormalidades, erros e principalmente a possibilidade de identificação e extração de ruído no sinal. Neste capítulo serão apresentados exemplos que ressaltam algumas das características desta técnica.

As potencialidades para tratamento do ruído antes mencionadas, tornam-se muito atraentes para os objetivos deste trabalho que visa realizar a reconstrução do Espaço de Imersão de um sistema gerador desconhecido aplicando o Teorema de Takens, uma vez que os sinais de energia elétrica são freqüentemente compostos de características ruidosas e o Teorema de Takens tem como uma de suas principais premissas a ausência de ruído no sinal para sua correta aplicação.

5.2 Premissas Utilizadas na Aplicação de *Wavelets*

O primeiro passo para a aplicação de *wavelets* é o nível para o qual é feita a decomposição. Nos exemplos apresentados neste capítulo, o nível de decomposição utilizado é definido em função do objetivo da análise, desta forma o grau de decomposição pode chegar até o nível 5 em função das necessidades do exemplo.

Com respeito à escolha das *wavelets*, e como é detalhado no ANEXO D, existem muitas alternativas de escolha, particularmente neste capítulo são utilizadas *wavelets* da família Daubachies (dbN). Realmente não existe uma regra geral que indique a priori qual deve ser a *wavelet* utilizada e o nível ótimo de decomposição, sendo que alguns fatores que influenciam na escolha podem ser: o comportamento da série e o tamanho dos dados. A experiência do pesquisador também vai exercer um papel importante nesta escolha. Como já foi mencionado antes, neste capítulo são utilizadas *wavelets* tipo dbN, mais especificamente do tipo db3 que tem sido utilizadas por alguns autores com sucesso em aplicações de séries temporais de carga [X1].

No referente à aplicação da transformada discreta de *wavelets*, se faz necessário a escolha de algum método para a “extensão do sinal”. Na literatura existem diversos métodos para resolver este problema sendo que usualmente é preferível usar esquemas simples, estendendo o sinal nos contornos do mesmo. No entanto, isto envolve o cálculo de alguns poucos coeficientes extras em cada estágio do processo de decomposição, com o objetivo de permitir posteriormente uma reconstrução perfeita do sinal [X2]. Neste capítulo todos os exemplos utilizam o método *Zero-padding*, que considera que o sinal é zero fora dos contornos do sinal original.

5.3 Exemplo 1: Análise de um Sinal Elétrico

O objetivo deste exemplo é fornecer uma descrição estatística de uma série de consumo de energia elétrica usando as ferramentas de decomposição de *wavelets* apresentadas no ANEXO D, particularmente a análise multiescalar. Os dados correspondem a uma série com coleta de dados de 1 em 1 minuto. Na Figura 5.1 é possível observar 3 dias de amostragem, num total de 4320 minutos [50].

Sabe-se que uma curva de consumo de energia elétrica é formada pelos dados apurados por inúmeros medidores, que além da informação de consumo, carregam consigo características ruidosas. Em geral pode-se afirmar que na maioria das empresas elétricas as curvas de consumo são constituídas por aproximadamente 50% de carga industrial com participação de poucos consumidores com alto nível de consumo – e o restante 50% por numerosos consumidores com baixos níveis de consumo. As componentes da curva de carga produzidas pelo consumo industrial apresentam um perfil regular com mudanças de baixa frequência, já os demais consumidores tem um perfil de consumo altamente irregular, carregando consigo componentes de alta frequência.

A Teoria de *Wavelets* proporciona a possibilidade de realizar uma análise multiescalar de um sinal pela soma de sinais ortogonais que correspondem a diferentes escalas de tempo (*time scales*).

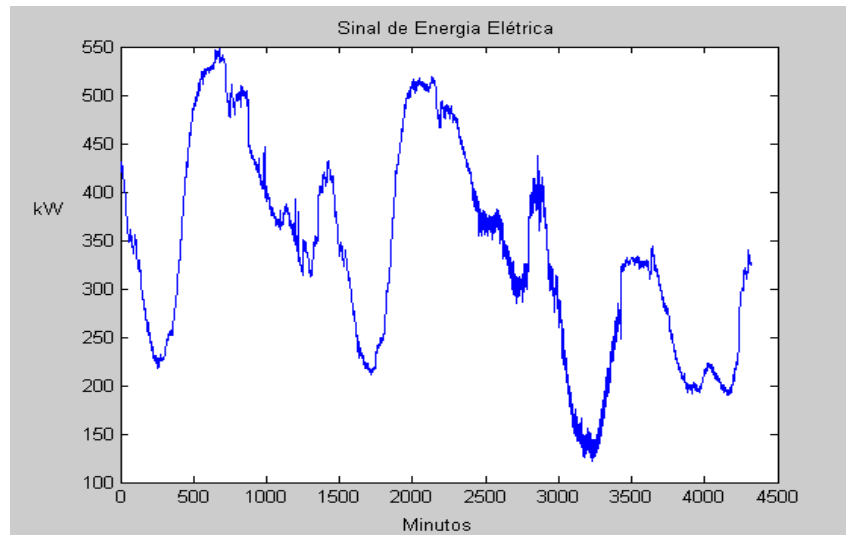


Figura 5.1. Sinal de Energia Elétrica

Observando a Figura 5.2 fica evidente que, para as observações 2400-3400, as componentes de ruído são particularmente altas. Com o objetivo de observar as características do sinal apresentado acima são aplicados os critérios de decomposição apresentados na Figura D.28 (do ANEXO D)

O primeiro nível de decomposição é apresentado na Figura 5.2, que mostra o nível de aproximação A1 e o nível de detalhes D1 junto com o sinal original S. O segundo nível de decomposição é apresentado na Figura 5.3. Os demais níveis de decomposição são apresentados nas três figuras subsequentes (5.4, 5.5 e 5.6)

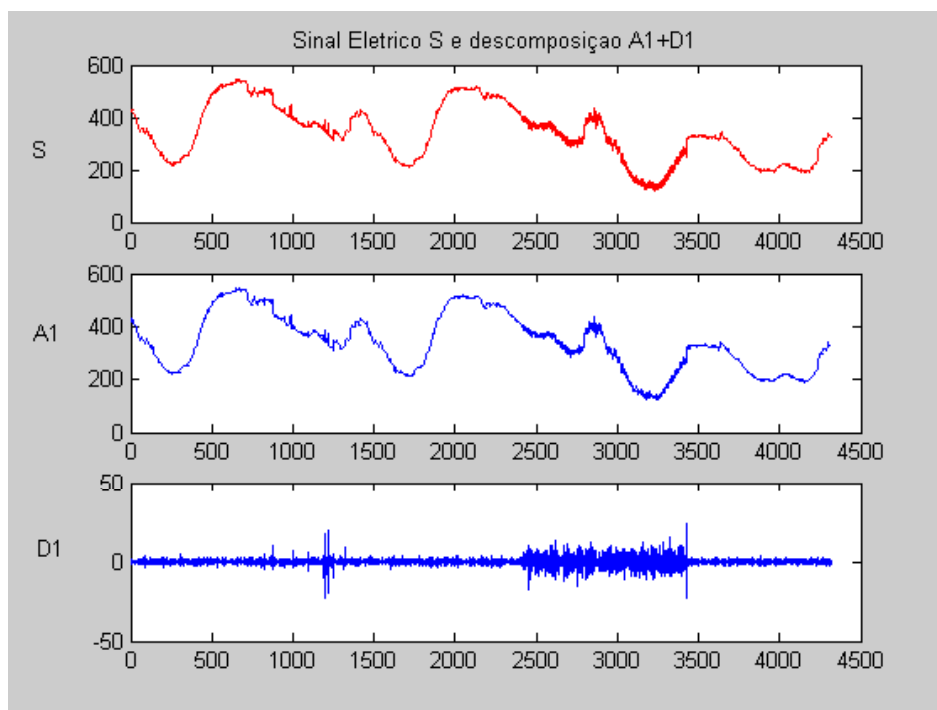


Figura 5.2. Sinal de Energia Elétrica. Decomposição de nível 1.

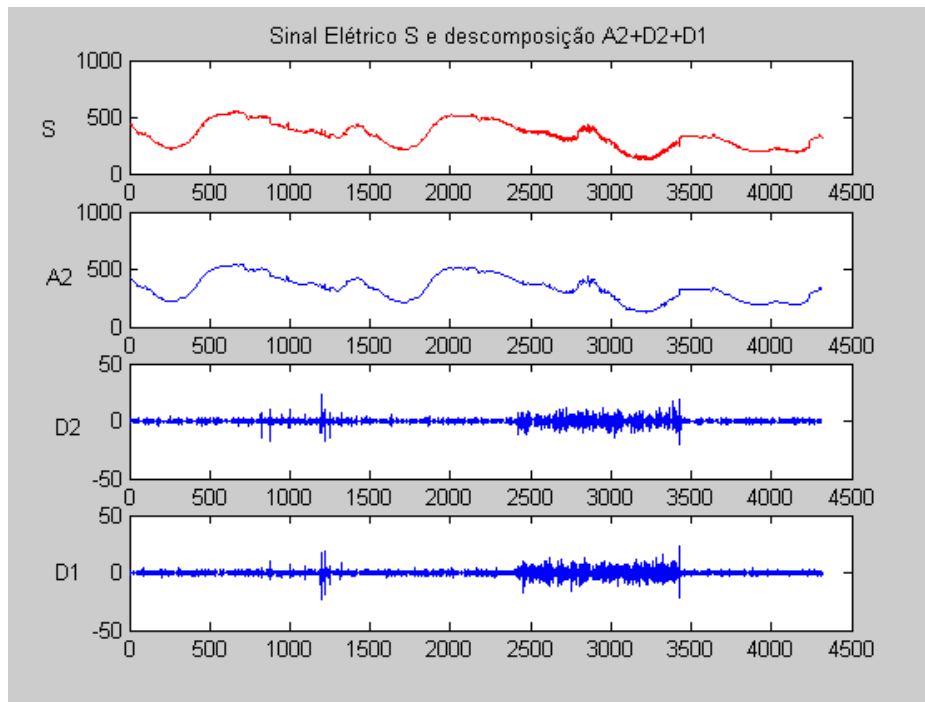


Figura 5.3. Sinal de Energia Elétrica. Decomposição de nível 2.

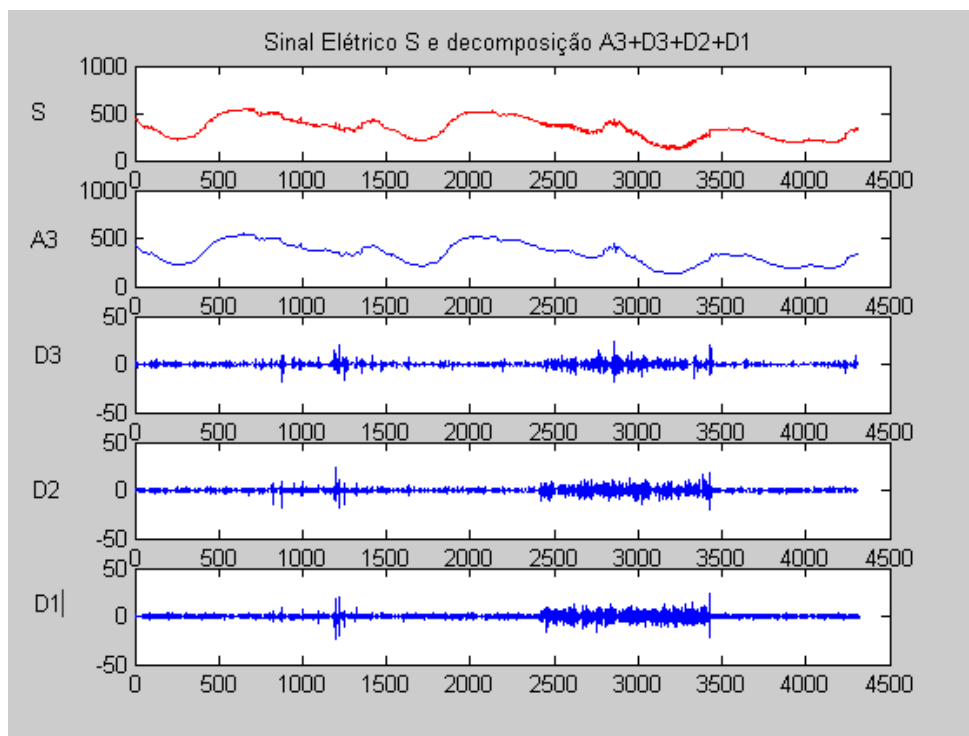


Figura 5.4. Sinal de Energia Elétrica. Decomposição de nível 3.

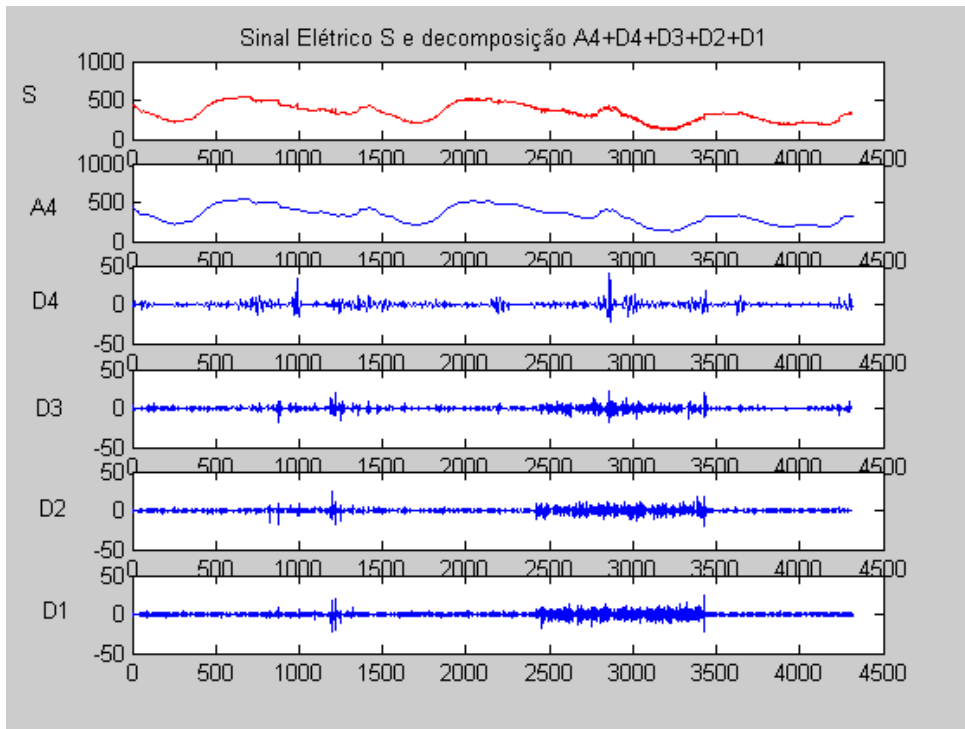


Figura 5.5. Sinal de Energia Elétrica. Decomposição de nível 4.

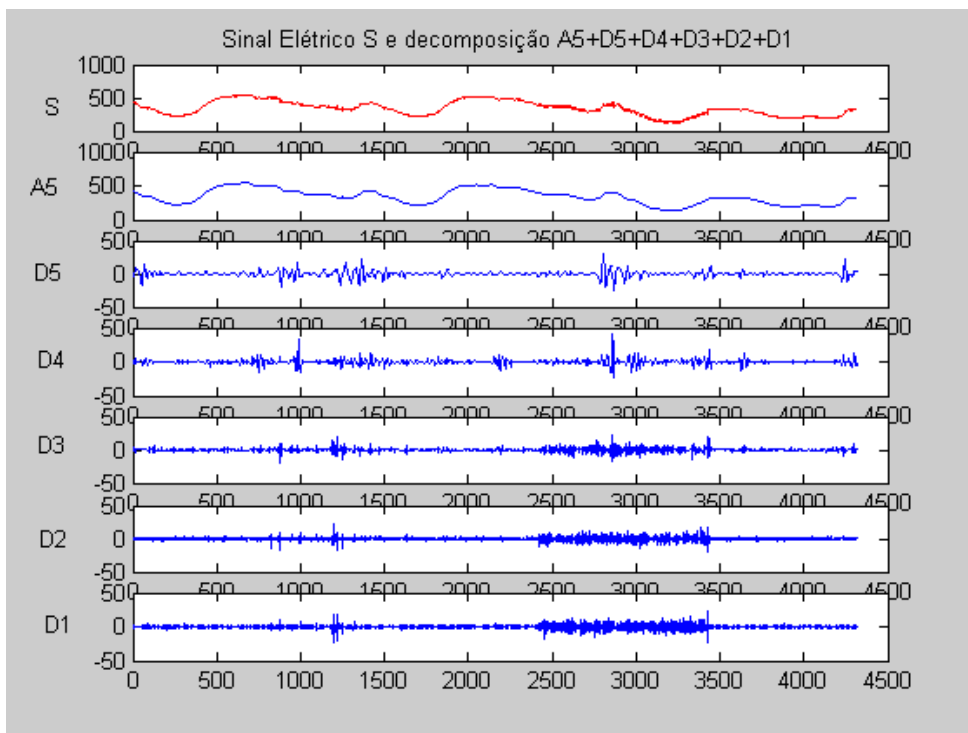


Figura 5.6. Sinal de Energia Elétrica. Decomposição de nível 5

Note que na Figura 5.7 são apresentadas as séries de aproximação A para cinco diferentes níveis de decomposição. Verifica-se que,, quanto maior o nível de decomposição, maior o grau de amortecimento ou suavização do sinal de aproximação.

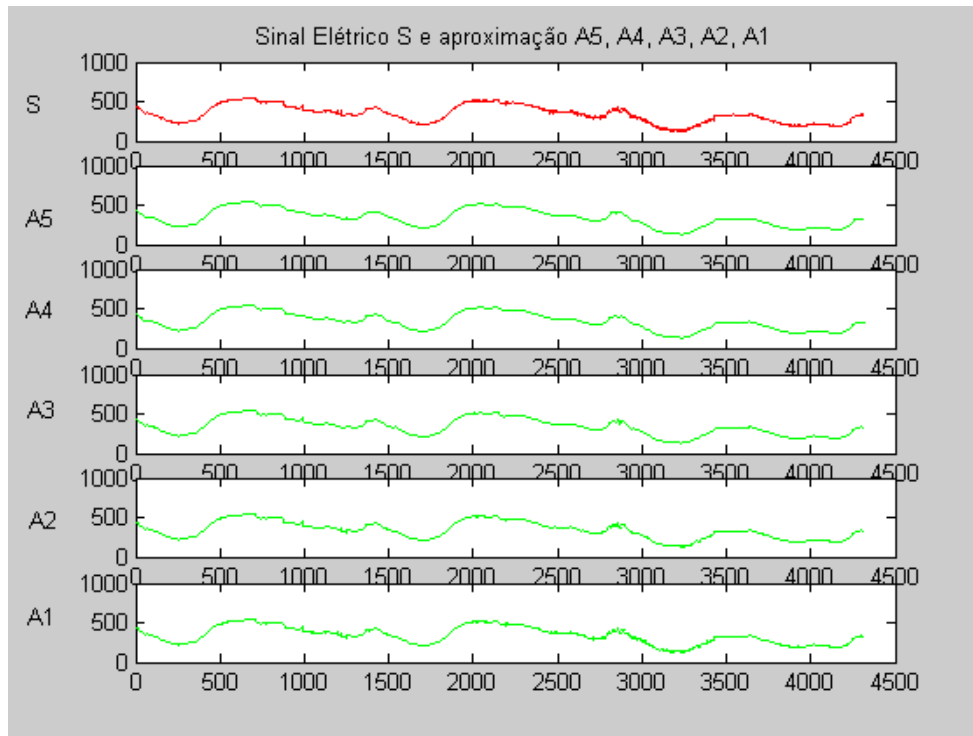


Figura 5.7. Sinal de Energia Elétrica. Sinal S e versões amortecidas em 5 níveis

A riqueza da informação fornecida por este tipo de metodologia permite realizar diversas análises como as citadas a seguir:

- identificação de erros na medição;
- análise de sinais por segmentos e de maneira independente;
- eliminação do ruído de um sinal;
- identificação de padrões regulares de um sinal através da análise dos detalhes;
- localização e eliminação de *outliers*;
- identificação de perda de dados e substituição, usando algum critério de interpolação.

Como exemplo de identificação de padrões regulares é apresentada a Figura 5.8, que contém 100 dados (minutos) entre o intervalo 1350-1450. Nela é possível observar mudanças bruscas no comportamento do sinal, possivelmente causadas pelo acionamento automático de aquecedores de água (países com baixa temperatura) ou de refrigeração climática (países com altas temperaturas). As mudanças abruptas acima mencionadas podem ser identificadas dos detalhes D2, D3, D4.

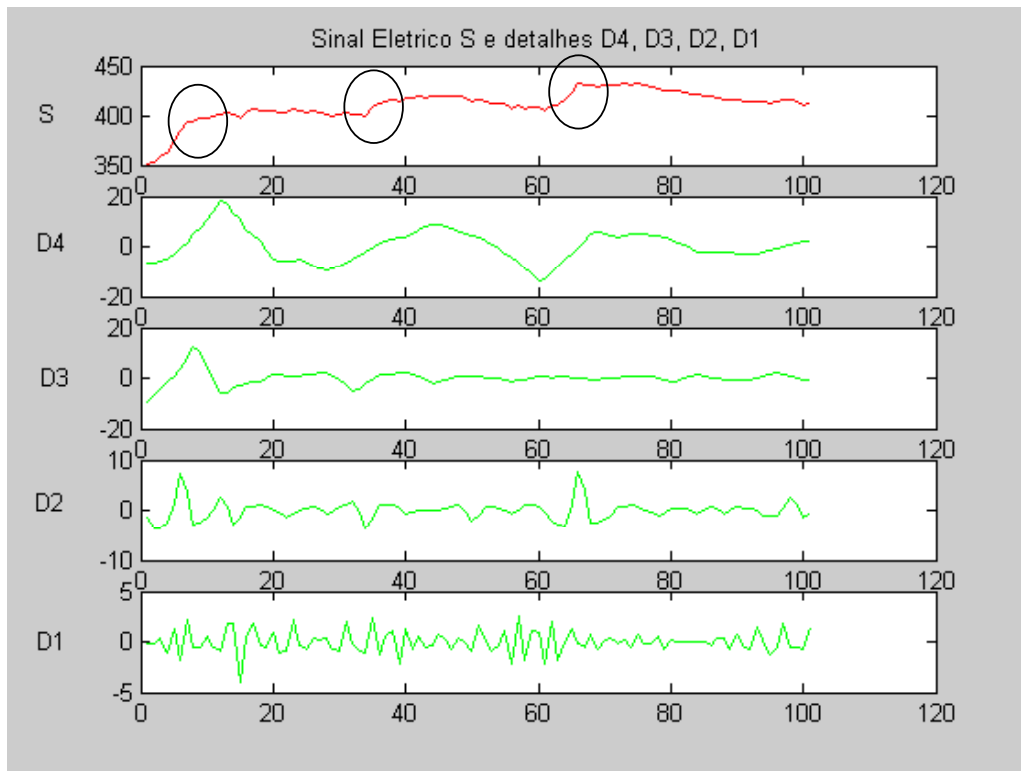


Figura 5.8. Sinal de Energia Elétrica (1350-1450)

Como exemplo de identificação de *outliers* é apresentada a Figura 5.9, que contém 100 dados (minutos) entre o intervalo 1150-1250, onde é possível observar dois *outliers*. Os efeitos podem ser vistos mais claramente quando analisados baixos níveis de detalhes como em D1 e D2, que estão em sincronia com o sinal S.

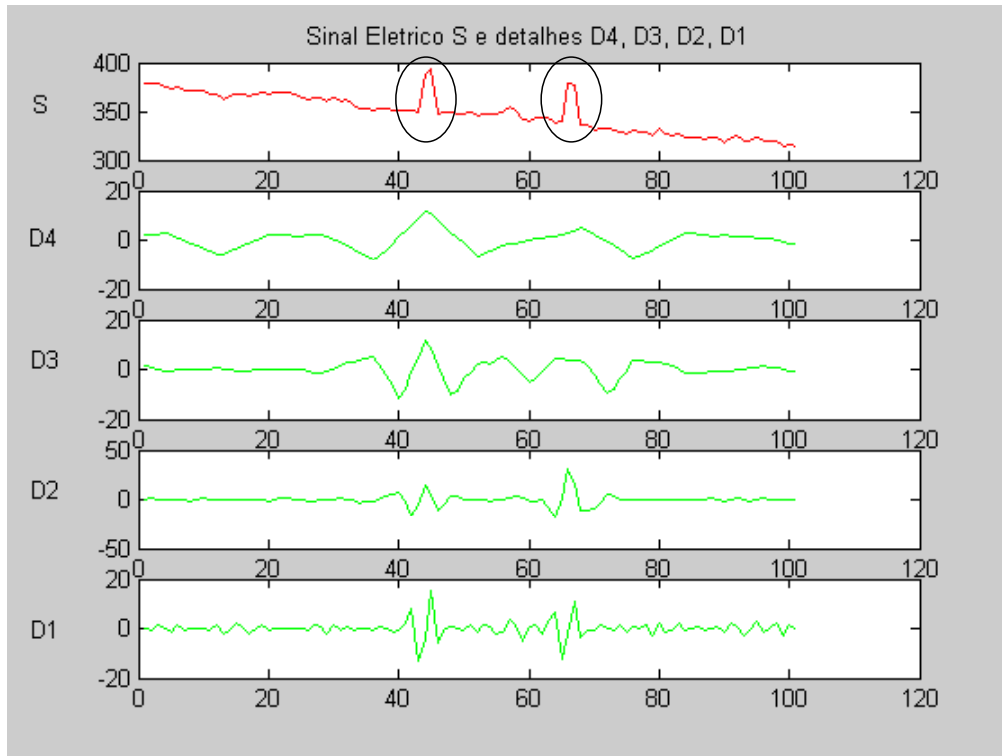


Figura 5.9. Sinal de Energia Elétrica (1150-1250)

Como exemplo de falha na medição é apresentada a Figura 5.10, que contém 600 dados (minutos) entre o intervalo 3000-3600, onde é possível identificar uma provável falha de medição. Os efeitos podem ser observados diretamente nos níveis de detalhes D1, D2 e D3.

Na Figura 5.11 são comparados intervalos de dados pertencentes a dois dias sucessivos: um com medição normal e um segundo corrompido por falhas na medição. É interessante notar a influência do grau de amortecimento na eliminação do ruído no sinal.

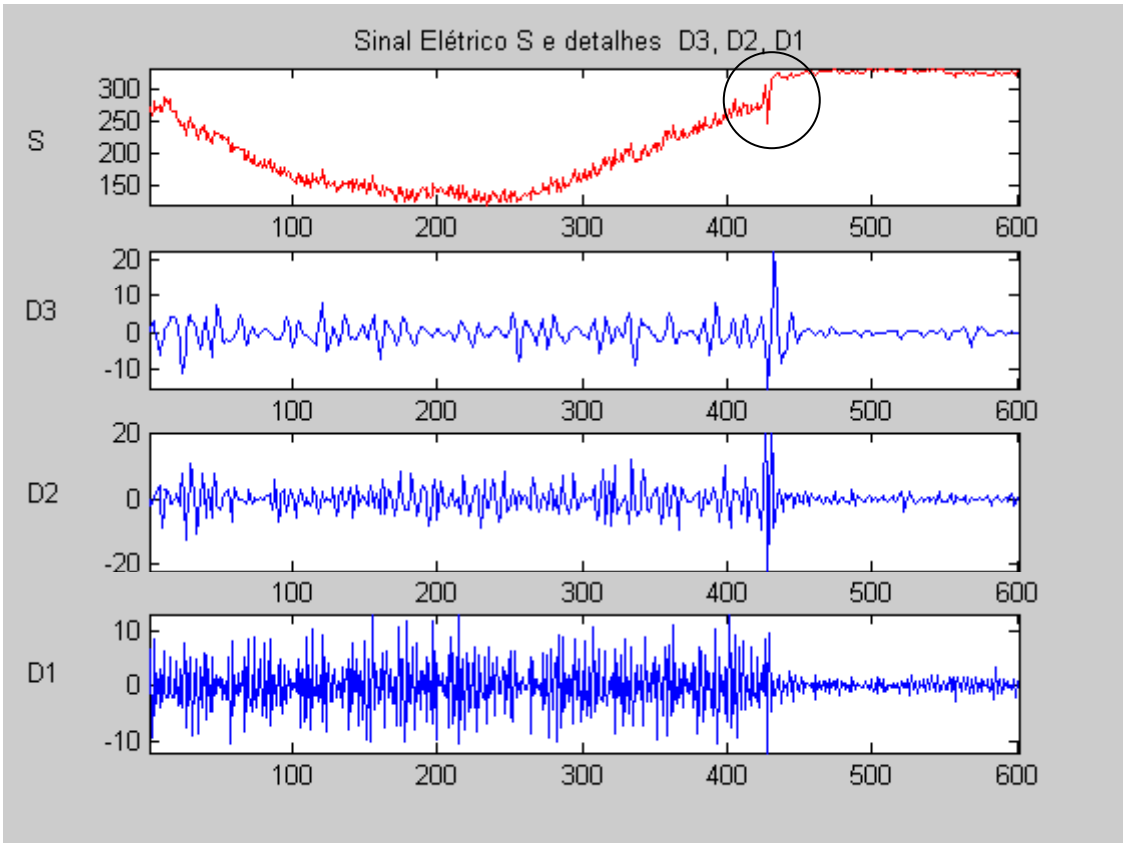


Figura 5.10. Sinal de Energia Elétrica (3000-3600).

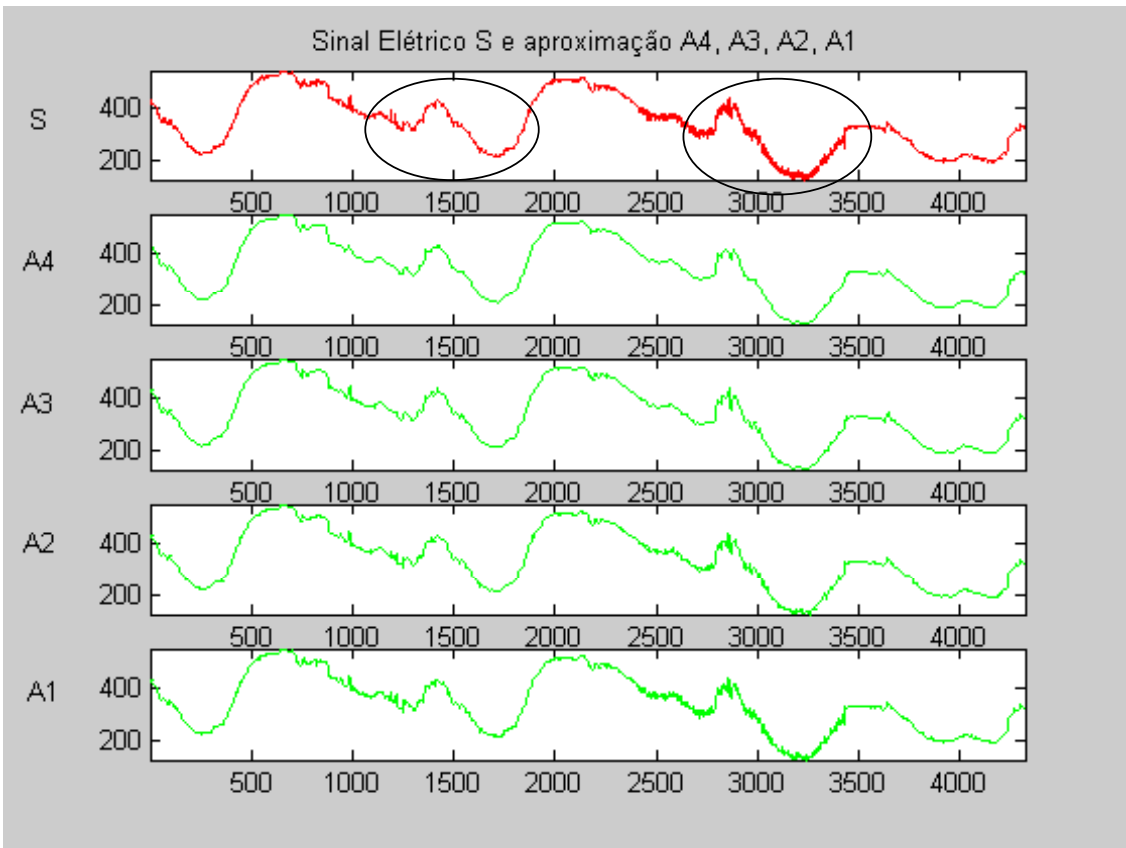


Figura 5.11. Sinal de Energia Elétrica (1-4320) e aproximações.

5.4 Exemplo 2: Extração de Ruído de um Sinal

Neste exemplo busca-se avaliar a habilidade da Teoria de *Wavelets* para extrair ruído de um sinal. Para isto utilizaremos o modelo AR(2) originalmente descrito e usado no capítulo 4, que é repetido a seguir na equação 5.3:

$$Y(t) = 102,84 + 1,19 \cdot Y(t-1) - 0,27 \cdot Y(t-2) \quad (5.3)$$

Agora será adicionado um sinal de ruído gaussiano no modelo determinístico definido acima. O sinal adicionado tem média zero e desvio padrão 50. A série simulada a partir do sistema acima descrito é mostrada na Figura 5.12 junto com a simulação de um sinal do ruído gaussiano. O resultado da adição do ruído pode ser observado na Figura 5.13.

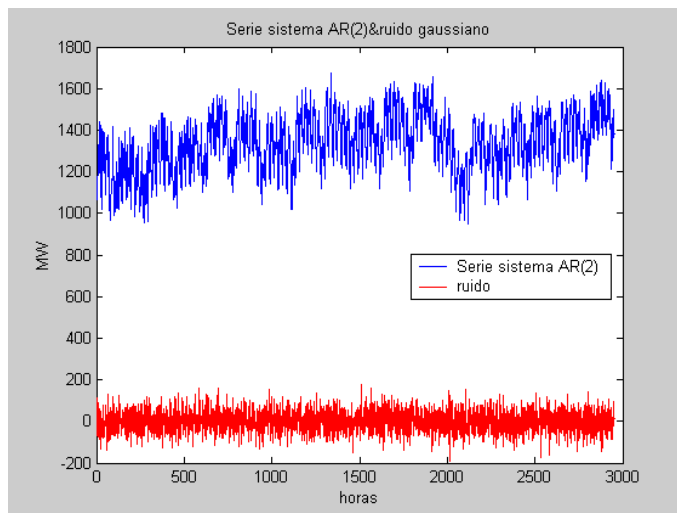


Figura 5.12. Série AR(2) e ruído gaussiano.

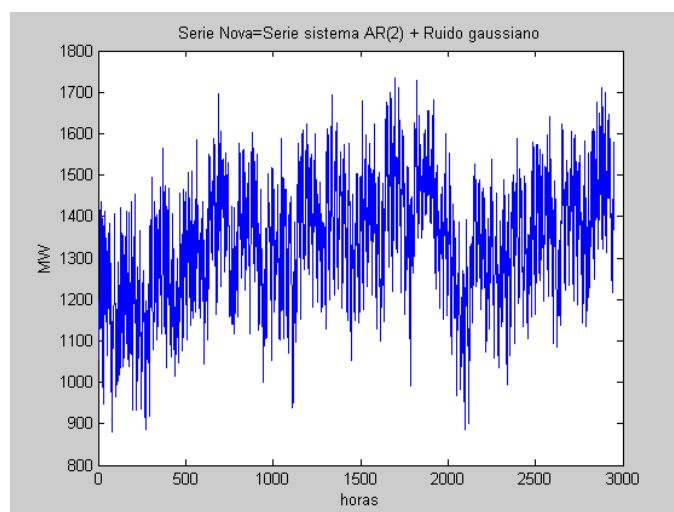


Figura 5.13. Série AR(2) e ruído gaussiano.

Na Figura 5.14 são exibidas as duas simulações: a do sistema AR(2) descrito no capítulo 4 junto com sua simulação, desta vez com a incorporação do ruído.

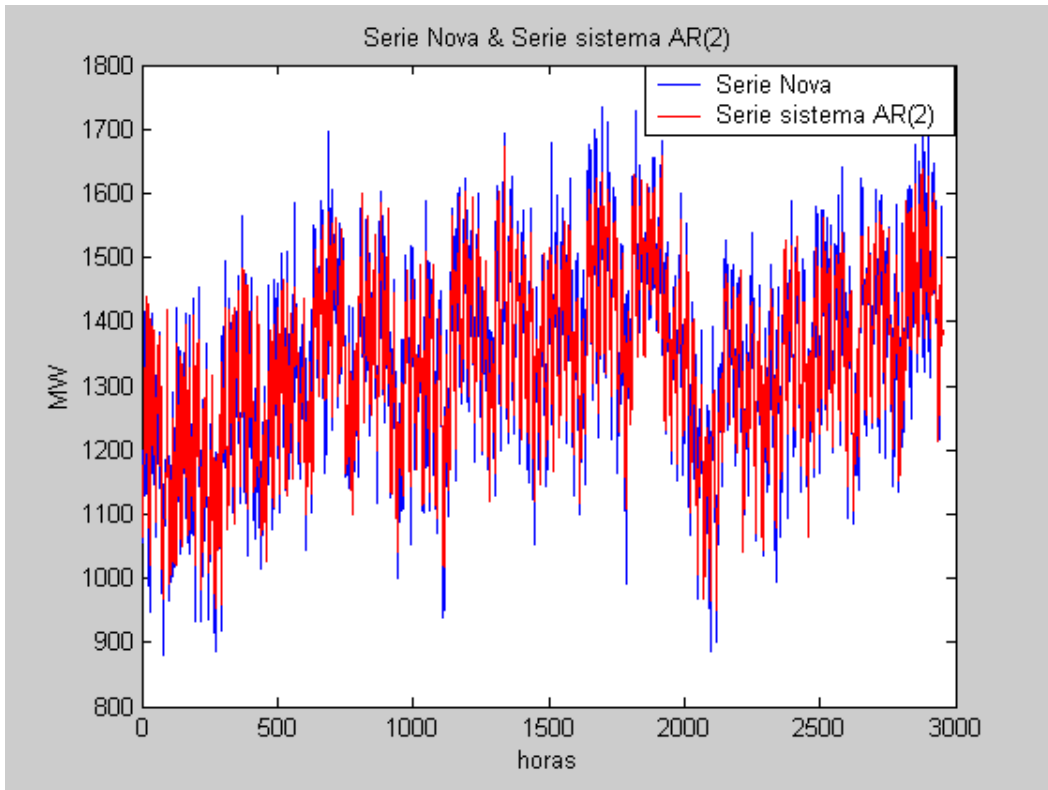


Figura 5.14. Compara simulações série AR(2) com série nova (AR(2)+ ruído).

Como já mencionado, o objetivo é extrair o ruído do sinal de tal forma que o sinal original seja recuperado da forma mais fiel possível. Simplificando bastante, o que se procura é que a diferença entre os dois sinais apresentados na Figura 5.15 seja minimizada. Para isto são aplicadas as ferramentas disponibilizadas pela Teoria de *Wavelets* previamente apresentadas neste capítulo.

Para começar, observe a Figura 5.16, onde são apresentados o sinal AR(2) mais a componente de ruído com cinco níveis de aproximação, A1 até A5. É evidente que quanto maior o nível de aproximação, maior o grau de amortecimento do sinal, isto porque no caso de A5, por exemplo, os detalhes D1 até D5 foram extraídos do sinal o que significa que nem altas frequências (D1) nem baixas frequências (D5) fazem mais parte do sinal. Na Figura 5.17 é exibida esta situação.

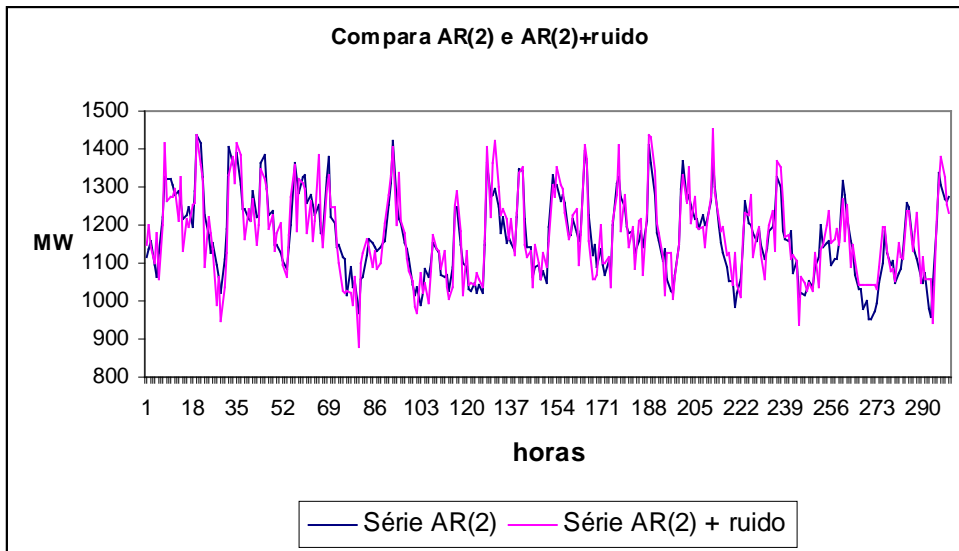


Figura 5.15. Comparação do sinal original AR(2) com a série AR(2)+ruído. Primeiros 300 dados.

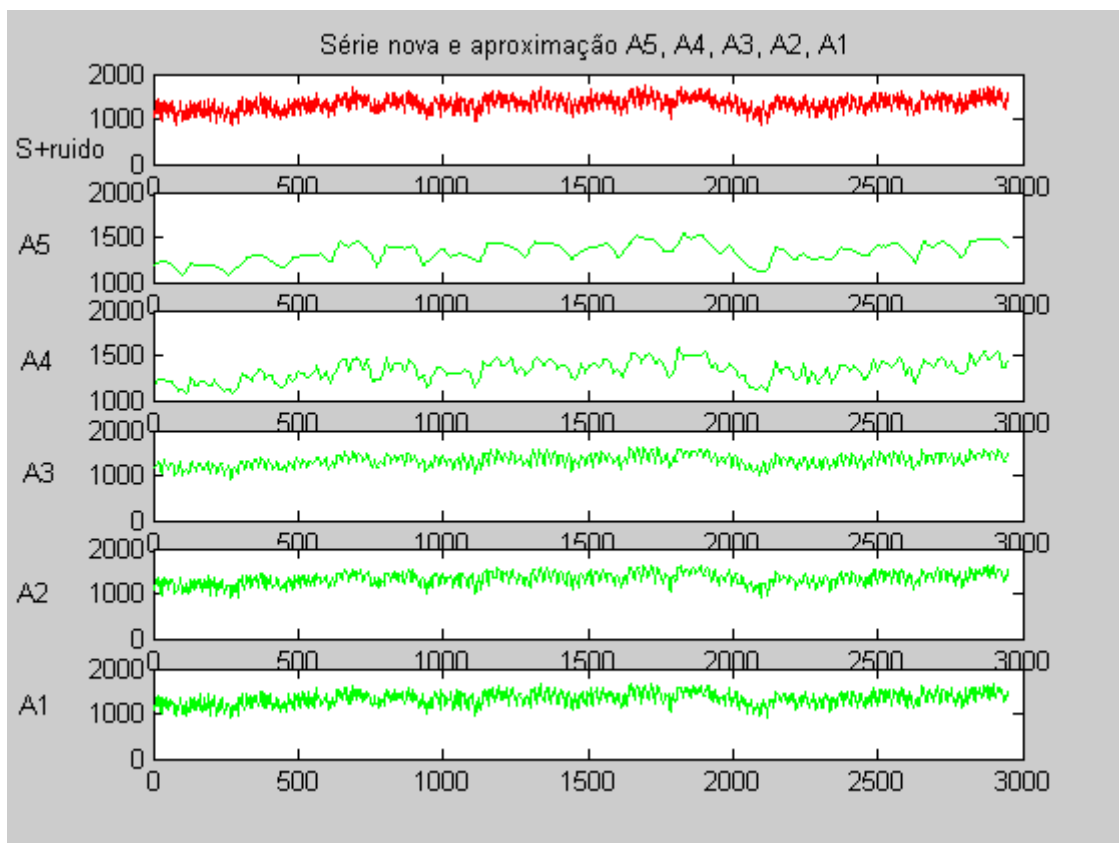


Figura 5.16. Comparação do sinal com ruído com os diferentes níveis de aproximação.

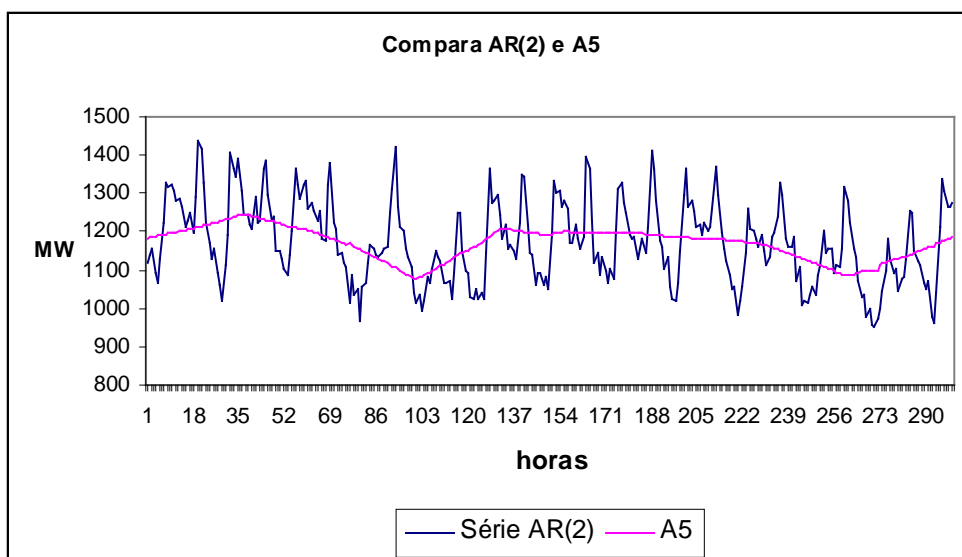


Figura 5.17. Comparação do sinal original AR(2) com a sinal de aproximação A5.

Primeiros 300 dados.

No outro extremo da análise, isto é, no primeiro nível, o sinal de aproximação A1 (Figura 5.18) é consequência da extração somente do nível de detalhes D1, que representa as maiores frequências na decomposição, mantendo-se ainda as frequências D2 até D4 no sinal. Vê-se que o resultado é muito mais aderente que o apresentado na Figura 5.17.

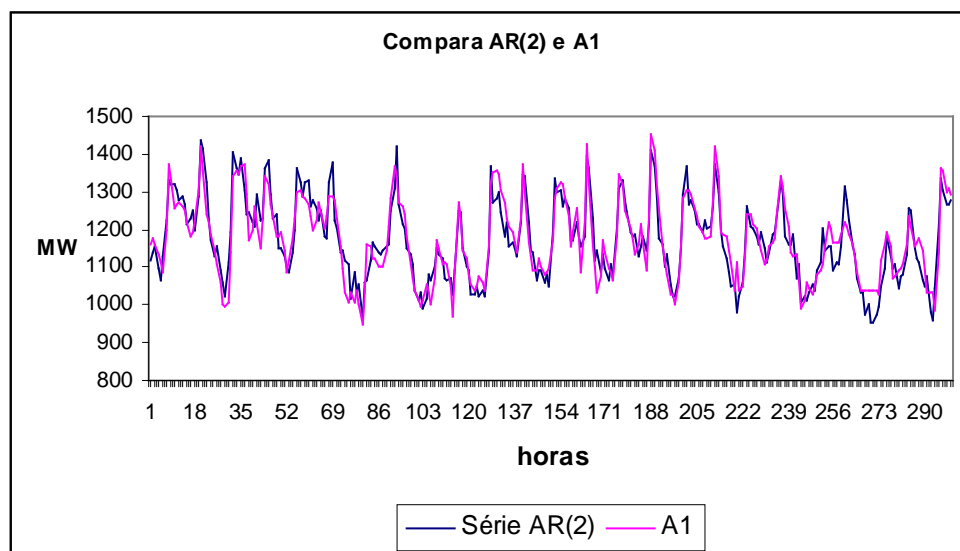


Figura 5.18. Comparação do sinal original AR(2) com o sinal de aproximação A1.

Primeiros 300 dados.

A chave reside em encontrar um equilíbrio entre a eliminação do ruído sem o comprometimento das informações úteis contidas nos sinais de alta frequência. Para decidir qual o nível de aproximação (entre A1 até A5) que apresenta as maiores

semelhanças com a série original AR(2) é aplicado o critério estatístico MAPE (*Mean Absolute Percentage Error*) que mede a qualidade do ajuste entre duas séries. Assim sendo, foi calculada dita estatística para todos os níveis de aproximação, sendo o menor valor o apresentado pelo sinal A2. Na Figura 5.19 podem ser observados os 300 primeiros dados da comparação entre o sinal original sem ruído e a componente A2 do sinal AR(2) com ruído.

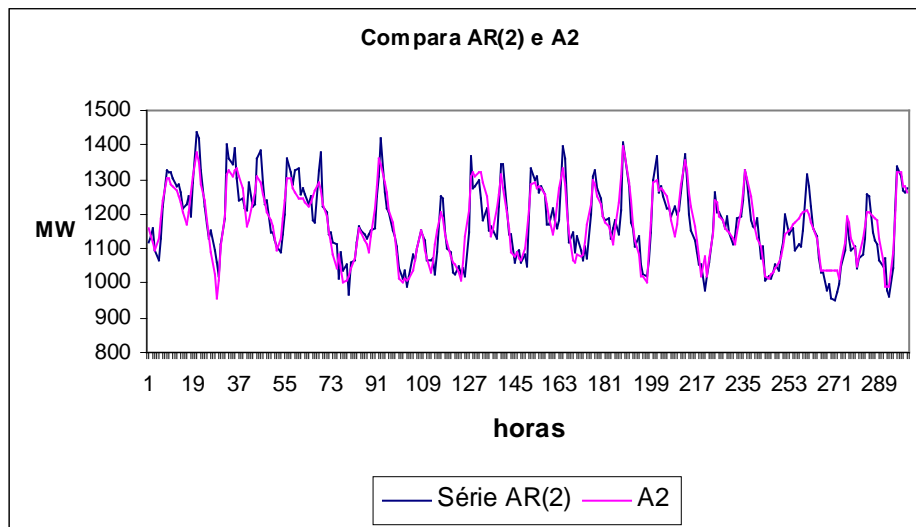


Figura 5.19. Comparação do sinal original AR(2) com o sinal de aproximação A2. Primeiros 300 dados.

Observa-se que o resultado acima apresentado significa que, para representar de maneira fiel a série original AR(2) a partir da série nova AR(2) com ruído, se faz necessário eliminar as componentes de frequência D1 e D2, mantendo as componentes de frequência D3, D4 e D5, vide Figura 5.20. O resultado desta análise é o sinal de A2, que é o que mais se assemelha à série AR(2) original.

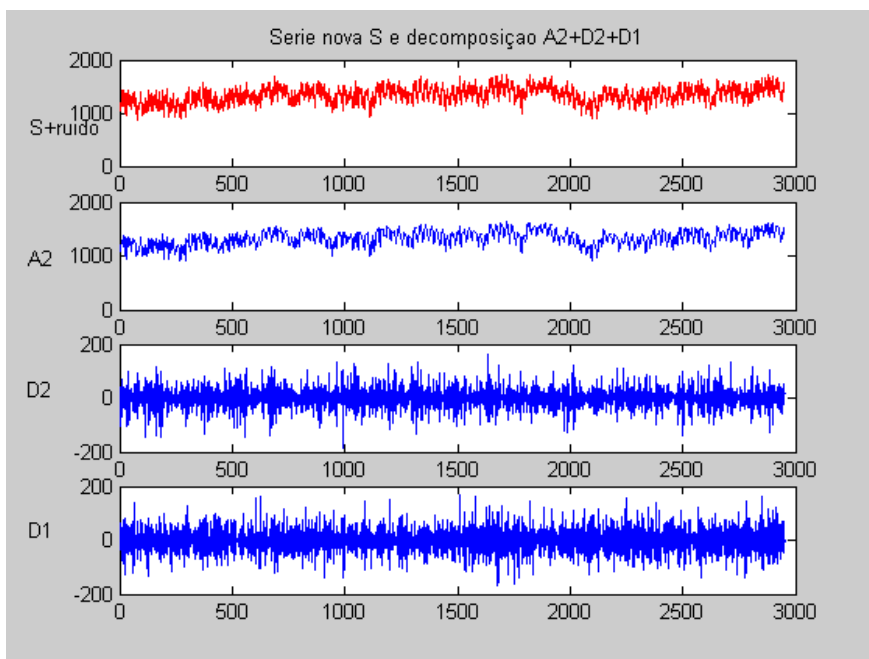


Figura 5.20. Série nova (AR(2)+ruído) e decomposição de segundo nível.

5.5 Conclusões

Neste capítulo são realizadas diversas aplicações da Teoria de *Wavelets* ressaltando suas potencialidades e vantagens. Foram aplicados conceitos apresentados no ANEXO D, como a Transformada Contínua de *Wavelets*, Transformada Discreta de *Wavelets*, filtragem, aproximação e detalhes. Também foram ressaltadas a capacidade de detecção de anormalidades, erros e principalmente a possibilidade de identificação e extração de ruído no sinal. Foi explorado um exemplo de extração de ruído a partir de um sinal determinístico contaminado com um ruído gaussiano. Esta potencialidade é muito importante para o objetivo deste trabalho que visa realizar a reconstrução do Espaço de Imersão de um sistema gerador desconhecido aplicando o Teorema de Takens, que possui como premissa a utilização de séries temporais com ausência de ruído.

No Capítulo 6 a seguir, são apresentados os resultados da aplicação de todos os conceitos introduzidos nesta tese. São apresentados diversos exemplos de reconstrução de Espaço de Imersão e sua avaliação considerando testes de previsão de carga de curto prazo.

CAPÍTULO 6

APLICAÇÕES e RESULTADOS

6.1 Introdução

Como mencionado anteriormente, o objetivo deste trabalho é realizar a reconstrução do Espaço Fase ou Espaço de Imersão, a partir da aplicação do Teorema de Takens em séries de carga de energia elétrica. Para isto, são aplicadas as técnicas exibidas nos capítulos 2 e 3: método dos Falsos Vizinhos mais Próximos, método da Dimensão da Correlação e método da Informação Mútua. São testados alguns critérios que podem melhorar o desempenho das técnicas de reconstrução, como são a diminuição das componentes de ruído via aplicação da Teoria de *Wavelets*, e a diminuição das componentes de tendência e sazonalidade presentes nas séries. Também neste capítulo é avaliado o efeito que a resolução dos dados causa na qualidade dos métodos de reconstrução.

São apresentados exemplos utilizando quatro séries de carga, três delas bastante utilizadas na literatura de previsão e uma série de carga de uma distribuidora de energia elétrica nacional.

6.2 Características das Redes Utilizadas

O padrão de entrada é formado pelos valores passados das variáveis explicativas e a saída desejada é o valor a observação da série temporal no horizonte de previsão. A Figura 6.1 ilustra como for construído o conjunto de treinamento:

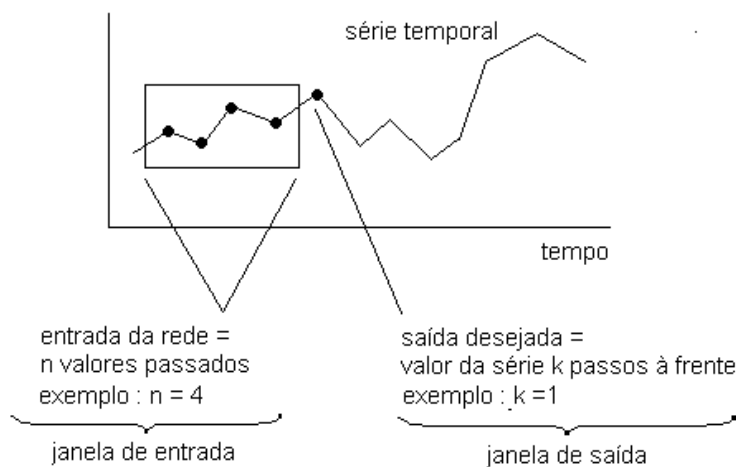


Figura 6.1. Montagem do conjunto de treinamento

O treinamento da rede consiste em mover as janelas de entrada e saída ao longo de toda série temporal. Cada par de janelas entrada/saída funciona como um padrão de treinamento e deve ser apresentado repetidas vezes até que o algoritmo de aprendizado alcance a convergência. Neste capítulo é realizada uma previsão $k=1$ passos à frente, onde as entradas (n valores passados) a serem utilizadas são fornecidas pelos métodos de extração de variáveis ou reconstrução do Espaço de Imersão. Uma vez treinada a rede é realizada uma previsão de 1-24 passos à frente

São os seguintes os passos para treinamento das redes utilizadas:

- o processo de treinamento visa realizar uma previsão de demanda serial de 1-24 passos a frente, utilizando uma amostra (entrada/saída) para treinamento;
- o processo de previsão é avaliado utilizando uma amostra pré-definida para testes com 25% dos dados. Os resíduos da previsão dos 24 dados são calculados e armazenados;
- o processo de treinamento e previsão se repete N iterações (dias), sendo que em cada uma das iterações são previstos mais 24 dados;
- no final de cada iteração, a amostra para treinamento é modificada, eliminando-se os 24 dados mais “antigos” da amostra de treinamento e incorporando-se os 24 dados mais “antigos” da amostra para testes, isto é, produz-se um deslocamento (de 24 passos) das duas amostras;
- após N iterações (dias) são avaliados os resultados de previsão *out sample* de todo o período, usando as estatísticas: MAPE, MAD, RMSE, Erro Máximo e Erro Relativo;
- todas as redes usadas neste trabalho possuem a mesma arquitetura: 3 camadas, sendo que a camada oculta está composta de 3 neurônios com função de ativação tipo tangente hiperbólica e um neurônio tipo linear na camada de saída. A quantidade de entradas depende dos resultados do processo de reconstrução;
- em todos os casos foi definido um critério que divide a série de dados disponível em duas amostras: uma para treinamento com 75% dos dados e uma segunda para testes, com o 25% restante;
- em todos os casos o critério de parada é a quantidade de épocas de treinamento que neste caso foi fixada em 500 épocas;

O processo descrito acima visa simular uma situação real em que diariamente é necessário realizar a previsão de carga do dia seguinte (próximas 24 horas). Assim todo dia é treinada uma rede neural, que incorpora na amostra de treinamento os últimos 24 dados realizados e elimina os 24 mais antigos, mantendo sempre constante o tamanho da amostra de treinamento. Cabe ressaltar que o procedimento descrito acima é idêntico para outros níveis de resolução: 30 minutos, 15 minutos e 5 minutos, substituindo somente a quantidade de dados 24 pela de 48, 96 e 288 respectivamente. Em todos os casos as redes neurais possuem as mesmas características de arquitetura, número de neurônios, número de camadas, critérios de parada, algoritmo de treinamento, etc.

A utilização de critérios simples para a definição da arquitetura e treinamento das redes, deve à necessidade de comparar “numa mesma base” os efeitos da reconstrução do Espaço de Imersão sugeridos pelos métodos de extração de variáveis utilizados. Ao fixar as características das redes o desempenho das mesmas do ponto de vista de previsão, pode ser atribuído de forma isolada à qualidade (ou a deficiência) do processo de reconstrução utilizado.

É claro que se trata de uma simplificação, porque a resolução de um problema de previsão de carga via rede neural é altamente dependente da arquitetura da rede, e que esta pode fazer a diferença quando se avalia a qualidade da previsão. No entanto, desde o ponto de vista prático, seria impossível testar diferentes arquiteturas em combinação com diferentes variáveis de entrada oriundas do processo de reconstrução. No final não seria possível saber se os resultados são atribuíveis à: arquitetura, ao processo de treinamento ou à qualidade da escolha das entradas da rede.

O uso de uma rede padrão de três camadas com três neurônios na camada oculta e um neurônio na camada de saída responde à premissa de que, sob determinadas condições, uma RNA simples pode aproximar qualquer função não linear com um determinado nível de precisão[4].

6.3 Estatísticas Utilizadas

A seguir são descritas as estatísticas utilizadas para avaliar a qualidade da previsão.

1. RMSE (“Root Mean Square Error”). Raiz Quadrada da Média dos Resíduos.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n R_i^2}{n}} \quad (6.1)$$

Onde n: tamanho da amostra
 R_i é o resíduo no instante i

2. MAD (“Mean Absolute Deviation”). Média dos Módulos dos Resíduos

$$MAD = \frac{\sum_{i=1}^n |R_i|}{n} \quad (6.2)$$

3. MAPE (“Mean Absolute Percentage Error”). Média dos Módulos dos Erros Percentuais.

$$MAPE = \frac{\sum_{i=1}^n |R_i| / X_i}{n} \quad (6.3)$$

Onde X_i : Dados da série histórica

4. Erro Máximo Percentual (EMAX)

$$EMAX = \text{Max} \left(|R_i| / X_i \right)_{i=1,n} \quad (6.4)$$

5. Erro Relativo (ER)

$$ER = \frac{\sum_{i=1}^n (R_i / X_i)}{n} \quad (6.5)$$

6.4 Diminuição das Componentes de Tendência e Sazonalidade

Utiliza-se um procedimento para eliminar a tendência e sazonalidade das séries de carga antes da aplicação dos métodos de reconstrução do Espaço de Imersão, cujos detalhes são fornecidos no ANEXO C. Em geral neste trabalho é utilizado o procedimento descrito a seguir:

$$\text{Seja } Z_t = T_t + S_t + a_t \quad (6.6)$$

onde Z_t : uma série de consumo de energia elétrica

T_t : componente de tendência da série de carga

S_t : componente de sazonalidade da série de carga

a_t : componente de ruído da série de carga

Para extração da componente de tendência é aplicada a seguinte expressão:

$$X_t = Z_t - Z_{t-1} \quad (6.7)$$

Já para extrair a componente de sazonalidade aplica-se a seguinte expressão:

$$Y_t = X_t - X_{t-lag} \quad (6.8)$$

onde X_t : é a série sem componente de tendência

lag : é a defasagem.

Ressalta-se que a variável lag varia em função da resolução da série utilizada, assim neste trabalho ela toma os seguintes valores:

lag : 24, para séries com resolução de 1 hora;

lag : 48, para séries com resolução de 30 minutos;

lag : 96, para séries com resolução de 15 minutos;

lag : 288, para séries com resolução de 05 minutos;

6.5 Minimização do Ruído Usando *Wavelets*

Como já foi mencionado no Capítulo 5, o primeiro passo para a aplicação de *wavelets* é o nível para o qual é feita a decomposição. Nos exemplos apresentados neste capítulo, o nível de decomposição utilizado é nível 1.

Com respeito à escolha das *wavelets*, e como é detalhado no ANEXO D, existem muitas alternativas de escolha, neste capítulo ao igual que no capítulo anterior também são utilizadas *wavelets* da família Daubachies (dbN), especificamente do tipo db3 que tem sido utilizadas por alguns autores com sucesso [X1].

No referente à aplicação da transformada discreta de *wavelets*, se faz necessário a escolha de algum método para a “extensão do sinal”. Neste capítulo e ao igual que no Capítulo 5, todos os exemplos utilizam o método *Zero-padding*, que considera que o sinal é zero fora dos contornos do sinal original.

6.6 Procedimentos Utilizados para Reconstrução do Espaço de Imersão

Nesta seção são apresentados os procedimentos de reconstrução do Espaço de Imersão testados nesta tese, cuja explicação é reforçada com um diagrama.

A. Procedimento 1. É realizada a reconstrução do Espaço Fase a partir de uma série de carga, sendo que a qualidade da reconstrução é avaliada em função do resultado das estatísticas de previsão fornecidas por um modelo neural. Os detalhes de treinamento e testes da RNA foram fornecidos na seção 6.2. Na Figura 6.2 é apresentado um diagrama do processo.

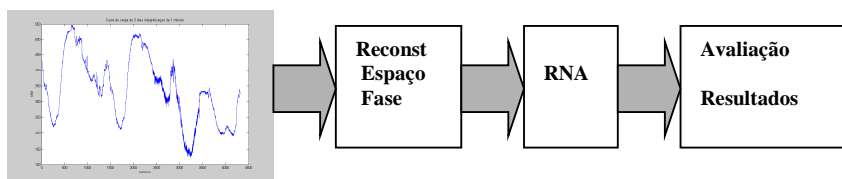


Figura 6.2. Diagrama do Procedimento 1

B. Procedimento 2. É realizada a reconstrução do Espaço Fase a partir de uma série de carga, sendo que a qualidade da reconstrução é avaliada em função do resultado das estatísticas de previsão fornecidas por um modelo neural. A diferença com o Procedimento 1 está no processamento da série anterior à reconstrução, objetivando a eliminação das componentes de tendência e sazonalidade. Os detalhes de treinamento e testes da RNA foram fornecidos na seção 6.2. Na Figura 6.3 é apresentado um diagrama do processo.

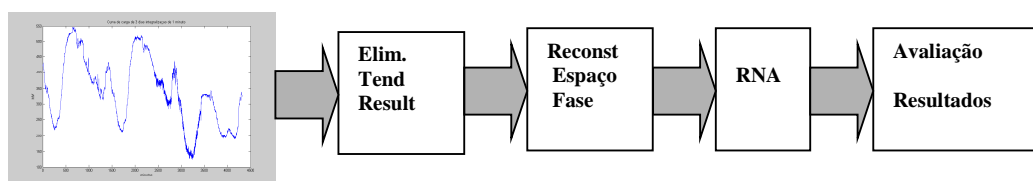


Figura 6.3. Diagrama do Procedimento 2

C. Procedimento 3. É realizada a reconstrução do Espaço Fase a partir de uma série de carga, sendo que a qualidade da reconstrução é avaliada em função do resultado das estatísticas de previsão fornecidas por um modelo neural. A diferença com o Procedimento 1 está no fato que anterior à reconstrução do Espaço Fase é realizada uma decomposição dos sinais (aproximação e detalhes) usando *Teoria de Wavelets*. Os detalhes de treinamento e testes da RNA foram fornecidos na seção 6.2. Na Figura 6.4 é apresentado um diagrama do processo.

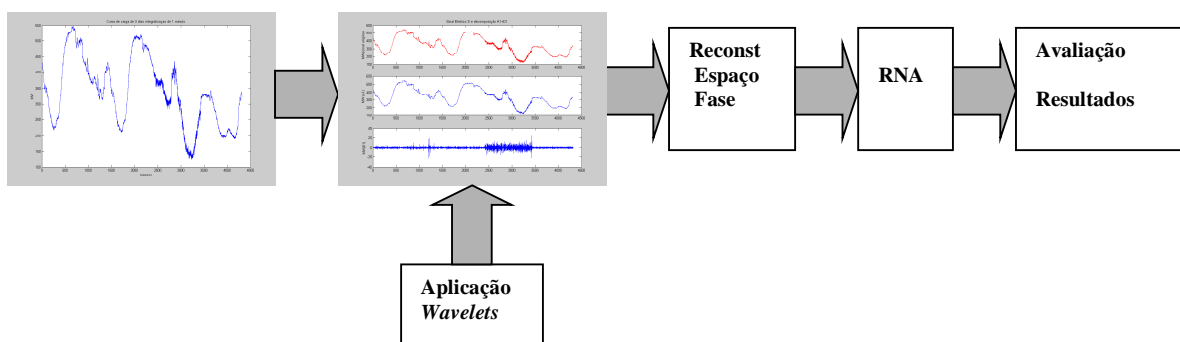


Figura 6.4. Diagrama do Procedimento 3

D. Procedimento 4. É realizada a reconstrução do Espaço Fase a partir de uma série de carga, sendo que a qualidade da reconstrução é avaliada em função do resultado das estatísticas de previsão fornecidas por um modelo neural. Antes do processo de reconstrução de Espaço Fase são aplicados dois procedimentos em série: decomposição dos sinais (aproximação e detalhes) usando Teoria de *Wavelets* e processamento para eliminação de tendência e sazonalidade. Os detalhes de treinamento e testes da RNA foram fornecidos na seção 6.2. Na Figura 6.5 é apresentado um diagrama do processo.

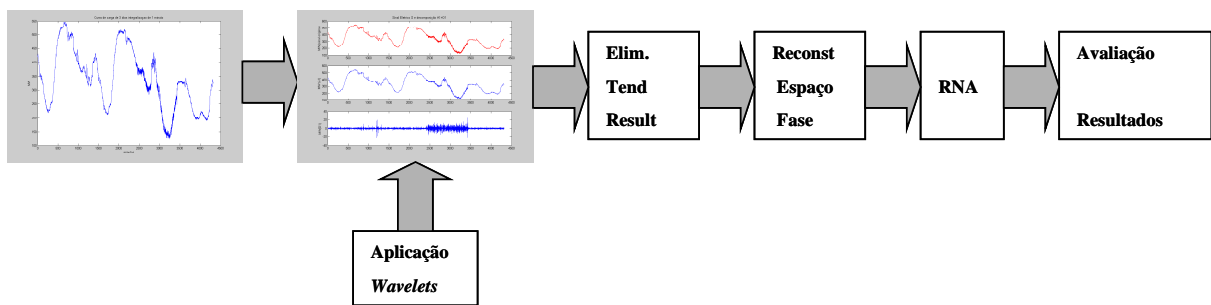


Figura 6.5. Diagrama do Procedimento 4

E. Procedimento de Reconstrução do Espaço Fase. A reconstrução do Espaço Fase ou Espaço de Imersão aplicada neste trabalho consiste primeiramente na aplicação do método da Informação Mútua para definição do Passo de Reconstrução. Posteriormente, para a definição da Dimensão da Imersão, são utilizadas duas alternativas - método dos Falsos Vizinhos mais Próximos e Dimensão da Correlação. As duas variáveis são aplicadas no Teorema de Reconstrução de Takens. Na Figura 6.6 é apresentado um diagrama do processo descrito acima.

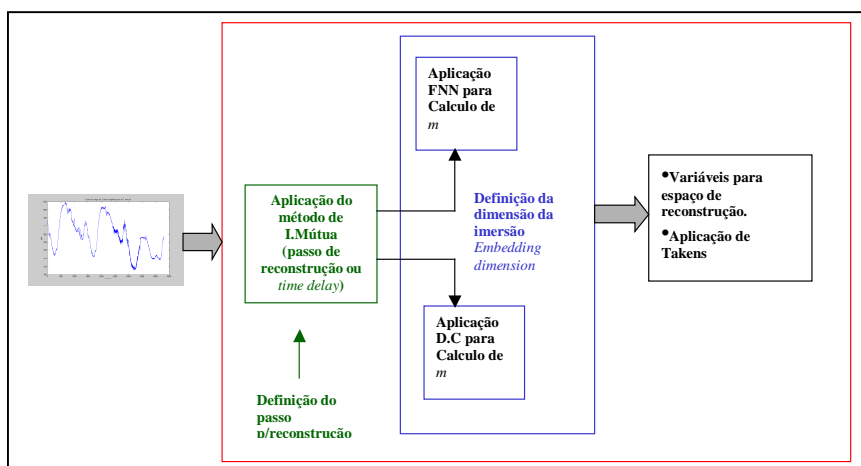


Figura 6.6. Diagrama do Procedimento de reconstrução.

6.7 Exemplo utilizando a série de carga EUNITE

Neste exemplo é realizada a aplicação das técnicas de reconstrução não lineares, definidas no capítulo 3, em uma série de carga real. Posteriormente são testados os Espaços de Imersão sugeridos, usando estes como entradas em modelos neurais.

6.7.1 Série de carga EUNITE

Os dados aqui utilizados correspondem a observações de carga e temperatura de uma distribuidora de energia elétrica europeia que foram utilizados numa competição promovida em 2001 pelo *European Network on Intelligent Technologies for Smart Adaptive Systems* mais conhecida como EUNITE (<http://neuron.tuke.sk/competition>). Nas tabelas 6.1 e 6.2 abaixo são apresentados o volume de dados disponíveis de carga e temperatura respectivamente. Embora os dados originais foram apurados em intervalos de 30 minutos, para a realização do exemplo elencado nesta seção da tese, são utilizadas informações horárias de carga. Desta forma a quantidade de dados utilizada é de 17.520.

Tabela 6.1. Dados de carga.

Período	Intervalo de apuração	Quantidade de dados
01/01/1997 até 31/01/1999	30 m	35.040

Tabela 6.2. Dados de temperatura.

Período	Intervalo de apuração	Quantidade de dados
01/01/1997 até 31/12/1998	diário	761

6.7.2 Análise dos dados

Na Figura 6.7 são apresentados os dados de consumo mensal e temperatura média mensal para todo o período de dados disponível. Observa-se claramente que na área de concessão da referida distribuidora existem dois períodos bem definidos: um de temperaturas baixas, algumas negativas e um período com temperaturas na faixa dos 20 graus Celsius, sendo que o pico de consumo acontece no inverno. Para reforçar a idéia de dois períodos durante o ano pode-se observar a Figura 6.8 que mostra os valores diários de temperatura do ano 1997. Como o objetivo deste trabalho é realizar previsão de curto prazo, é conveniente trabalhar com amostras separadas, uma para o inverno e outra para o verão. Assim para os testes deste capítulo, são considerados dados do período de 01/10/97 ao 31/03/1998 que fazem parte do período mais frio do ano.

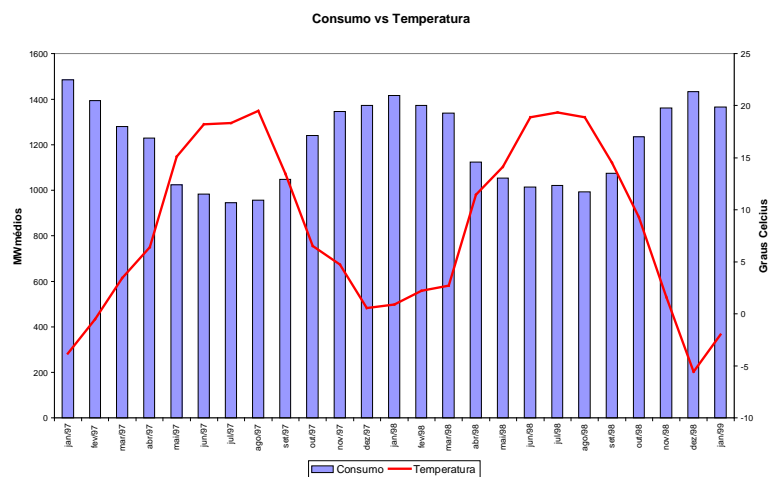


Figura 6.7. Consumo mensal e Temperatura (média mensal).

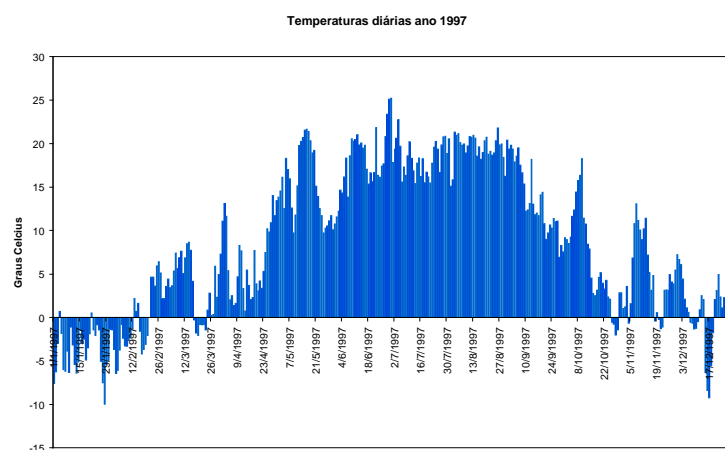


Figura 6.8. Evolução da temperatura diária ano 1997.

A amostra utilizada possui 4.368 dados de carga horários e 182 dias de temperatura máxima. Na Figura 6.9 é apresentada a série de carga EUNITE para este período de observações e na Figura 6.10 tem-se a mesma série, mas desta vez, com

dados processados aplicando os critérios definidos no Anexo D, que visam a diminuição ou eliminação das componentes de tendência (T) e sazonalidade (S). Cabe ressaltar que o processamento para eliminação da tendência e sazonalidade da Figura 6.10 foi realizado com 2.952 dados, isto é, considerando-se apenas o período entre o 01/10/97 e 31/01/98. Estes dados, ao serem processados, converteram-se em 2.902 observações.

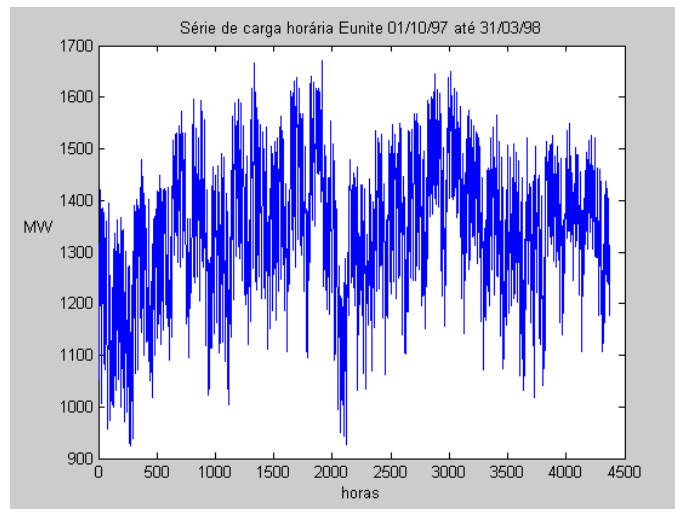


Figura 6.9. Série de carga EUNITE - período frio.

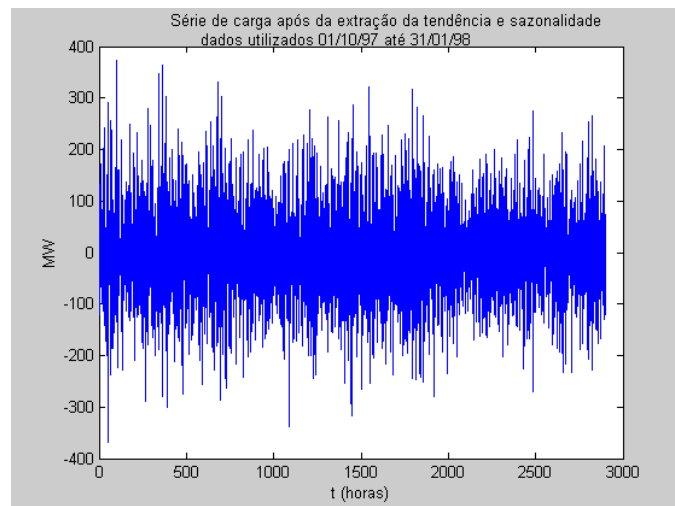


Figura 6.10. Série de carga EUNITE após eliminação de T e S.

6.7.3 Reconstrução do Espaço Fase Série EUNITE

A primeira tentativa de reconstrução é realizada usando as observações 1-2952. São aplicados os métodos de análise não linear na série EUNITE para a totalidade dos dados, sejam estes processados ou não. Os resultados são exibidos nos gráficos a seguir.

Na Figura 6.11 pode-se observar que o método de Informação Mútua, para a série processada, aponta para um valor de defasagem de $\tau=3$, que corresponde ao primeiro mínimo.

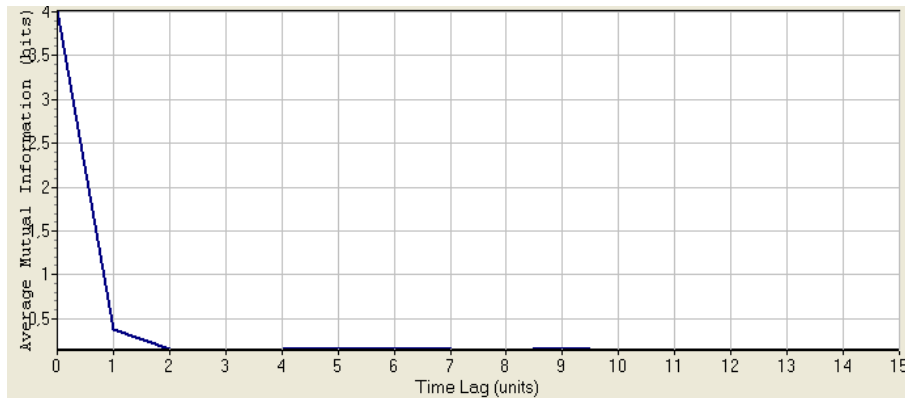


Figura 6.11. Cálculo da Informação Mútua com a série processada $\tau=3$.

O cálculo da Dimensão da Correlação para a série processada é mostrado na Figura 6.12 e denota uma certa dificuldade na definição de qual o valor de m para o qual a curva satura. De fato, visualmente observam-se dois níveis de saturação: o primeiro (entre $m=5$ e $m=7$) e o segundo em $m=14$.

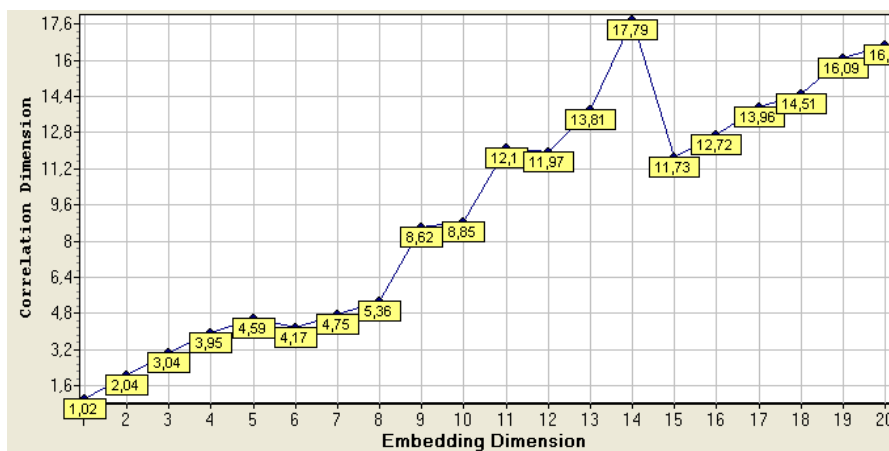


Figura 6.12. Cálculo da Dimensão da Correlação com a série processada.

A Figura 6.13 mostra os resultados após a aplicação do Método dos Falsos Vizinhos mais Próximos na série processada, que apontam claramente para um valor de $m=7$.

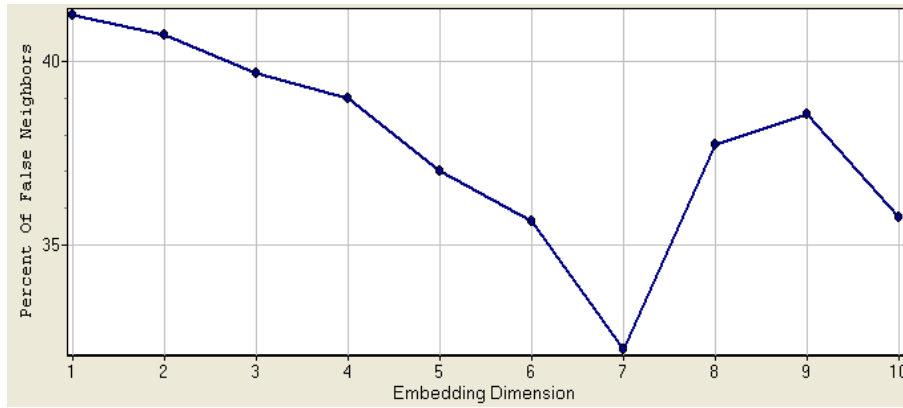


Figura 6.13. Cálculo dos Falsos Vizinhos com a série processada.

O mesmo tratamento é dado para a série temporal sem processamento de tendência e sazonalidade, ou seja, são aplicados os três métodos de análise não linear de séries temporais. Na Figura 6.14 é possível observar o cálculo do Passo de Reconstrução τ , que apresenta um incremento considerável quando comparado com a série processada. Já no caso das Figuras 6.15 e 6.16 mostra-se o cálculo do percentual de falsos vizinhos e da dimensão da correlação, que tendem a sinalizar uma dimensão da imersão $m = 7$.

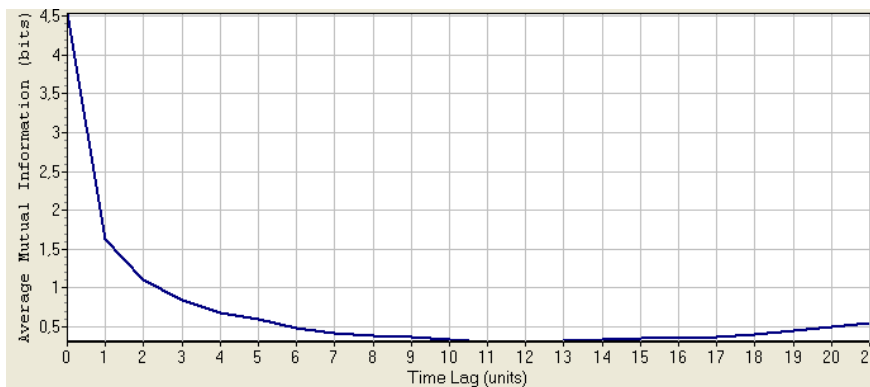


Figura 6.14. Cálculo da Informação Mútua com a série não processada ($\tau = 10$).

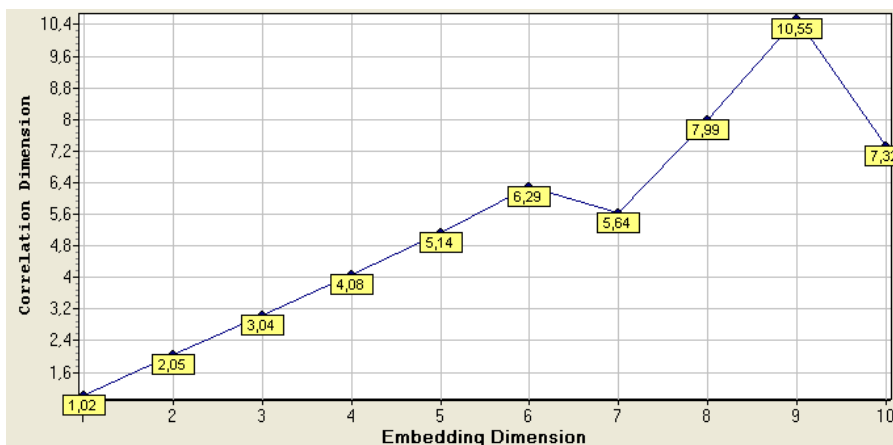


Figura 6.15. Cálculo da Dimensão da Correlação com a série não processada.

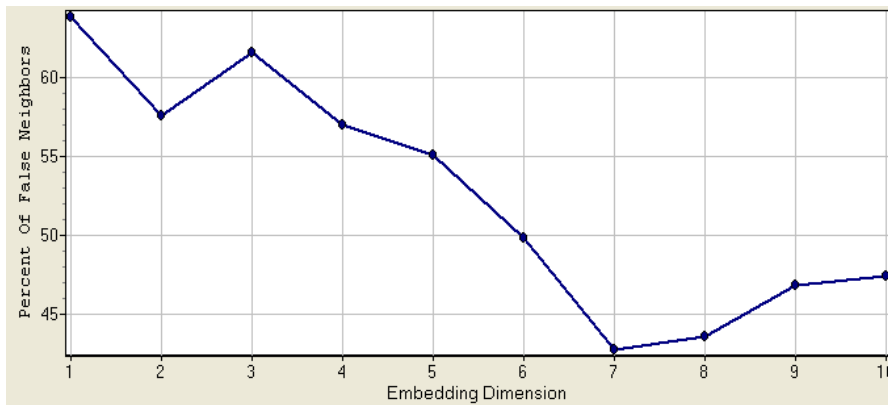


Figura 6.16. Cálculo dos Falsos Vizinhos com a série não processada.

6.7.4 Previsão de Carga usando RNA

Com o objetivo de avaliar a qualidade do processo de seleção de variáveis efetuam-se testes de previsão utilizando redes neurais artificiais. Como definido em 6.2, em todos os casos as redes neurais possuem as mesmas características de arquitetura, número de neurônios, número de camadas, critérios de parada, algoritmo de treinamento, etc.

6.7.5 Reconstrução Espaço de Imersão Série EUNITE 1-2952

Os casos rodados com esta amostra de 2.952 observações foram o seguintes:

A. Rodando com os dados EUNITE usando a variável de Carga: amostra de dados utilizada foi de 182 (01/10/97-31/03/98) dias, sendo 123 para treinamento (01/10/97-31/01/98) com *delays* fornecidos pelo Método dos Falsos Vizinhos mais Próximos ($m=7$) e pelo Método de Informação Mútua ($\tau=10$), reconstruindo-se o espaço fase com os *delays* (10, 20, 30, 40, 50, 60), 500 iterações, RN=[3 1]. Os dados para os testes foram os de 01/02/98 a 31/03/98 (N=51). Não houve nenhum tipo de pré-processamento nos dados antes da aplicação das técnicas de reconstrução (Procedimento 1).

B. Rodando com os dados EUNITE usando a variável de Carga: a amostra de dados utilizada foi de 182 (01/10/97-31/03/98) dias, sendo 123 para treinamento (01/10/97-31/01/98) com *delays* fornecidos pelo método dos Falsos Vizinhos mais Próximos ($m=7$) e pelo Método de Informação Mútua ($\tau=3$), reconstruindo-se o espaço fase usando os *delays* (3, 6, 9, 12, 15, 18), 500 iterações, RN=[3 1]. Os dados para testes utilizados foram (01/02/98-31/03/98) (N=51). A diferença com o Caso A reside no pré-

processamento da série antes da aplicação dos critérios de seleção de variáveis relevantes. É claro que no treinamento da rede os dados também foram processados (Procedimento 2).

C. Rodando com os dados EUNITE usando a variável de Carga e temperatura com os mesmos *lags* definidos para a carga (3, 6, 9, 12, 15, 18): este caso é idêntico ao caso B com a única diferença causada pela incorporação da variável temperatura.

Tabela 6.3. Resultados casos A, B e C

Caso	MAPE (%)	RMSE	ERRO MAX. (%)	ERRO RELAT. (%)	MAD	Nome do Arquivo
A	4,2047	71,316	22,786	0,72786	55,642	Rede_carga_ext_N_process_1_2952.m
B	2,4161	41,071	16,579	-0,20201	31,207	Rede_carga_ext_process_1_2952.m
C	2,2124	37,256	11,122	0,48331	29,213	Rede_carga_Temp_1_2952.m

A seguir alguns comentários sobre os resultados exibidos na Tabela 6.3:

- as redes treinadas com a série temporal de carga “estacionada” apresentam um melhor desempenho do ponto de vista das estatísticas de previsão. Isto pode ser observado quando se compara o Caso A com o Caso B, ambos treinados com os mesmos critérios (arquitetura, algoritmo de treinamento, numero de iterações, critérios de parada, etc) ;
- a inclusão da variável temperatura causa uma melhoria na performance de previsão das redes, como exemplo pode-se confrontar os resultados do Caso B com o Caso C, ambos treinados com os mesmos critérios (arquitetura, algoritmo de treinamento, número de iterações, critérios de parada, etc) e as mesmas entradas de carga, sendo que no caso C também se utilizou dos mesmos *lags* para a variável temperatura. Deve ser ressaltado que neste caso não é necessário possuir previsões de temperatura, porque são utilizados dados históricos desta variável.

Nas simulações a seguir o objetivo é mostrar a sensibilidade dos métodos de extração de variáveis em função da amostra escolhida. No exemplo realizam-se diversas simulações usando a série de carga EUNITE, que foi dividida em 08 intervalos diferentes. Primeiro será realizada a reconstrução de diversos segmentos da série EUNITE sem processamento. Posteriormente efetua-se a reconstrução para os mesmos segmentos anteriores, desta vez aplicando-se os critérios descritos na seção 6.4. Finalmente são treinados 16 redes neurais, 8 para os segmentos da série processada e 8 para a série original. Os resultados são comparados por segmento (processado com não processado),

o que permite novamente a possibilidade de se auferir a sensibilidade do métodos de reconstrução ao processamento de tendência e sazonalidade.

6.7.6 Reconstrução por Segmentos da Série EUNITE sem Processamento

Os resultados obtidos com os métodos de análise não linear da série EUNITE sem processamento podem ser utilizados para reconstruir o Espaço de Imersão. As diversas possibilidades de reconstrução dos sistemas geradores da série são exibidos na Tabela 6.4. Vale ressaltar que foram aplicados dois métodos para calcular a dimensão da imersão (m), sendo necessário, no entanto, escolher o resultado de um deles para efeito de reconstrução. O critério utilizado neste trabalho é o de parcimônia, isto é, sempre é escolhido o método que fornece o menor valor de m .

Tabela 6.4. Espaços de fases reconstruído para série EUNITE sem processamento.

CASO	Dados da série	τ	m	Espaço de Imersão reconstruído
1	1-500	6	6	$Y(t)=\{x(t), x(t-6), x(t-12), x(t-18),x(t-24), x(t-30)\}$
2	1-1000	7	6	$Y(t)=\{x(t), x(t-7), x(t-14), x(t-21),x(t-28), x(t-35)\}$
3	1-1500	8	6	$Y(t)=\{x(t), x(t-8), x(t-16), x(t-24),x(t-32), x(t-40)\}$
4	1-2000	9	6	$Y(t)=\{x(t), x(t-9), x(t-18), x(t-27),x(t-36), x(t-45)\}$
5	1-2500	7	9	$Y(t)=\{x(t), x(t-7), x(t-14), x(t-21),x(t-28), x(t-35), x(t-42), x(t-49), x(t-56)\}$
6	1-3000	11	10	$Y(t)=\{x(t), x(t-11), x(t-22), x(t-33),x(t-44), x(t-55), x(t-66), x(t-77), x(t-88), x(t-99)\}$
7	2000-2500	11	8	$Y(t)=\{x(t), x(t-11), x(t-22), x(t-33),x(t-44), x(t-55), x(t-66), x(t-77)\}$
8	2000-3000	8	10	$Y(t)=\{x(t), x(t-8), x(t-16), x(t-24),x(t-32), x(t-40), x(t-48), x(t-56), x(t-64), x(t-72)\}$

Em geral observa-se na Tabela 6.4 que quanto maior o intervalo de dados, maior o valor de τ e m .

6.7.7 Reconstrução de Segmentos da Série EUNITE com Processamento.

Na Tabela 14 são mostrados os resultados de reconstrução para os intervalos definidos com a série EUNITE, mas desta vez as séries foram processadas visando diminuir as componentes de tendência e sazonalidade. À diferença do exemplo anterior, os valores da defasagem τ variam menos (entre 1 e 2), sendo que 7 entre 8 casos apresentam como resultado $\tau=2$.

Seguindo a mesma lógica do exemplo anterior, os resultados obtidos com os métodos de análise não linear de séries temporais podem ser utilizados para reconstruir o Espaço de Imersão. As diversas possibilidades de reconstrução dos sistemas geradores

da série são exibidos na Tabela 6.5, onde foi aplicado um critério de parcimônia para a definição do valor de m utilizado.

Tabela 6.5. Espaços de fases reconstruído para série EUNITE processada.

CASO	Dados da série	τ	m	Espaço de Imersão reconstruído
1	1-500	2	6	$Y(t)=\{x(t), x(t-2), x(t-4), x(t-6),x(t-8), x(t-10)\}$
2	1-1000	2	8	$Y(t)=\{x(t), x(t-2), x(t-4), x(t-6),x(t-8), x(t-10),x(t-12), x(t-14)\}$
3	1-1500	1	9	$Y(t)=\{x(t), x(t-1), x(t-2), x(t-3),x(t-4), x(t-5), x(t-6), x(t-7) ,x(t-8)\}$
4	1-2000	2	7	$Y(t)=\{x(t), x(t-2), x(t-4), x(t-6),x(t-8), x(t-10) , x(t-12)\}$
5	1-2500	2	2	$Y(t)=\{x(t), x(t-2)\}$
6	1-3000	2	7	$Y(t)=\{x(t), x(t-2), x(t-4), x(t-6),x(t-8), x(t-10) , x(t-12)\}$
7	2000-2500	2	7	$Y(t)=\{x(t), x(t-2), x(t-4), x(t-6),x(t-8), x(t-10) , x(t-12)\}$
8	2000-3000	2	6	$Y(t)=\{x(t), x(t-2), x(t-4), x(t-6),x(t-8), x(t-10)\}$

6.7.8 Qualidade da Reconstrução para Série EUNITE

Na Tabela 6.4 e 6.5 foram apresentadas as diversas possibilidades de reconstrução do Espaço de Imersão e, em consonância com o objetivo deste trabalho que é previsão, faz-se necessário testar quais alternativas revelam-se mais adequadas. A metodologia utilizada para testar as alternativas são:

- usar redes neurais como metodologia de previsão. Desta forma para cada caso das tabelas 6.4 e 6.5, existirá uma rede neural;
- todas as redes usadas devem possuir a mesma arquitetura, usar os mesmos algoritmos de treinamento, os mesmos critérios de parada, a mesma quantidade de épocas para o treinamento, etc;
- as entradas utilizadas são as fornecidas pelos métodos de reconstrução do espaço fase das tabelas 6.4 e 6.5;
- as observações utilizadas como dados de treinamento correspondem aos mesmos dados do processamento utilizados pelos métodos de extração de variáveis mais relevantes;
- como em geral cada intervalo de dados tem quantidades diferentes de observações e como o objetivo é comparar capacidade de previsão em condições semelhantes,

procurou se fixar a relação de dados de treinamento e de testes em aproximadamente 75% e 25% respectivamente;

As estatísticas de previsão são apresentadas nas tabelas 6.6 e 6.7. Pretende-se comparar os resultados da qualidade da reconstrução em função da qualidade da previsão, particularmente os resultados para cada segmento da série não processada com o mesmo segmento desta vez processado, assim por exemplo deve-se comparar o caso 1 da Tabela 6.6 com o caso 1 da Tabela 6.7.

Os resultados mostram uma melhoria em praticamente todas as estatísticas quando existe processamento dos dados utilizados.

Tabela 6.6. Tabela com as estatísticas *out sample* série EUNITE sem processamento.

CASOS	Dados da série	MAPE (%)	RMSE	ERRO MAX. (%)	ERRO RELAT. (%)	MAD
1	1-500	3,7680	64,158	15,004	-0,81044	51,354
2	1-1000	3,5534	62,162	12,502	-0,63962	49,845
3	1-1500	3,5544	62,724	16,021	-0,12449	48,861
4	1-2000	3,8844	66,028	17,936	-0,40116	51,047
5	1-2500	3,5737	64,054	17,784	-0,40207	50,249
6	1-3000	3,9859	66,583	18,378	0,58571	52,900
7	2000-2500	3,8176	63,124	11,001	-0,89985	53,318
8	2000-3000	2,8947	49,274	14,719	1,4993	38,592

Tabela 6.7. Tabela com as estatísticas *out sample* série EUNITE processada.

CASOS	Dados da série	MAPE (%)	RMSE	ERRO MAX. (%)	ERRO RELAT. (%)	MAD
1	1-500	3,0653	50,637	11,383	-0,061511	41,025
2	1-1000	2,3737	41,862	12,306	0,222690	32,597
3	1-1500	2,3024	40,080	8,9444	0,115550	31,822
4	1-2000	2,1012	36,015	13,971	-0,045519	27,397
5	1-2500	2,0744	37,723	11,527	0,029062	28,945

6	1-3000	2,4020	41,405	13,178	0,056006	31,69 6
7	2000-2500	2,4385	44,808	17,828	0,121590	33,29 7
8	2000-3000	2,4855	44,405	12,831	0,094521	33,51 4

As figuras 6.17 e 6.18 exibem duas das estatísticas (MAPE e Erro Máximo) comparando cada caso com e sem processamento, onde se evidencia a importância do processamento das séries antes da aplicação dos métodos de reconstrução de Espaço de Imersão, como forma de melhorar a capacidade de previsão dos modelos neurais. Nos resultados apresentados na Figura 6.18 deve-se excepcionar o Caso 7, onde existe um Erro Máximo atípico completamente fora do padrão.

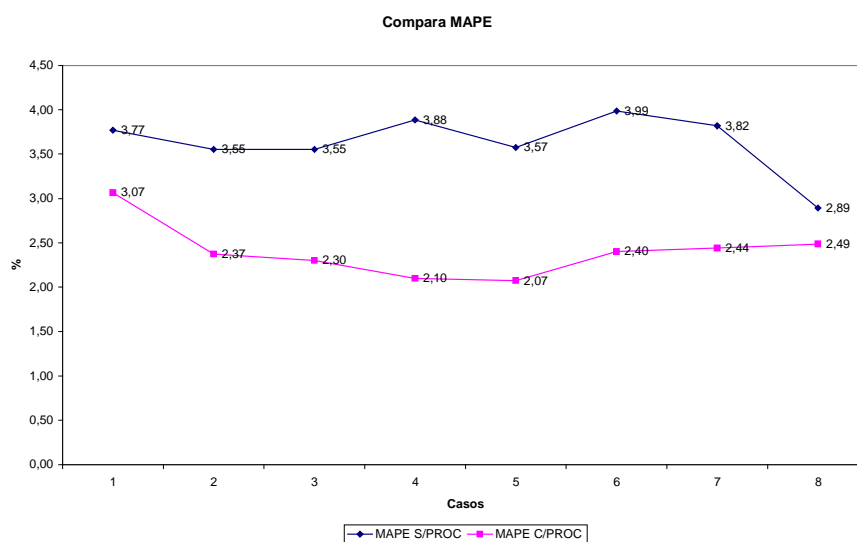


Figura 6.17. Comparação da estatística MAPE .

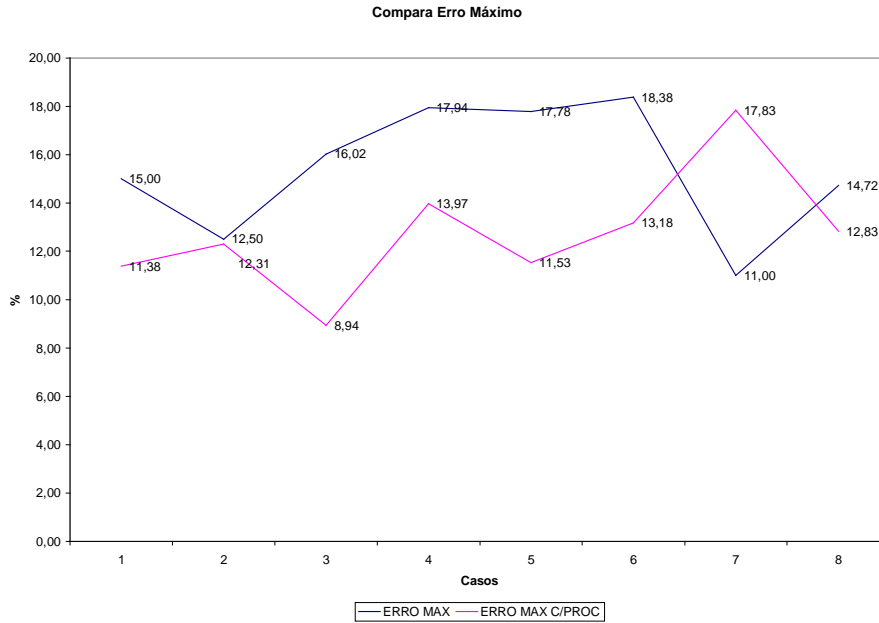


Figura 6.18. Compara a estatística de Erro Máximo.

6.8 Exemplo Usando Série Teórica com Resolução de 1 minuto

Neste exemplo é utilizada a série de carga com integralização de 1 minuto já introduzida no capítulo 5, lembrando que a mesma possui 4320 dados equivalentes a 72 horas ou 3 dias. Aqui o objetivo deste exemplo é avaliar a eventual melhoria, consequência da aplicação de *wavelets* no processo de reconstrução do Espaço Fase.

Na Figura 6.19 é tem-se a série de carga já mencionada. Cabe observar que a mesma tem um intervalo de dados com forte presença de ruído, o que a torna muito interessante para o teste proposto.

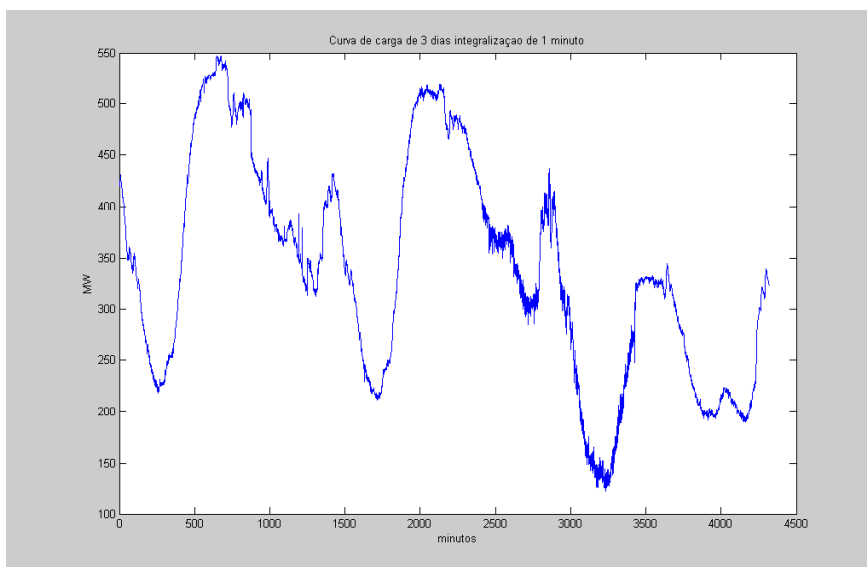


Figura 6.19. Série de carga com dados de 1 em 1 minuto

Na Figura 6.20 é possível observar a aplicação de decomposição para o primeiro nível, o que significa que a série original S (segundo terminologia de *wavelets*) é decomposta em duas componentes, uma conhecida como aproximação $A1$ e uma segunda conhecida como detalhe $D1$. Nota-se que a presença de ruído evidencia-se quando analisada a componente $D1$.

6.8.1 Reconstrução do Espaço de Imersão

Neste exemplo é realizada a reconstrução das três séries: a principal S e suas componentes $A1$ e $D1$. Os resultados são apresentados na Tabela 6.8.

A Tabela 6.8 revela a semelhança entre as variáveis de reconstrução para as séries S e $A1$, o que era já esperado, uma vez que as mesmas são muito parecidas. Observe também que a Dimensão da Imersão m é consideravelmente superior no caso da reconstrução de $D1$.

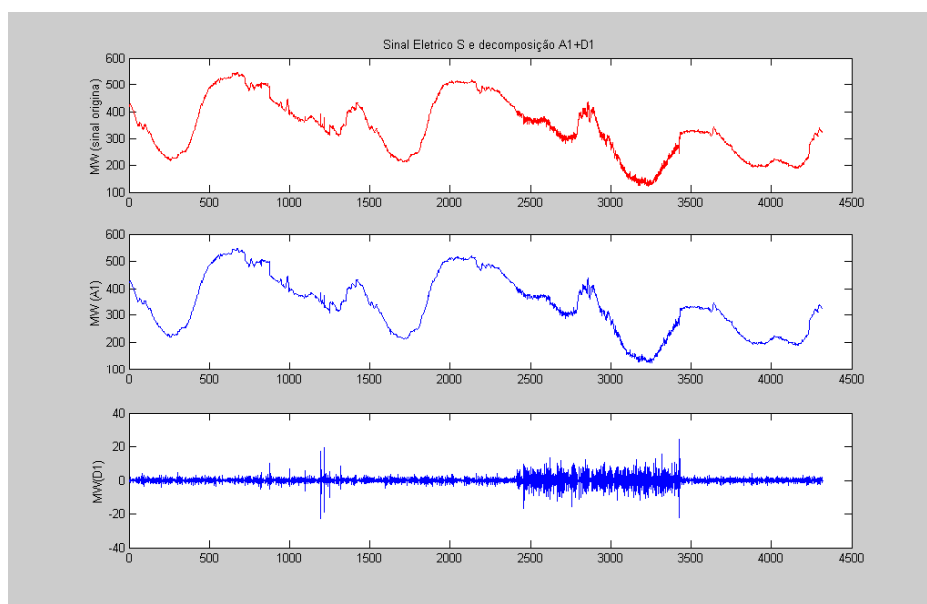


Figura 6.20. Série de carga. Decomposição de primeiro nível.

Tabela 6.8. Reconstrução do Espaço de Imersão

Série	τ	m	Espaço de Imersão reconstruído
S	48	5	$Y(t)=\{x(t), x(t-48), x(t-96), x(t-144),x(t-192)\}$
A1	47	4	$Y(t)=\{x(t), x(t-47), x(t-94), x(t-141)\}$
D1	5	9	$Y(t)=\{x(t), x(t-5), x(t-10), x(t-15),x(t-20), x(t-25), x(t-30), x(t-35) ,x(t-40)\}$

6.8.2 Avaliação da Reconstrução via Previsão

No treinamento da rede foram usados 75% da amostra de dados para treinamento (54 horas) e 16 horas para testes. Observe que na Figura 6.21 a amostra para teste contém um alto nível de ruído nos primeiros 130 minutos de previsão (duas primeiras horas), fato que sem dúvida aumenta o grau de dificuldade na previsão.

Abaixo descreve-se a metodologia adotada para este exemplo:

- A. treinamento de uma rede neural utilizando a reconstrução sugerida para a série original S na Tabela 6.8;
- B. treinamento de uma rede neural utilizando a reconstrução sugerida para A1 na Tabela 6.8;
- C. treinamento de uma rede neural utilizando a reconstrução sugerida para D1 na Tabela 6.8, o que implica numa realização de uma tentativa de modelar o ruído;
- D. comparar os valores previstos pelo modelo obtido em A com os valores realizados. Calcular as estatísticas de previsão;
- E. somar os valores previstos em B e C e confrontar com os valores realizados. Calcular as estatísticas de previsão.

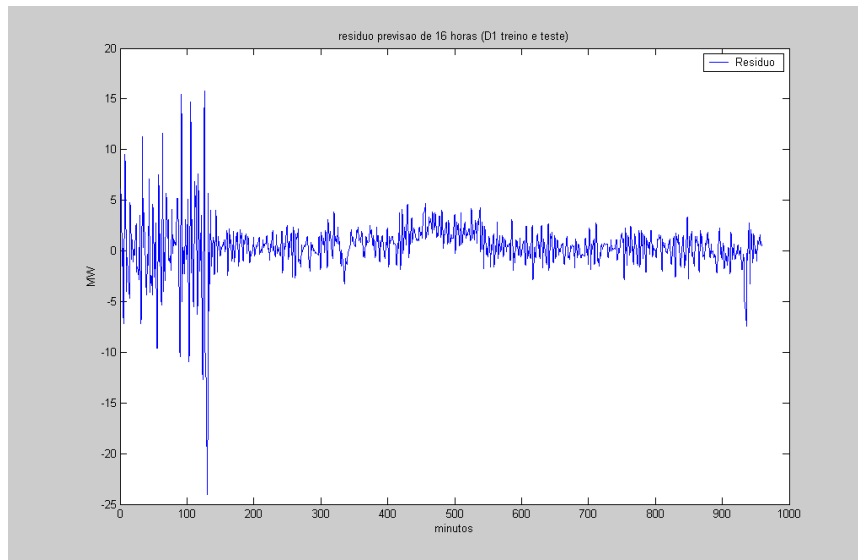


Figura 6.21. Últimas 16 horas da série. Componente D1.

Na tabela 6.9 são apresentados os resultados das estatísticas de previsão dos itens D e E acima listados. Observa-se que utilizar a reconstrução de A1 e D1 de maneira independente, para posteriormente somar os resultados da previsão, aparentemente produz melhores resultados. Na Figura 6.22 tem-se observam as três curvas: as duas previsões junto com os valores realizados para as últimas 16 horas.

Não obstante o anterior, foi realizado um último teste onde foram comparados os valores previstos pelo modelo obtido em A com os valores realizados. Este procedimento é o mesmo utilizado em E, porém sem adicionar os resultados previstos para D1. Ao calcular as estatísticas de previsão da Tabela 6.10 restou constatado que as mesmas são superiores em termos de qualidade de previsão às da Tabela 6.9 (caso A1+D1), o que leva à conclusão de que adicionar os valores previstos para D1 de forma independente não contribui para um incremento no desempenho de previsão, podendo ao contrário prejudica-lo. Em outras palavras, aparentemente não faz sentido tentar “modelar” a componente de alta frequência D1.

Tabela 6.9. Tabela com as estatísticas *out sample* série de 1 minuto.

Caso	MAPE (%)	RMSE	ERRO MAX.	ERRO RELAT. (%)	MAD	Nome do Arquivo
			(%)			
A1+D1	3,4411	13,322	20,027	0,74365	9,0801	Estatística_S_D1_A1.m
Série S	12,443	39,223	30,622	7,4932	32,468	Série_1_M_reconst_S.m

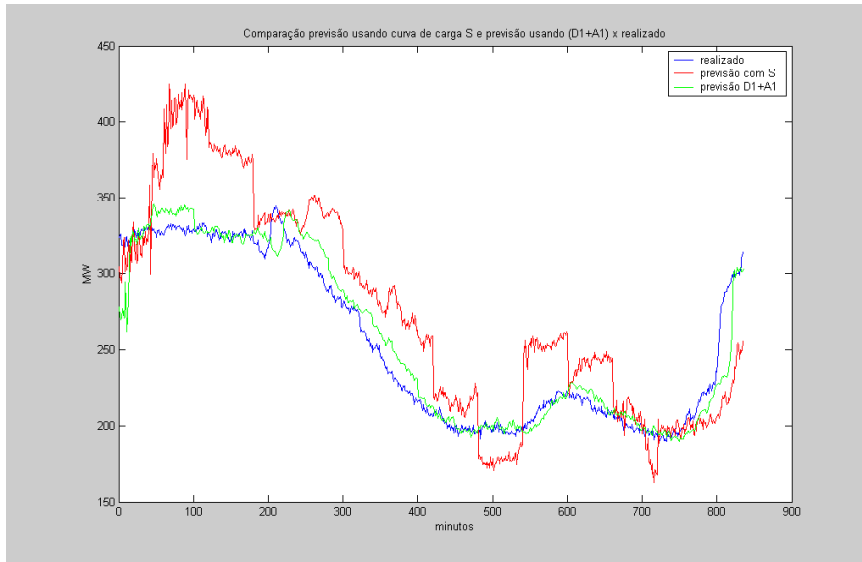


Figura 6.15. Comparação de resultados de previsão

Figura 6.22. Comparação de previsões.

Tabela 6.10. Tabela com as estatísticas *out sample* série A1

Caso	MAPE (%)	RMSE	ERRO MAX. (%)	ERRO RELAT. (%)	MAD	Nome do Arquivo
Série A1	2,0997	7,497	20,625	0,78225	5,3476	Série_1_M_reconst_A_1.m

6.9 Exemplo Usando Série PUGET

Os dados aqui utilizados correspondem a observações de carga e temperatura da *Puget Sound Power and Light Company*, distribuidora de energia elétrica da cidade de Seattle, EUA que neste trabalho é denominada simplesmente PUGET (<http://www.ee.washington.edu/class/559/2002spu>). Nas tabelas abaixo é apresentado o volume de dados disponíveis de carga e temperatura.

Tabela 6.11. Total de dados disponíveis.

Variável	Período	Intervalo apuração	Quantidade de dados
Carga	01/01/85-10/12/92	horário	68.204
Temperatura	01/01/85-10/12/92	horário	68.204

Neste capítulo, nos exemplos relacionados com a série PUGET serão utilizados dados históricos dos janeiros entre os anos 1989 e 1992, os quais são apresentados na Tabela 6.12.

Tabela 6.12. Dados mês de Janeiro entre 1989-1992

Variável	Período	Intervalo apuração	Quantidade de dados
Carga	Janeiros 89-92	horário	2.976
Temperatura	Janeiros 89-92	horário	2.976

6.9.1 PUGET: Exemplo 1

Neste primeiro Exemplo é realizada uma comparação da qualidade da reconstrução utilizando os procedimentos 1 e 2 descritos na seção 6.6. Lembre-se que a única diferença entre estes procedimentos é o processamento da série para tendência e sazonalidade utilizado no procedimento 2. As tabelas 6.13 e 6.14 exibem os resultados para as variáveis de reconstrução para duas situações: série sem processamento e com processamento. Observe que os valores de τ e m escolhidos para reconstruir o Espaço de Imersão estão destacados em azul.

TABELA 6.13.Reconstrução espaço estado usando série PUGET

Dados da série	Informação mutua τ	FNN (precisão Max) m	FNN (precisão Med) m	FNN (precisão Min) m	D ₂ m
PUGET	(τ) min= 5	4	5	9	9
	(τ) med= 5	4	5	9	9
	(τ) Max= 5	4	5	9	9

TABELA 6.14.Reconstrução espaço estado usando serie PUGET pré-processada

Dados da série	Informação mutua τ	FNN (precisão Max) m	FNN (precisão Med) m	FNN (precisão Min) m	D ₂ m
PUGET_PROC	(τ) min= 2	9	9	9	13
	(τ) med= 6	10	10	3	12
	(τ) Max= 6	10	10	3	12

Os resultados de previsão obtidos pelas RNA com as entradas sugeridas pelas tabelas 6.13 e 6.14 estão disponíveis na Tabela 6.15. Novamente indicam que é recomendável utilizar séries de carga com pré-processamento de tendência e sazonalidade. Os critérios utilizados para treinar e testar as redes são os descritos na seção 6.2.

TABELA 6.15. Resultados de previsão Exemplo 1.

h	Série	MAPE (%)	ERRO MAX. (%)	ERRO RELAT. (%)	MAD	Nome do Arquivo
	S	4,758	29,33	0,25	123,468	COMPARA2.XLS
	S estac	1,674	19,05	0,033	44,697	COMPARA2.XLS

(*) série S=PUGET original

6.9.2 PUGET: Exemplo 2

Neste segundo Exemplo compara-se a qualidade da reconstrução utilizando os procedimentos 1 e 3 descritos na seção 6.6. Neste caso, a única diferença entre estes procedimentos é a aplicação da decomposição de *wavelets* antes da reconstrução utilizada no procedimento 3. A tabelas 6.16 mostra os resultados da previsão para ambas situações. Cabe ressaltar que segundo a terminologia de *wavelets*, S é a série original, neste caso PUGET, e A1 é a componente de aproximação para o primeiro nível de decomposição da série. Neste caso é reconstruído, o Espaço Fase apenas para A1 e desconsiderado o “ruído” original (ou componente de alta frequência) D1. As series não foram pré-processadas antes da aplicação dos métodos de reconstrução. Os resultado indicam que ao se eliminar o ruído D1 e realizar a reconstrução somente com A1, a qualidade da reconstrução melhora refletindo nos resultados da previsão.

TABELA 6.16. Resultados de previsão Exemplo 2.

Série	MAPE (%)	ERRO MAX. (%)	ERRO RELAT. (%)	MAD	Nome do Arquivo
S	4,660	29,33	0,154	120,27	COMPARA1.XLS
A1	4,614	20,99	0,555	116,91	COMPARA1.XLS

6.9.3 PUGET: Exemplo 3

Neste terceiro Exemplo é realizada uma comparação da qualidade da reconstrução utilizando todos os procedimentos descritos na seção 6.6. Sobre os resultados apresentados na Tabela 6.17 podem ser tecidos os seguintes comentários:

- a utilização do procedimento 1 apresenta os piores resultados;
- a aplicação do procedimento 3 (*wavelets*) na série provoca uma leve melhoria nos resultados;
- a aplicação do procedimento 2 (processamento) melhora consideravelmente os resultados;
- a aplicação do procedimento 4 (*wavelets*+processamento), neste caso, não supera os resultados do procedimento 2.

TABELA 6.17. Resultados de previsão Exemplo 2.

Procedime	MAP E (%)	ERRO MAX. (%)	ERRO RELAT. (%)	MAD	Nome do Arquivo
PROC_1	4,75	29,33	0,248	123,49	COMPARA4.XLS
PROC_2	1,68	19,05	0,036	44,775	COMPARA4.XLS
PROC_3	4,74	20,99	0,735	120,464	COMPARA4.XLS
PROC_4	1,70	14,37	-0,038	44,797	COMPARA4.XLS

6.9.4 PUGET: Exemplo 4 - Influência da Temperatura

Neste quarto e último Exemplo usando a série PUGET é avaliada a influência da variável temperatura no desempenho da previsão. Para isto são testadas as seguintes alternativas ou casos:

Caso A. Corresponde ao caso base, ou seja, à aplicação do procedimento 1, cujos resultados já foram expostos na primeira linha da Tabela 6.15, ou seja, neste caso não há presença da variável temperatura.

Caso B. Corresponde ao caso base, incluindo-se aqui a variável temperatura com os mesmos *lags* definidos para a carga após a reconstrução do Espaço de Imersão. Em outras palavras, são incluídas tantas variáveis de temperaturas quanto variáveis de carga. Ressalte-se que as variáveis de temperatura fazem parte do histórico, isto é, não é necessário ter a disposição uma previsão de temperatura.

Caso C. Semelhante ao Caso A, diferindo apenas pela inclusão de uma variável de temperatura no *target* da RNA. Na prática isto significa a necessidade de dispor de uma previsão de temperatura horária confiável.

Caso D. Semelhante ao Caso B, agora com a utilização de variáveis de temperatura defasadas em função dos *lags* definidos para a carga e inclusão de uma variável de

temperatura no *target* da RNA. Igualmente ao Caso C, há necessidade de dispor de uma previsão de temperatura confiável.

Os resultados apresentados na Tabela 6.18 mostram que há piora no resultado de Caso A para o Caso B, o que evidencia que nem sempre a inclusão da variável temperatura pode ser um fator benéfico ou talvez que a inclusão, da variável temperatura defasada em função dos mesmos *lags* definidos para a carga, não seja a melhor opção.

A melhoria exposta no Caso C é “tímida” e tem como premissa uma previsão de temperatura horária perfeita, o que a converte numa alternativa pouco plausível. A mesma análise realizado para o Caso C serve para o Caso D.

TABELA 6.18. Resultados de previsão Exemplo 4.

Caso	MAPE (%)	ERRO MAX. (%)	ERRO RELAT. (%)	MAD	Nome do Arquivo
A	4,758	29,33	0,25	123,468	COMPARA2.XLS
B	4,7948	22,961	-0,52398	125,79	Rede_Puget_serie_S_temp.m
C	3,9932	22,077	-1,4304	104,79	Rede_Puget_serie_S_temp_2.m
D	3,9786	22,936	-0,80101	104,87	Rede_Puget_serie_S_temp_3.m

6.10 Exemplo Usando Série de Carga Nacional com Resolução Horária

A seguir são realizados diversos Exemplos de reconstrução e previsão utilizando uma série de carga de uma distribuidora de energia nacional. A série em questão corresponde a um grupo de alimentadores da região metropolitana de Recife, que corresponde a aproximadamente 21% do mercado total atendido pela CELPE- Companhia Energética de Pernambuco. Em termos de demanda, significa aproximadamente 200 MW médios. Os dados originais foram integralizados de 5 em 5 minutos, desta forma existe a possibilidade de analisar as séries com diversos níveis de resolução: 5 minutos, 15 minutos, 30 minutos e 1 hora. Cabe ressaltar que independente da resolução dos dados da série, o objetivo é realizar previsão de demanda horária. Isto significa que, por exemplo, quando utilizados os dados de 5 minutos são realizadas 12 previsões para cada intervalo de 1 hora, no entanto para avaliar a qualidade da previsão,

é utilizado somente o último valor entre os 12 previstos. Analogamente, no caso de dados de 15 minutos, são realizadas 4 previsões no intervalo de 1 hora, no entanto para avaliar a qualidade da previsão, é utilizado somente o último valor entre os 4 previstos. O mesmo critério se aplica aos dados de 30 minutos, onde são previstos dois valores para cada hora.

6.10.1 Influência do Processamento da Série na Reconstrução

Nesta seção é avaliada a influência isolada que o processamento da série CELPE (para tendência e sazonalidade) causa no processo de reconstrução. A seguir são realizados 04 Exemplos de reconstrução e previsão utilizando uma série de carga CELPE com diferentes níveis de resolução.

1. Caso Série 01 hora:

Neste caso é utilizada a série de carga horária sem processamento prévio, isto é, sem aplicação dos procedimentos para diminuição das componentes de tendência e sazonalidade. Os resultados do processo de reconstrução de Espaço Fase são apresentados na tabela abaixo. Observe que os valores escolhidos estão destacados em azul.

TABELA 6.19. Variáveis para reconstrução espaço de imersão usando série horária

Dados da série	Informação mutua τ	FNN (precisão Max) m	FNN (precisão Med) m	FNN (precisão Min) m	D ₂ m
(*) Série S 1 hora	(τ) min= 8	7	4	7	13
	(τ) med= 7	6	6	6	13
	(τ) Max= 7	6	6	6	13

(*) Série S: neste caso corresponde à série CELPE

Os resultados do processo de reconstrução de Espaço Fase para a série horária com processamento prévio, isto é, aplicando os procedimentos para diminuição das componentes de tendência e sazonalidade, são exibidos na Tabela 6.20.

TABELA 6.20. Variáveis para reconstrução espaço de imersão usando série horária processada.

Dados da série	Informação mutua τ	FNN (precisão Max) m	FNN (precisão Med) m	FNN (precisão Min) M	D ₂ m
Série S 1 hora_Proc	(τ) min= 1	10	4	10	10
	(τ) med= 6	10	10	10	18
	(τ) Max= 6	10	10	10	18

TABELA 6.21. Espaço de reconstrução.

Série usada	Passo de Reconstrução τ	Dimensão da Imersão m	Espaço de Imersão
Serie S_1h	7	6	$Y(t)=\{ x(t-7),x(t-14),\dots,x(t-35) \}$
Serie S_1h_proc	6	10	$Y(t)=\{ x(t-6),x(t-12),\dots,x(t-54) \}$

TABELA 6.22. Resultados de previsão usando dados da Tabela 6.21

Série usada	MAPE (%)	ERRO MAX. (%)	ERRO RELAT. (%)	MAD	Nome do Arquivo
Serie S_1h	8,328	51,241	1,067	131,99	Rede_1h_serie_S_completa.m
Serie S_1h_pr	3,290	30,480	0,107	55,19	Rede_1h_serie_S_proc.m

2. Caso Série 30 minutos:

Neste caso é utilizada a série de carga CELPE com resolução de 30 minutos com e sem aplicação dos procedimentos para diminuição das componentes de tendência e sazonalidade. Os resultados do processo de reconstrução de Espaço Fase são apresentados na Tabela 6.23, já na Tabela 6.24 são apresentados os resultados de previsão para cada um das séries utilizadas.

TABELA 6.23. Espaço de reconstrução.

Série usada	Passo de Reconstrução τ	Dimensão da Imersão m	Espaço de Imersão
Serie S_30m	13	10	$Y(t)=\{ x(t-13),x(t-26),\dots,x(t-117) \}$
Serie S_30m_pr	7	13	$Y(t)=\{ x(t-7),x(t-14),\dots,x(t-84) \}$

TABELA 6.24. Resultados de previsão usando dados da Tabela 6.23

Série usada	MAPE (%)	ERRO MAX.	ERRO RELAT.	MAD	Nome do Arquivo
-------------	----------	-----------	-------------	-----	-----------------

		(%)	(%)		
Serie S_30m	6,351	72,255	1,0721	103,68	Rede_30_serie_S_completa.m
Serie S_30m_pr	2,558	23,73	0,056	44,25	RESUMO_RESULTADOS CASO10.xls

3. Caso Série 15 minutos:

Neste caso é utilizada a série de carga CELPE com resolução de 15 minutos com e sem aplicação dos procedimentos para diminuição das componentes de tendência e sazonalidade. Os resultados do processo de reconstrução de Espaço Fase são apresentados na Tabela 6.25, já na Tabela 6.26 são apresentados os resultados de previsão para cada um das séries utilizadas.

TABELA 6.25. Espaço de reconstrução.

Série usada	Passo de Reconstrução τ	Dimensão da Imersão m	Espaço de Imersão
Serie S_15m	25	10	$Y(t)=\{ x(t-25),x(t-50),\dots,x(t-225) \}$
Serie S_15m_pr	4	17	$Y(t)=\{ x(t-4),x(t-8),\dots,x(t-64) \}$

TABELA 6.26. Resultados de previsão usando dados da Tabela 6.25

Série usada	MAPE (%)	ERRO MAX. (%)	ERRO RELAT. (%)	MAD	Nome do Arquivo
Serie S_15m	6,8775	69,516	1,2135	109,76	Rede_15_serie_S_completa.m
Serie S_15m_pr	2,063	45,58	0,060	35,14	RESULTADOS CASO 8

4. Caso Série 05 minutos:

Neste caso é utilizada a série de carga CELPE com resolução de 05 minutos com e sem aplicação dos procedimentos para diminuição das componentes de tendência e sazonalidade. Os resultados do processo de reconstrução de Espaço Fase são apresentados na Tabela 6.27, já na Tabela 6.28 são apresentados os resultados de previsão para cada um das séries utilizadas.

TABELA 6.27. Espaço de reconstrução.

Série usada	Passo de Reconstrução τ	Dimensão da Imersão m	Espaço de Imersão
Serie S_05m	67	10	$Y(t)=\{ x(t-67),x(t-134),\dots,x(t-603) \}$
Serie S_05m_pr	2	9	$Y(t)=\{ x(t-2),x(t-4),\dots,x(t-16) \}$

TABELA 6.28. Resultados de previsão usando dados da Tabela 6.27.

Série usada	MAPE (%)	ERRO MAX. (%)	ERRO RELAT. (%)	MAD	Nome do Arquivo
Serie S_05m	6,65	149,81	1,2757	107,3	Rede_5_serie_S_completa.m
Serie S_05m_pr	1,50	7,54	-0,005	25,64	RESUMO_RESULTADO S CASO10.xls

6.10.2 Influência da Resolução da Série na Reconstrução do Espaço de Imersão

No exemplo apresentado nesta seção, é testado o impacto isolado que ocasiona reconstruir o Espaço de Imersão usando diferentes níveis de resolução da série de carga CELPE. Cabe ressaltar que independente da resolução dos dados da série, o objetivo é realizar previsão de demanda horária para um mesmo intervalo de tal forma a viabilizar a comparação.

Na Tabela 6.29 são apresentados os resultados de previsão onde em geral é possível observar que quanto maior o grau de resolução utilizado melhores os resultados obtidos. É interessante observar que no caso da série com resolução de 30 minutos os resultados foram melhores que para 15 e 05 minutos, resultado que de alguma forma se configura como uma surpresa.

TABELA 6.29. Comparação de resultados previsão usando diferente nível de resolução.

	MAPE (%)	ERRO MAX. (%)	ERRO RELAT. (%)	MAD	Nome do Arquivo
Serie S_1h	7,888	61	0,897	127,46	RESUMO_RESULTADO S CASO1
Serie S_30n	6,218	72	1,069	101,46	
Serie S_15n	6,851	69	1,217	110,01	
Serie S_5m	6,761	59	1,111	109,26	

6.10.3 Influência da Aplicação de *Wavelets* na Reconstrução do Espaço de Imersão

Nesta seção é avaliada a influência isolada que o processamento via Teoria de *Wavelets* da série CELPE produz no processo de reconstrução. A seguir são realizados 04 Exemplos de reconstrução e previsão utilizando uma série de carga CELPE com diferentes níveis de resolução. Cabe ressaltar que em todos os casos foi utilizado somente o primeiro nível de decomposição, sendo eliminado a componente de alta frequência D1. Desta forma a reconstrução foi realizado utilizando apenas a componente de aproximação A1.

5. Caso Série 01 hora:Aplicação de *wavelets*

Neste caso é utilizada a série de carga CELPE com resolução de 1 hora com e sem aplicação da decomposição via *wavelets*. Os resultados de previsão para cada um das séries utilizadas são apresentados na Tabela 6.30.

TABELA 6.30.Resultados de previsão Caso série 01 hora

Série us	MAPE (%)	ERRO MAX. (%)	ERRO RELAT. (%)	MAD	Nome do Arquivo
Serie S_1h	8,144	61,42	1,481	129,30	RESUMO_RESULTADOS CASO2.XLS
Serie A1_1h	7,861	73,94	1,312	119,60	

6. Caso Série 30 minutos:Aplicação de *wavelets*

Aqui é utilizada a série de carga CELPE com resolução de 30 minutos com e sem aplicação da decomposição via *wavelets*. Os resultados de previsão para cada um das séries utilizadas são apresentados na Tabela 6.31.

TABELA 6.31.Resultados de previsão Caso série 30 minutos.

Série us	MAPE (%)	ERRO MAX. (%)	ERRO RELAT. (%)	MAD	Nome do Arquivo
Serie S_30	6,351	72,25	2,461	103,67	RESULTADOS CASO3.xls
Serie A1_30	6,174	51,07	3,348	101,07	

7. Caso Série 15 minutos:Aplicação de *wavelets*

Neste caso é utilizada a série de carga CELPE com resolução de 15 minutos com e sem aplicação da decomposição via *wavelets*. Os resultados de previsão para cada um das séries utilizadas são apresentados na Tabela 6.32.

TABELA 6.32. Resultados de previsão Caso série 15 minutos.

Série us	MAPE (%)	ERRO MAX. (%)	ERRO RELAT. (%)	MAD	Nome do Arquivo
Serie S_15	6,857	69,48	1,226	110,24	RESULTADOS CASO4
Serie A1_15	6,572	74,08	0,350	107,22	

8. Caso Série 05 minutos: Aplicação de *wavelets*

Finalmente é utilizada a série de carga CELPE com resolução de 05 minutos com e sem aplicação da decomposição via *wavelets*. Os resultados de previsão para cada um das séries utilizadas são apresentados na Tabela 6.33.

TABELA 6.33. Resultados de previsão Caso série 05 minutos.

Série us	MAPE (%)	ERRO MAX. (%)	ERRO RELAT. (%)	MAD	Nome do Arquivo
Serie S_05	6,651	149,79	1,276	107,29	RESULTADOS CASO5.xls
Serie A1_05	5,990	107,18	0,458	97,89	

6.10.4 Influência do Processamento de T_t e S_t na série + *Wavelets* na Reconstrução do Espaço de Imersão

Nesta seção é avaliada a influência que o processamento de tendência (T_t) e sazonalidade (S_t) aliado a decomposição via Teoria de *Wavelets* causam na série CELPE e no seu processo de reconstrução. Em outras palavras esta sendo testado o Procedimento 4 descrito na Figura 6.4. A seguir são realizados 04 Exemplos de reconstrução e previsão utilizando uma série de carga CELPE com diferentes níveis de resolução.

9. Caso Série 01 hora: Aplicação do processamento de T_t e S_t + *wavelets*

É utilizada a série de carga CELPE com resolução de 1 hora. Observe que as duas séries estão processadas em T_t e S_t com a diferença que na segunda linha da Tabela 6.34 foi aplicada a decomposição de *wavelets*.

TABELA 6.34. Resultados de previsão.

Série usada	MAPE (%)	ERRO MAX. (%)	ERRO RELAT. (%)	MAD	Nome do Arquivo
Serie S_Proc_1h	3,290	30,48	0,107	55,19	RESULTADOS CASO6.xls
Serie A1_Proc_1h	2,523	25,60	-0,100	42,42	

10. Caso Série 30 minutos: Aplicação do processamento de T_t e S_t + *wavelets*

É utilizada a série de carga CELPE com resolução de 30 minutos. Observe que as duas séries estão processadas em T_t e S_t com a diferença que na segunda linha da Tabela 6.35 foi aplicada a decomposição de *wavelets*.

TABELA 6.35. Resultados de previsão.

Série usada	MAPE (%)	ERRO MAX. (%)	ERRO RELAT. (%)	MAD	Nome do Arquivo
Serie S_Proc_30m	2,735	31,28	0,047	45,90	RESULTADOS CASO7
Serie A1_Proc_30m	1,914	18,36	0,041	31,73	

11. Caso Série 15 minutos: Aplicação do processamento de T_t e S_t + *wavelets*

É utilizada a série de carga CELPE com resolução de 15 minutos. Observe que as duas séries estão processadas em T_t e S_t com a diferença que na segunda linha da Tabela 6.36 foi aplicada a decomposição de *wavelets*.

TABELA 6.36. Resultados de previsão.

Série usada	MAPE (%)	ERRO MAX. (%)	ERRO RELAT. (%)	MAD	Nome do Arquivo
Serie S_Proc_15m	2,063	45,58	0,060	35,14	RESULTADOS CASO 8
Serie A1_Proc_15m	1,229	17,22	0,045	20,65	

12. Caso Série 05 minutos: Aplicação do processamento de T_t e S_t + *wavelets*

É utilizada a série de carga CELPE com resolução de 05 minutos. Observe que as duas séries estão processadas em T_t e S_t com a diferença que na segunda linha da Tabela 6.37 foi aplicada a decomposição de *wavelets*.

TABELA 6.37. Resultados de previsão.

Série usada	MAPE (%)	ERRO MAX. (%)	ERRO RELAT. (%)	MAD	Nome do Arquivo
Serie S_Proc_05	1,498	139,89	0,023	25,34	RESULTADOS CASO 9.xls
Serie A1_Proc_05	0,847	81,00	0,024	14,29	

6.11 Comparação Resultados Competição EUNITE 2001

Nesta seção é apresentada uma comparação de resultados de previsão entre a técnica apresentada em [64] e a aplicada no desenvolvimento desta tese. Os dados utilizados são os já apresentados na seção 6.7.1 deste capítulo. O objetivo desta competição foi a realização de previsões de demanda máxima diária durante o mês de janeiro de 1999. Para isto foram fornecidos dados históricos de carga do período 01/01/1997 até 31/12/1998 e dados de temperatura média do mesmo período. Cabe ressaltar que os autores de [64] foram os vencedores de dita competição.

Da mesma forma que os autores em [64] nesta tese foram definidas duas amostras de dados diferentes, a saber: Amostra de Inverno (W), que considera os meses de Janeiro-Março além dos dados de Outubro-Dezembro e a Amostra Janeiro-Fevereiro (JF) que considera somente os meses de Janeiro e Fevereiro. Em nenhum dos testes apresentados foram utilizados os dados de temperatura. A seguir serão apresentados os resultados obtidos.

6.11.1 Resultados Utilizando Amostra de Inverno (W)

Como foi mencionado antes a amostra utilizada para este primeiro teste está constituída de dados de carga de 30 minutos dos meses de Janeiro-Março além dos dados de Outubro-Dezembro dos anos 1997 e 1998. Como tem sido rotina nesta tese, as séries são analisadas com ou sem processamento das componentes de tendência (T) e sazonalidade (S).

6.11.1.1 Série Amostra de Inverno (W) sem processamento T e S

A seguir são apresentados os resultados usando a série original, isto é, sem nenhum tipo de processamento de T e S. Na Tabela 6.38 é mostrada a aplicação dos métodos de reconstrução utilizados nesta tese. Observe que como de costume são calculados dois valores diferentes para a variável dimensão da imersão (m), uma usando FNN e uma segunda utilizando D_2 . Na Tabela 6.39 é exibido o Espaço de Imersão reconstruído, isto é, as entradas que serão utilizadas na rede neural visando realizar a previsão de demanda máxima diária.

TABELA 6.38. Resultados das variáveis de reconstrução

Dados da série	Informação mutua τ	FNN (precisão Max) m	D_2 m
Série w_s/proc	(τ) med=28	23	8

TABELA 6.39. Resultados do Espaço de reconstrução.

Série usada	Passo de Reconstrução τ	Dimensão da Imersão m	Espaço de Imersão
Série w_s/proc	28	8	$Y(t) = \{ x(t-28), x(t-56), \dots, x(t-196) \}$
Série w_s/proc	28	23	$Y(t) = \{ x(t-28), x(t-56), \dots, x(t-616) \}$

Os resultados da Tabela 6.40 indicam um melhor desempenho ao utilizar, junto com o Passo de Reconstrução 28, a Dimensão da Imersão 23 fornecida pelo método dos falsos vizinhos FNN. Nas figuras 6.23 e 6.24 é apresentada a variação entre a demanda realizada e a prevista pelas RNA utilizadas.

TABELA 6.40. Resultados das estatísticas de previsão

MAPE	Erro Máximo	Técnica Dimensão da Imersão m
3,715%	13,197%	D_2
2,918%	8,535%	FNN

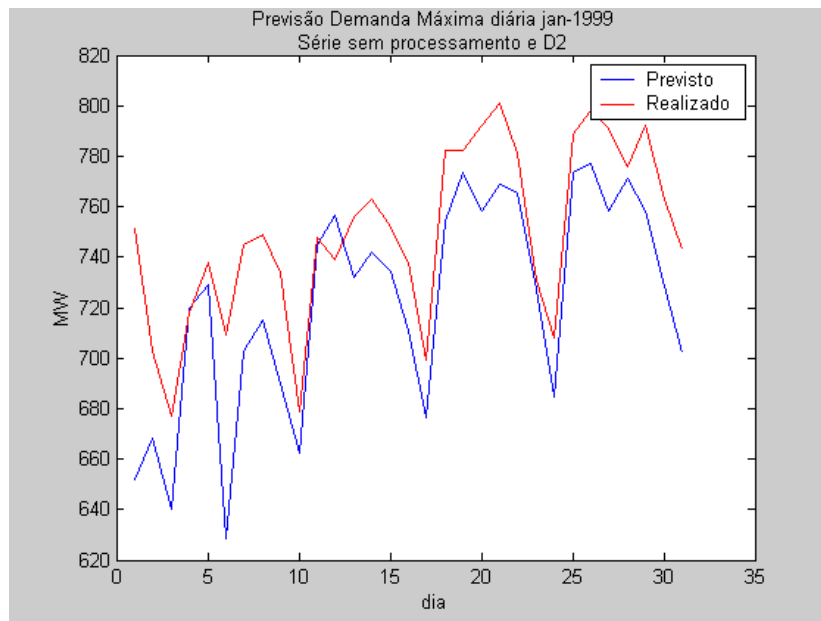


Figura 6.23. Previsto x Realizado usando Série W s/proc (D_2)

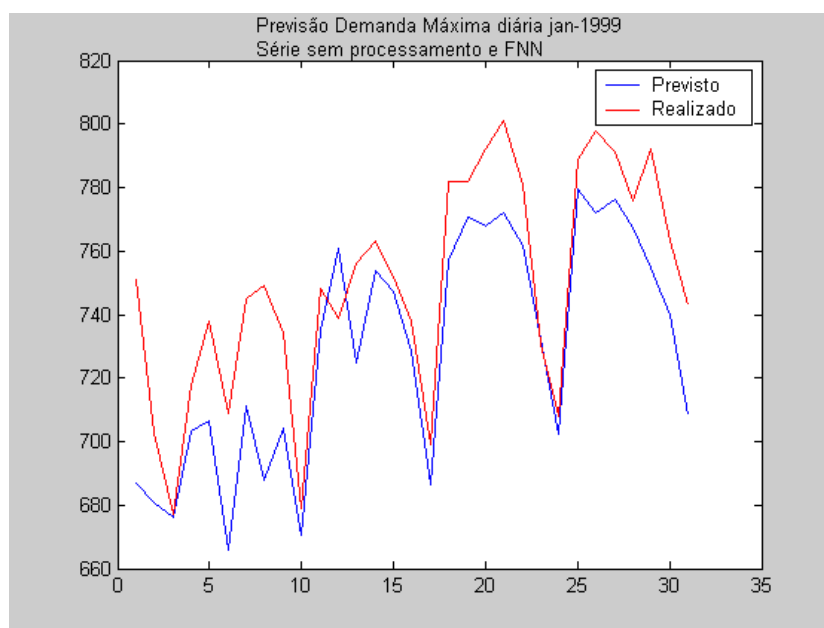


Figura 6.24. Previsto x Realizado usando Série W s/proc (FNN)

6.11.1.2 Série Amostra de Inverno (W) com processamento T e S

Agora são apresentados os resultados utilizando a série original com pré-processamento de T e S. Na Tabela 6.41 é mostrada a aplicação dos métodos de reconstrução utilizados nesta tese. Observe que ao igual que na seção anterior, são calculados dois valores diferentes para a variável dimensão da imersão (m), uma usando FNN e uma segunda utilizando D_2 . Na Tabela 6.42 são exibidos os Espaços de Imersão

reconstruídos, isto é, as entradas que serão utilizadas na rede neural visando realizar a previsão de demanda máxima diária.

TABELA 6.41. Resultados das variáveis de reconstrução

Dados da série	Informação mutua τ	FNN (precisão Max) m	D_2 m
Série w_C/proc	(τ) med=2	36	12

TABELA 6.42. Resultados do Espaço de reconstrução.

Série usada	Passo de Reconstrução τ	Dimensão da Imersão m	Espaço de Imersão
Série w_C/proc	2	12	$Y(t) = \{ x(t-2), x(t-4), \dots, x(t-22) \}$
Série w_C/proc	2	36	$Y(t) = \{ x(t-2), x(t-4), \dots, x(t-70) \}$

Os resultados da Tabela 6.43 indicam um melhor desempenho ao utilizar, junto com o Passo de Reconstrução 2, a Dimensão da Imersão 36 fornecida pelo método dos falsos vizinhos FNN. Nas figuras 6.25 e 6.26 é apresentada a variação entre a demanda realizada e a prevista pelas RNA utilizadas.

TABELA 6.43. Resultados das estatísticas de previsão

MAPE	Erro Máximo	Técnica Dimensão da Imersão m
1,616%	4,523%	D_2
0,835%	3,327%	FNN

Observe que ao comparar as Tabelas 6.40 com a Tabela 6.43 fica evidente o benefício de realizar o pré-processamento das séries de carga antes da aplicação das técnicas de reconstrução do Espaço de Imersão. Outro fato interessante é o melhor desempenho dos valores fornecidos pelo método FNN, que nas duas tabelas mencionadas apresentou valores mais adequados.

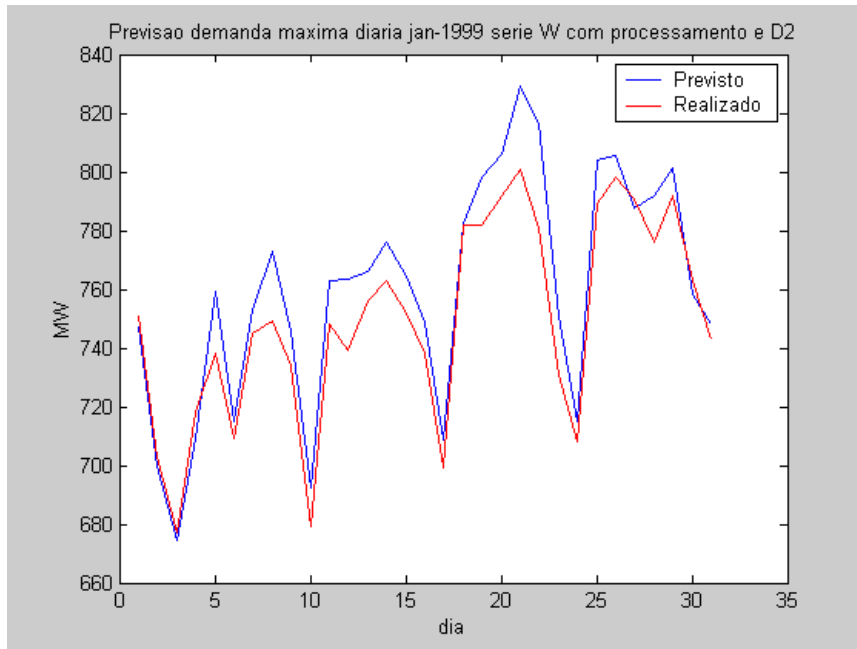


Figura 6.25. Previsto x Realizado usando Série W com/ proc (D_2)

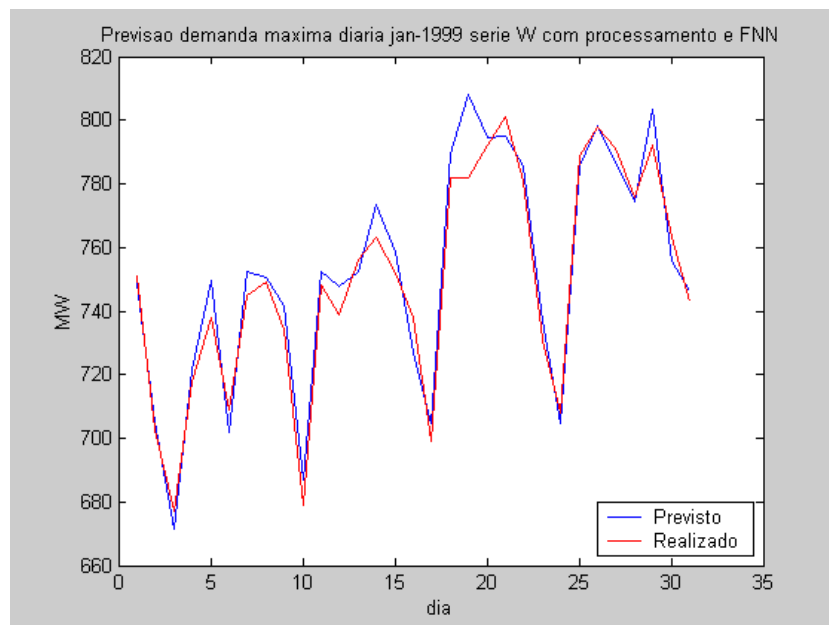


Figura 6.26. Previsto x Realizado usando Série W com/ proc (FNN)

6.11.2 Resultados Utilizando Amostra de Janeiro-Fevereiro (JF)

Nesta seção são utilizados de dados de carga de 30 minutos dos meses de Janeiro e Fevereiro dos anos 1997 e 1998. Em particular, são apresentados os resultados obtidos ao trabalhar com a série processada nas suas componentes de tendência (T) e sazonalidade (S).

Tabela 6.44 é mostrada a aplicação dos métodos de reconstrução utilizados nesta tese. Observe que ao igual que nas seções anteriores, são calculados dois valores diferentes para a variável dimensão da imersão (m), uma usando FNN e uma segunda utilizando D_2 . Na Tabela 6.45 são exibidos os Espaços de Imersão reconstruídos, isto é, as entradas que serão utilizadas na rede neural visando realizar a previsão de demanda máxima diária.

TABELA 6.44. Resultados das variáveis de reconstrução

Dados da série	Informação mútua τ	FNN (precisão Max) m	D_2 m
Série_JF _C/proc	(τ) med= 4	39	19

TABELA 6.45. Resultados do Espaço de reconstrução.

Série usada	Passo de Reconstrução τ	Dimensão da Imersão m	Espaço de Imersão
Série_JF_C/proc	4	39	$Y(t)=\{ x(t-4),x(t-8),\dots,x(t-152) \}$
Série_JF_C/proc	4	19	$Y(t)=\{ x(t-4),x(t-8),\dots,x(t-72) \}$

Os resultados da Tabela 6.46 indicam um melhor desempenho ao utilizar, junto com o Passo de Reconstrução 4, a Dimensão da Imersão 39 fornecida pelo método dos falsos vizinhos FNN. Nas figuras 6.27 e 6.28 é apresentada a variação entre a demanda realizada e a prevista pelas RNA utilizadas, com resultados bastantes satisfatórios em ambos os casos.

TABELA 6.46. Resultados das estatísticas de previsão

MAPE	Erro Máximo	Técnica Dimensão da Imersão m
1,058%	3,228%	D_2
0,905%	3,200%	FNN

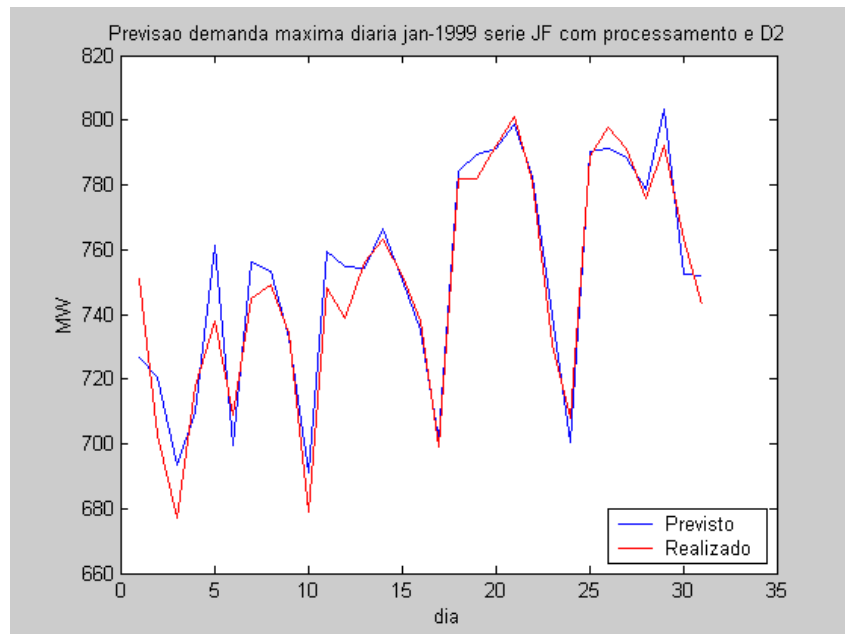


Figura 6.27. Previsto x Realizado usando Série JF com/ proc e (D_2)

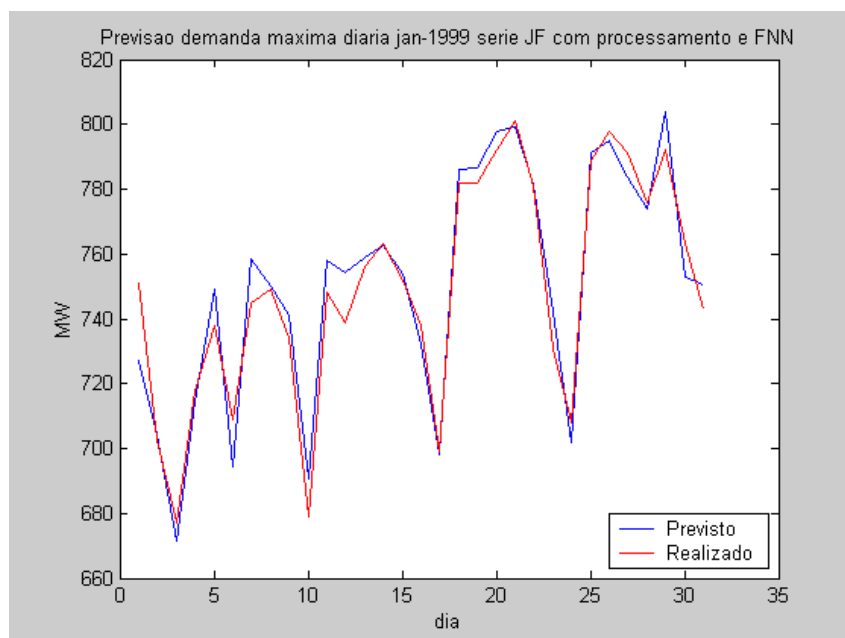


Figura 6.28. Previsto x Realizado usando Série JF com/ proc e (FNN)

6.11.3 Comparação de Resultados

A seguir são apresentadas duas tabelas com os resultados obtidos nesta tese e sua comparação com os apresentados pelos autores em [64]. Na Tabela 6.47, são apresentados os resultados obtidos ao utilizar a amostra W, onde é possível observar a superioridade das RNA que foram treinadas com entradas fornecidas pelos métodos de reconstrução, prévio processamento em T e S. A mesma afirmação pode ser realizada ao observar os resultados da Tabela 6.48, ressaltando que independente do método para cálculo da Dimensão da Imersão escolhido, os resultados foram superiores aos apresentados em [64]

Tabela 6.47. Resultados usando amostra de inverno (W)

Amostra p/treinamento e reconstrução	Método para Dimensão da Imersão m	MAPE
Resultado para série W no artigo [64]	-----	1,950%
Série W sem processamento	D ₂	3,715%
Série W sem processamento	FNN	2,918%
Série W com processamento	D ₂	1,616%
Série W com processamento	FNN	0,835%

Tabela 6.48. Resultados usando amostra de jan-fev (JF)

Amostra p/treinamento	Método para Dimensão da Imersão m	MAPE
Resultado para série JF no artigo [64]	-----	2,540%
Série JF com processamento	D ₂	1,058%
Série JF com processamento	FNN	0,905%

As RNA utilizadas nesta seção seguem o padrão simples definido na seção 6.2, isto é, redes com três neurônios (tangente hiperbólica) na camada escondida, um neurônio linear na camada de saída e treinadas durante 500 épocas. A previsão realizada foi de um passo a frente para um período de 1488 dados de 30 minutos, isto é, um mês. Neste exemplo não foram utilizados os dados de temperatura, no entanto pelos resultados apresentados em [64], os mesmos não contribuíram para a melhoria nos resultados, muito pelo contrário, foram os que tiveram pior desempenho.

6.12 Conclusões

Neste capítulo objetivou-se a aplicação do Teorema de Takens em séries de carga de energia elétrica. Para isto, foram aplicadas as técnicas exibidas nos capítulos 2 e 3: método dos Falsos Vizinhos mais Próximos, método da Dimensão da Correlação e método da Informação Mútua. Foi testado o efeito no desempenho das técnicas de reconstrução, procedimentos que permitem a diminuição das componentes de ruído via aplicação da Teoria de *Wavelets*, e a diminuição das componentes de tendência e sazonalidade presente nas séries. Também neste capítulo foi avaliado o efeito que a resolução dos dados causa na qualidade dos resultados dos métodos de reconstrução. Foram apresentados diversos exemplos utilizando quatro séries de carga, três delas bastante utilizadas na literatura de previsão e uma série de carga de uma distribuidora nacional. Sobre os resultados dos exemplos podem-se fazer os seguintes comentários:

- os resultados dos métodos de reconstrução variam em função da amostra utilizada. O que significa que deve-se ter especial cuidado na escolha da amostra utilizada no processo de reconstrução, porque embora os dados são oriundos de um único “sistema gerador”, o mesmo pode apresentar dinâmicas muito diferentes ao longo do tempo;
- em todos os exemplos apresentados neste capítulo mostrou-se a conveniência de utilizar uma série de carga processada em T_t e S_t , antes do processo de reconstrução. Neste caso estima-se que a não estacionariedade das séries de carga de alguma forma “mascaram” a verdadeira dinâmica dos dados, desta forma se faz necessário “purificar” os dados visando que os métodos de reconstrução sejam capazes de enxergar a verdadeira dinâmica presente na série temporal;
- no exemplo 6.8 mostrou-se que a decomposição de *wavelets* é benéfica do ponto de vista de reconstrução, já que ajuda a eliminar o ruído presente na série de carga. Este fato mostra coerência com as premissas teóricas levantadas por Takens para a aplicação de seu teorema;
- no entanto no exemplo 6.9 mostrou-se que o critério de utilização da variável temperatura com os mesmos *lags* usados para carga, não sempre produz bons resultados, no exemplo apresentado na seção 6.7.5 foi verificado que a inclusão da variável de temperatura, usando os mesmos *lags* definidos para a carga,

proporcionou melhoria nos resultados de previsão. Estes fatos talvez um pouco contraditórios levam à conclusão que o tratamento da variável temperatura deve observar cuidados e metodologias especiais. A relação entre as variáveis de carga e temperatura possui características não lineares bastante complexas que devem ser abordadas de maneira multivariada e independente;

- foi verificado no exemplo 6.10 que o nível de resolução fornece maior grau de informação para os métodos de reconstrução, beneficiando seu desempenho;
- o exemplo 6.9 também mostrou melhorias quando aplicado critérios de processamento de T_t e S_t e decomposição de *wavelets* de maneira conjunta;
- o exemplo 6.10 permitiu a avaliação isolada dos efeitos de *wavelets*, processamento de T_t e S_t e nível de resolução da série;
- numa série de carga sem processamento de T_t e S_t , em geral, quanto maior o intervalo de dados da amostra, maior o valor do Passo de reconstrução;
- o valor do Passo de reconstrução diminui bastante quando a série de carga é processada em T_t e S_t ;
- ainda que utilizando uma arquitetura e um critério de treinamento simples, no exemplo 6.11 a técnica aplicada nesta tese mostrou-se superior à aplicada pelos autores em [64], o que mostra o potencial da metodologia.

CAPÍTULO 7

CONCLUSÕES

O objetivo da tese aqui apresentada foi a investigação de processos de extração de variáveis relevantes objetivando o aperfeiçoamento da previsão de carga de curto prazo em sistemas elétricos. O foco foi concentrado na utilização de métodos físicos que permitam a extração de variáveis relevantes para seu uso em previsão de carga. Para isto foram utilizadas exclusivamente, as informações contidas nas séries temporais de carga analisadas. Foram estudados, avaliados e comparados alguns métodos de reconstrução do Espaço Fase ou Espaço Estado baseados no Teorema de Takens (método da Dimensão de Correlação, método dos Falsos Vizinhos mais Próximos) além de métodos baseados na Teoria da Informação (método da Informação Mútua). Os métodos estudados tiveram seu origem na física, particularmente no estudo de sistemas determinísticos com características caóticas, no entanto neste trabalho os mesmos foram aplicados em sistemas não lineares geradores de séries de carga que possuem natureza estocástica não linear e multivariada.

Em relação à influência que a não estacionariedade das séries de carga causam nos métodos de extração de variáveis, ficou evidente em repetidas ocasiões, que eliminar ou diminuir a presença da não estacionariedade, ajuda consideravelmente aos resultados fornecidos pelos métodos de reconstrução. O motivo é que, a não estacionariedade das séries de carga de alguma forma “mascaram” a verdadeira dinâmica dos dados, assim se faz necessário “purificar” os dados visando que os métodos de reconstrução, sejam capazes de enxergar a verdadeira dinâmica presente na série temporal;

No referente às conseqüências que o nível de resolução das séries de carga causam no desempenho dos métodos de extração, foi mostrado que um nível de resolução alto nos dados da série, fornece maior grau de informação para os métodos de reconstrução, beneficiando assim seu desempenho.

Visando realizar um tratamento na presença de ruído nas séries de carga, foi aplicada a Teoria de *Wavelets* a qual permitiu, em todos os casos testados, uma melhoria nos resultados de reconstrução do Espaço de Imersão. Como está técnica permite a decomposição em vários níveis e conseqüentemente a observação de diversos padrões de ruído, é necessário tomar cuidado para não extrair variáveis de alta freqüência que

contenham informações úteis para a reconstrução do processo. Respeitar esse ponto de equilíbrio é chave para o uso correto das vantagens desta metodologia.

Foi investigada a sensibilidade dos métodos físicos de extração de variáveis, quando aplicadas variações na quantidade de dados amostrais (comprimento da série). Os resultados indicam que os métodos de reconstrução variam em função da amostra utilizada, o que significa que deve-se ter especial cuidado na escolha da amostra usada no processo de reconstrução, porque embora os dados são oriundos de um único “sistema gerador”, o mesmo pode apresentar diversas dinâmicas ao longo do tempo. Por exemplo, a dinâmica na estação de inverno é diferente da dinâmica do verão, assim como os janeiros são diferentes dos fevereiroiros. Além disso, dentro de uma amostra contendo, por exemplo, somente janeiros é possível encontrar várias dinâmicas diferentes. A escolha deve estar associada ao horizonte de previsão que se deseja atingir e à experiência do pesquisador.

O critério simplista de utilização da variável temperatura histórica com os mesmos *lags* usados para carga, em alguns casos produz bons resultados, no entanto em outros fracassa, inclusive produzindo resultados piores que redes que, nem sequer consideram a presença desta variável.

Quando incluída na RNA a variável temperatura prevista, os resultados são sempre superiores que quando comparado com redes que não trabalham com esta variável climatológica. No entanto a grande dificuldade é depender de uma previsão de temperatura confiável, o que não um problema simples de ser resolvido.

Estes fatos levam à conclusão que o tratamento da variável temperatura devem-se observar cuidados especiais e metodologias apropriadas. A relação entre as variáveis de carga e temperatura apresenta características não lineares bastante complexas que devem ser abordadas de maneira multivariada e independente;

O Teorema de Takens tem como premissa a reconstrução do Espaço Estado a partir de um sinal unidimensional. No entanto Takens colocou como condição a existência de séries infinitas e sem ruído. A primeira condição é impossível de ser cumprida quando se trabalha com séries de carga reais, desta forma fica evidente que para aproximar os métodos das hipóteses, deve se utilizar a maior quantidade de dados disponível. A presença de ruído, também é freqüente em séries de carga, assim novamente se faz necessária a aplicação de técnicas de redução de ruído no sinal. Neste trabalho foi utilizada a teoria de *Wavelets* com resultados satisfatórios.

Os métodos para calcular a variável Dimensão da Imersão foram criados acima de métodos determinísticos com características caóticas, no entanto sabe-se que as séries de carga não tem características totalmente determinísticas, mas também não são completamente aleatórias, de fato estas séries possuem um pouco de cada uma destas características. Este fato teórico é importante, porque se não tratado o ruído, o desempenho dos métodos pode ficar comprometido. Como já mencionado antes, neste trabalho foi utilizada a teoria de *Wavelets* com resultados satisfatórios.

Foi mostrado que em geral, para o cálculo do Passo de Reconstrução τ , o método de Informação Mútua tende a apresentar valores maiores, na medida que é incrementado o tamanho da amostra. Por outro lado, este método não é muito sensível ao aumento no grau de detalhes, isto é, à diminuição da partição.

No cálculo da Dimensão da Imersão m , foi observado que o método dos Falsos Vizinhos é menos sensível ao tamanho da amostra, e que a diminuição da variável de tolerância R pode em alguns casos atrasar o processo de convergência. Já o método da Dimensão da Correlação, apresenta valores de m que tendem a crescer com o tamanho amostral. Em geral, os métodos utilizados para o cálculo da Dimensão da Imersão não apresentam resultados semelhantes entre se, de fato nesta tese recomenda-se que para escolher os resultados da variável Dimensão da Imersão entre os dois métodos disponíveis seja aplicado um critério de parcimônia.

A Dimensão da Correlação é uma forma aproximada de calcular a dimensão de uma atrator e pode também fornecer informações sobre o grau de aleatoriedade de um sinal, de fato este método não converge quando existe presença de componentes aleatórias. Isto ajuda a justificar a aplicação de técnicas para redução de ruído antes da aplicação desta técnica.

Embora não aplicado nesta tese, uma utilidade importante do cálculo dos invariantes geométricos, como por exemplo dos expoentes de Lyapunov, é verificar o grau de determinismo ou não presente num sinal. Isto acontece de fato, quando existe pelo menos um expoente positivo, a partir de aí pode-se afirmar que trata-se de um sistema caótico.

Esta tese apresentou as potencialidades de utilizar uma técnica de reconstrução de Espaço de Imersão baseada em informações contidas exclusivamente numa série temporal univariada. Foram aplicadas técnicas não lineares de extração de variáveis de entrada visando reconstruir o Espaço de Estado de um sistema desconhecido. Embora ainda é necessário aprofundar em vários dos aspectos apresentados nesta tese, a

metodologia mostra-se útil e aplicável à seleção de variáveis de entradas para previsão de carga de curto prazo.

7.1 Trabalhos futuros

A seguir alguns dos aspectos que devem ser pesquisados em trabalhos futuros:

- O teorema de Takens que é a base de sustentação desta tese, está baseado em sistemas caóticos determinísticos. As aplicações nesta tese, foram realizadas em séries de carga finitas com características estocásticas. Assim, é necessário investigar exaustivamente quais os limites “no relaxamento” das premissas determinísticas realizadas neste trabalho.
- Avaliação do grau de violação das hipóteses usando para isto séries sintéticas. Ditas séries podem ser geradas a partir de modelos auto-regressivos.
- Utilizar outras metodologias para remoção da sazonalidade diferentes das propostas por Box e Jenkins.
- Investigar uma metodologia de extração de variáveis relevantes que permita o tratamento simultâneo de carga, temperatura e eventualmente de outras variáveis explicativas.

ANEXO A

GLOSSÁRIO DE TERMOS

Atrator: região no Espaço Fase dos sistemas dissipativos até a qual convergem as trajetórias numa determinada região chamada de bacia de atração. Os atratores mais simples são o ponto e o ciclo limite, que correspondem a movimentos periódicos e que, por tanto são curvas fechadas.

Atrator de Lorenz: atrator estranho com forma de borboleta. Originado a partir de um modelo meteorológico desenvolvido por Edward Lorenz. Possui três equações e três variáveis. Foi um dos primeiros atratores estranhos estudados.

Auto similaridade: descreve a geometria de objetos nos quais uma pequena parte quando expandido mostra-se como um todo.

Atrator Estranho: atrator de N pontos com N igual a infinito. Geralmente (talvez sempre) com auto similaridade.

Bacia de Atração: região do Espaço Fase associado com um determinado atrator. A bacia de atração de um atrator é o conjunto (inicial) de pontos que converge ao atrator.

Bifurcação: mudança qualitativa no comportamento da dinâmica de um sistema (atrator) associada a mudanças no parâmetro de controle.

Caos: comportamento de um sistema dinâmico que tem (a) um número de atratores muito grande (possivelmente infinito) e b) sensível às condições iniciais.

Condição Inicial: ponto de partida de um sistema dinâmico.

Ciclo Limite: atrator que é periódico no tempo. Sistemas que apresentam oscilações auto sustentadas mesmo na ausência de um forçamento periódico externo.

Diagrama de Bifurcação: representação visual dos sucessivos períodos duplos produzidos por mudanças no parâmetro de controle.

Dimensão: vide dimensão da imersão, dimensão de box-counting, dimensão da correlação, dimensão da informação, dimensão de um sistema.

Dimensões de um Sistema: conjunto de variáveis de um sistema dinâmico.

Dimensão da Imersão: Vide Espaço Fase.

Dimensão Fractal: medida de um objeto geométrico que pode ter valores fracionais. Tem como sinônimo a dimensão de Hausdorff. A dimensão fractal é geralmente usada para medir a velocidade com que uma distância, área, ou volume aumentam ou diminuem.

Dimensão de Hausdorff: medida de um objeto geométrico que pode possuir valores fracionários (vide dimensão fractal).

Espaço Fase: é o espaço de definição das variáveis dependentes que descrevem o sistema. Por exemplo, no sistema dinâmico $\dot{x} = f(x)$, x é um ponto no Espaço Fase cuja evolução temporal está definida pelo vetor f que define a derivada temporal de x . Dito de outra maneira é o espaço constituído pelas variáveis do sistema dinâmico. Cada ponto do Espaço Fase representa um possível estado do sistema. A evolução no tempo do sistema representa-se como uma trajetória no Espaço Fase.

Expoente de Lyapunov: é uma quantidade que serve para caracterizar a divergência das trajetórias no tempo num processo caótico.

Equação de Diferencias: função que define a mudança de uma variável de um ponto a outro no domínio do tempo.

Equação Diferencial: equação que especifica a taxa de mudança em variáveis contínuas devido a mudanças em outras variáveis (por exemplo no domínio do tempo)

Espaço Fase (espaço estado): um espaço abstrato usado para representar o comportamento de um sistema. Suas dimensões são as variáveis do sistema. Então um ponto no Espaço Fase define o potencial estado do sistema.

Fluxo: as soluções de um sistema de equações diferenciais podem ser conhecidas como uma função que para um determinado tempo t e para um determinado ponto x no Espaço Fase, $x \in M$, produz um único ponto $x(t)$ no Espaço Fase. Formalmente, o fluxo F é definido como a função contínua:

$$F : M \times \mathfrak{R} \rightarrow M \text{ com } F(x,t) = x(t) \text{ para todo } x \in M \text{ e } t \in \mathfrak{R}$$

onde x pode ser reconhecida como a condição inicial a $t=0$, e $x(t)$ é a evolução temporal desta condição após de um tempo t .

Fractais: são conjuntos de pontos ou curvas que mostram uma estrutura cada vez mais detalhada na medida que são observadas com maior amplificação. Muitos fractais são autosimilares, isto é, a figura tem aspecto similar em cada passo de amplificação em distintas escalas. A dimensão geométrica de um fractal pode não ser inteira.

Fractal: forma irregular que possui autosimilaridade. Tem infinitos detalhes que não podem ser diferenciados.

Iteração: aplicação repetida de uma função, usando as suas saídas como entradas para as próximas aplicações.

Mapa de retorno: gráfico dos valores n vs. $n+1$ de uma série temporal.

Número de Lyapunov: valor de um expoente, coeficiente de tempo, que reflete a taxa de partida de órbitas dinâmicas. Mede a sensibilidade às condições iniciais.

Órbita (trajetória): uma seqüência de posições (caminhos) de um sistema em seu Espaço Fase. O caminho vai desde o ponto inicial (condições iniciais) até dentro do atrator.

Parâmetro de Controle: parâmetro das equações de um sistema dinâmico, caso sejam permitidas mudanças no parâmetro de controle a dinâmica do mesmo será alterada mudará. Mudanças que ultrapassem determinados valores podem provocar bifurcações.

Período Duplo: mudanças na dinâmica na qual um atrator de N -pontos é substituído por um atrator de $2N$ pontos.

Retrato de Fases: coleção de todas as trajetórias a partir de todos os possíveis pontos de partida no Espaço Fase de um sistema dinâmico.

Sistema Dinâmico: conjunto de equações que definem a forma como determinadas variáveis mudam no tempo. As equações especificam como calcular os novos valores em função de seus variáveis e seus parâmetros de controle. As funções quando explícitas, são equações diferenciais ou equações de diferenças. Os sistemas dinâmicos podem ser estocásticos ou determinísticos. Num sistema estocástico, os novos valores são oriundos de uma distribuição de probabilidade. Num sistema determinístico, um novo valor está associado com os valores das variáveis do sistema.

Sensibilidade às Condições Iniciais: é uma propriedade de sistemas caóticos. Um sistema dinâmico tem sensibilidade às condições iniciais quando muito pequenas diferenças nos valores iniciais resultam em comportamentos muito diferentes. Se as órbitas de pontos vizinhos divergem, o sistema é sensível às condições iniciais.

Torus: um atrator com N oscilações independentes. Plotado no Espaço Fase, um torus de 2-oscilações lembra um *donut*.

Vetor: dois valores medidos e associados com o Espaço Fase de um sistema dinâmico. O primeiro indica a direção do sistema a partir de um ponto e o segundo indica a velocidade.

Vetor de Fluxo: conjunto de todos os vetores no Espaço Fase de um sistema dinâmico. Para um dado sistema contínuo, o vetor de fluxo é especificado por um conjunto de equações diferenciais.

ANEXO B

MAPA LOGÍSTICO

B.1 Introdução

Neste anexo são apresentadas as simulações realizadas com o mapa logístico [28]. Este sistema é muito usado em diversas áreas principalmente em biologia, e embora matematicamente simples possui características que ajudam a ilustrar os conceitos apresentados no capítulo 2.

$$x(n+1) = r \cdot [1 - x(n)] \quad (\text{B.1})$$

B.2 Para valores de $r < 1$

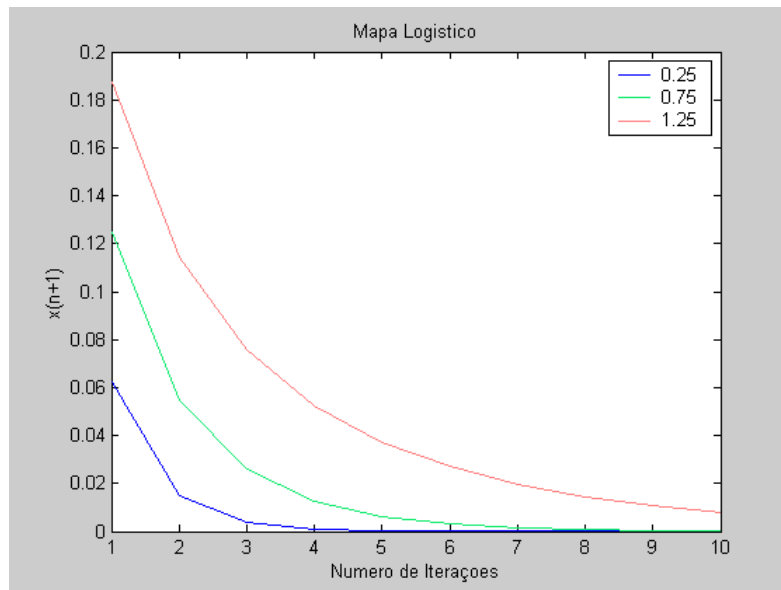


Figura B.1. Comportamento do Mapa Logístico para $r = 0.25$, 0.50 , e 0.75 . Em todos os casos $x(1) = 0.5$.

Quanto mais distante de 1 o valor de r , mais rápida a convergência para o valor de $x = 0$. A situação anterior representa um atrator de um ponto.

B.3 Comportamento para $1 < r < 3$

Agora independentemente do valor inicial de x , temos um atrator de um ponto com valor diferente de zero.

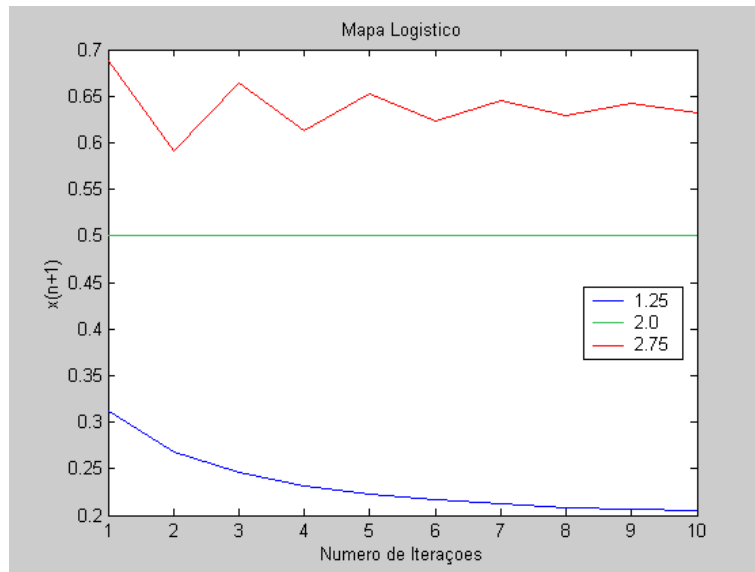


Figura B.2. Comportamento do mapa logístico para $r=1.25$, 2.00 , e 2.75 . Em todos os casos $x(1)=0.5$.

B.4 Quando r é maior que 3

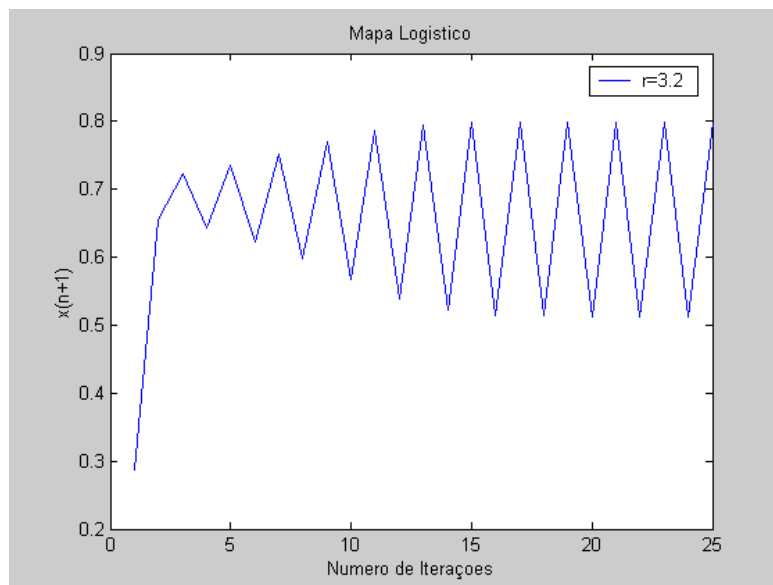


Figura B.3. Comportamento do mapa logístico para $r=3.2$.

Quando r é maior que 3, o sistema se alterna entre dois pontos. Desta forma pode-se concluir que neste caso existem um atrator de dois pontos. Se configura assim, uma bifurcação ou período duplo,

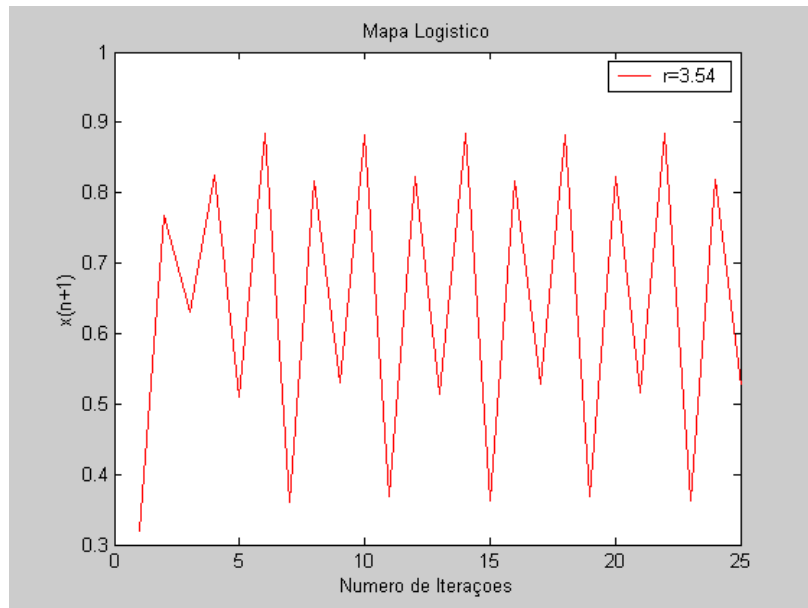


Figura B.4. Comportamento do mapa logístico para $r= 3.54$. Atrator de 4 pontos.

B.5 Atrator de N pontos

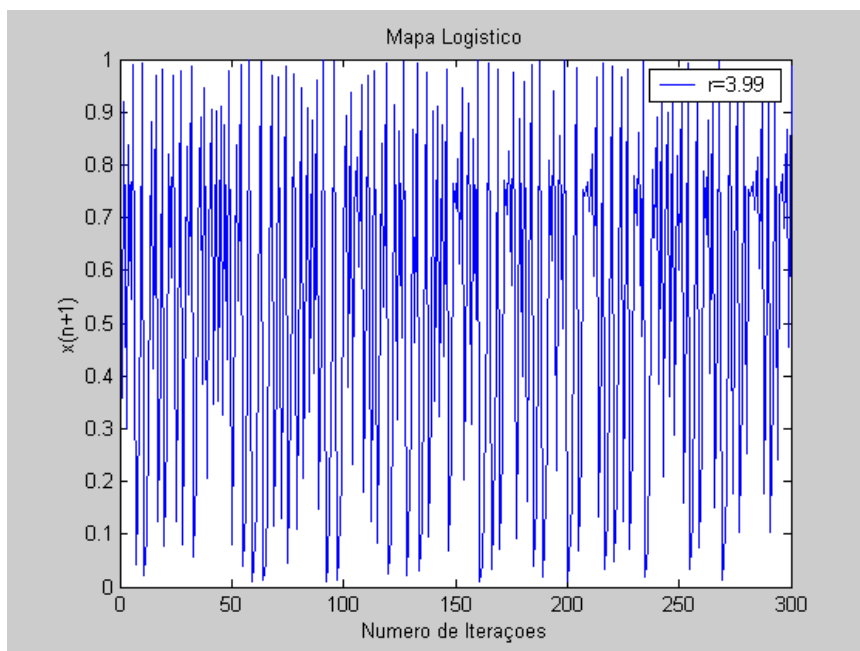


Figura B.5. Comportamento caótico do mapa logístico $r= 3.99$

B.6 Diagrama de Bifurcação

Uma bifurcação e um periodo-duplo, um cambio de um atrator de N pontos a um atrator de $2N$ pontos de atratores, o qual ocorre quando o parâmetro de controle e

mudado. Um diagrama de bifurcação e um resumo visual da sucessão de um período duplo causado por incrementos de r .

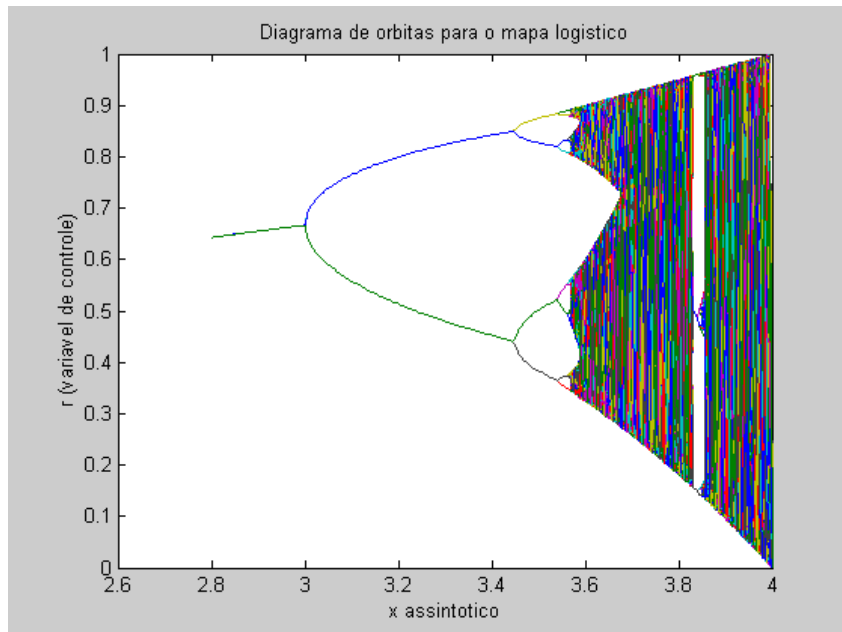


Figura B.6. Diagrama de bifurcação entre 0 e 4

Pode-se observar que para r menor que 1 todos os valores plotados são zero. Para r menor que 1, o ponto de atração é zero. Para r entre 1 e 3, ainda existe um atrator, porém o valor atraído de x aumenta com o incremento de r , até que $r=3$. As bifurcações ocorrem para $r=3$, $r=3.45$, 3.54 , 3.564 , 3.569 (aproximadamente), etc., até que r ultrapassa o valor de 3.57, onde o sistema é caótico. No entanto, o sistema não é caótico para todos os valores superiores a 3,57.

B.7 Sensibilidade as condições Iniciais

É uma outra importante característica que emerge de uma região caótica. Observe que As duas series permanecem juntas nas primeiras 10 iterações. Porém após deste período, não é possível afirmar o mesmo. Na seguinte figura as duas series são inicializadas com valores menos distantes que no caso anterior. Neste segundo caso $x_1=0.3$ e comparada com $x_1=0.3000001$

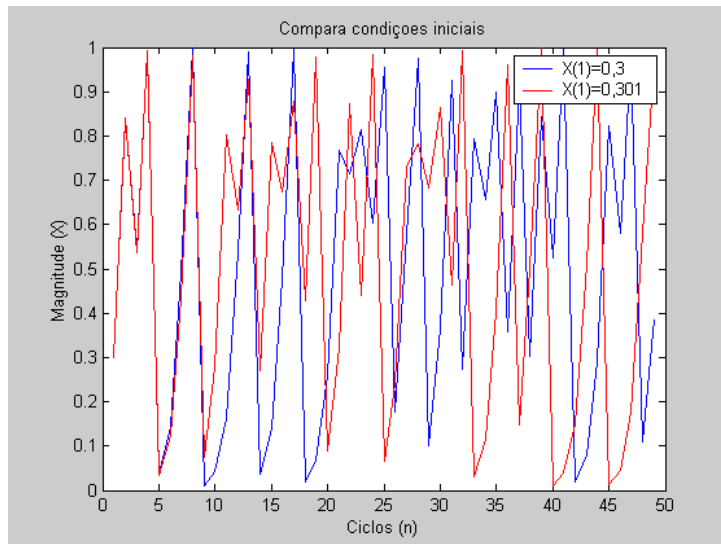


Figura B.7. Duas series temporais para $r=3.99$, $x_1=0.3$ comparada com $x_1=0.301$

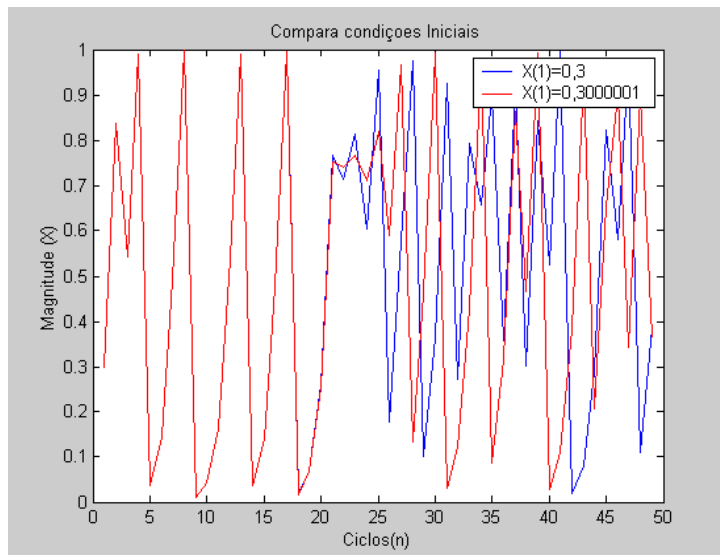


Figura B.8. Duas series temporais para $r=3.99$, $x_1=0.3$ comparada com $x_1=0.3000001$

Desta vez as series permanecem juntas durante um tempo maior, porem após das primeiras 24 iterações as series se distanciam. Para saber quão independente as duas series são, a seguir é mostrado um gráfico de dispersão para as duas series, antes e após das primeiras 24 iterações.

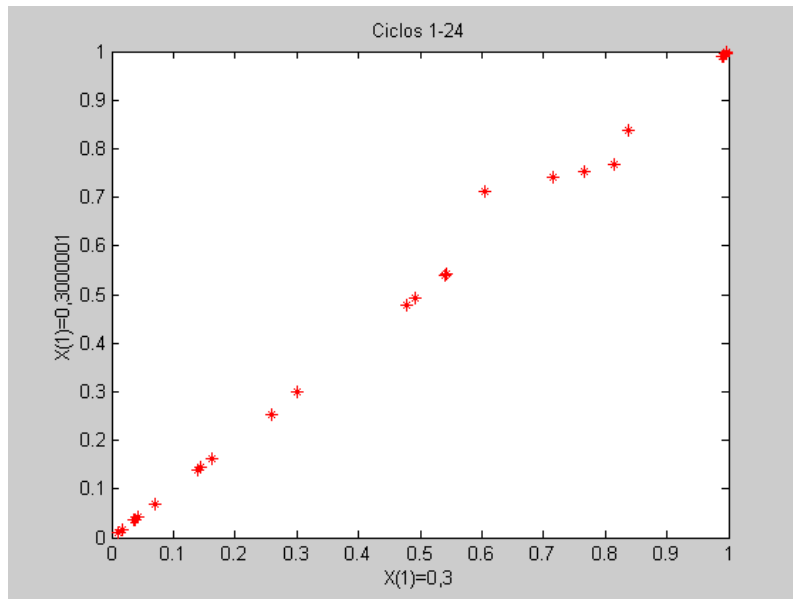


Figura B.9. Gráfico de dispersão de series começando em 0.3 vs. series começando em 0.3000001. Primeiros 24 ciclos

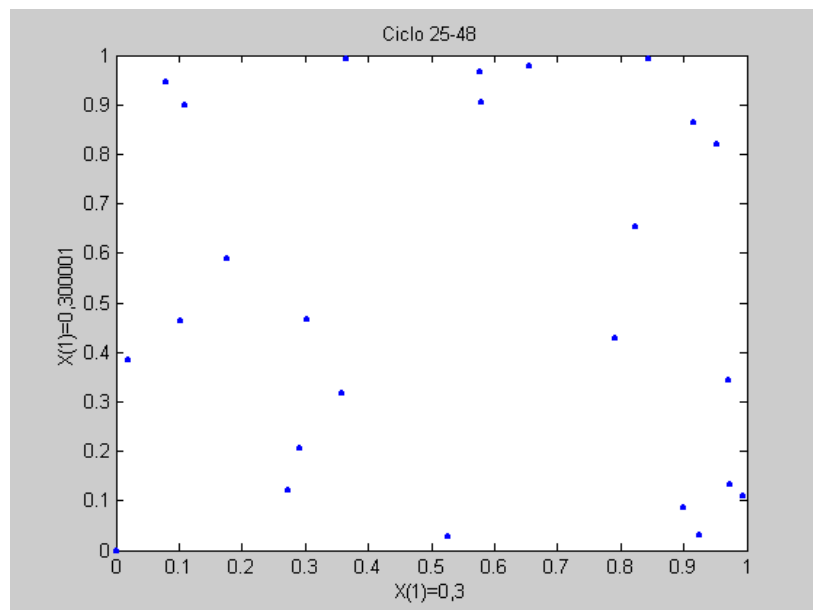


Figura B.10. Gráfico de dispersão de series começando em 0.3 vs. series começando em 0.3000001. Últimos 24 ciclos

A correlação após das primeiras 24 iterações é aproximadamente zero. Ilustramos aqui um dos sintomas do caos. Um sistema caótico é aquele para o qual a distância entre duas trajetórias vizinhas no espaço estado divergem no tempo. A magnitude da divergência aumenta exponencialmente num sistema caótico.

Significa que um sistema caótico, definido por uma regra simples é em principio imprevisível. Temos então aqui um exemplo em que uma diferença de $0,1 \cdot 10^{-6}$, produz como resultado falhas na previsão a partir da vigésima quarta iteração.

ANEXO C

TENDÊNCIA e SAZONALIDADE: Conceitos Básicos

C.1 Introdução

O objetivo deste anexo é apresentar de forma sucinta algumas das técnicas utilizadas neste trabalho para eliminação ou diminuição das componentes de tendência e sazonalidade presentes em séries de carga.

Considere as observações $\{Z_t, t = 1, \dots, N\}$ de uma série temporal. Assumindo-se que um modelo de decomposição consiste em escrever Z_t como uma soma de três componentes não-observáveis,

$$Z_t = T_t + S_t + a_t \quad (\text{C.1})$$

onde T_t e S_t representam a tendência e sazonalidade, respectivamente, enquanto a_t é uma componente aleatória, de média zero e variância constante δ_a^2 . Se $\{a_t\}$ for um ruído branco, então $E(a_t a_s) = 0$, $s \neq t$; mas pode-se, eventualmente, relaxar esta suposição, tomando-se $\{a_t\}$ como um processo estacionário. Segue-se que $\{Z_t\}$ será uma série não-estacionária.

O interesse principal em considerar um modelo do tipo (C.1) será o de estimar S_t e construir a série livre de sazonalidade ou sazonalmente ajustada. Isto é, se \hat{S}_t for uma estimativa de S_t ,

$$Z_t^{SA} = Z_t - \hat{S}_t \quad (\text{C.2})$$

será a série sazonalmente ajustada. Há varias razões para considerar este procedimento de ajustamento sazonal. As componentes T_t e S_t são, em geral, bastante relacionadas e a influência da tendência sobre a componente sazonal pode ser muito forte, por dois motivos [56]:

- (a) métodos de estimação de S_t podem ser bastante afetados se não levarmos em conta tendência;
- (b) a especificação de S_t depende da especificação de T_t .

Por isso, não poderemos isolar uma das componentes sem tentar isolar a outra. Estimando-se T_t e S_t e subtraindo de Z_t obteremos uma estimativa da componente aleatória a_t .

C.2 Tendências

Inicialmente considere-se vamos supor que a componente sazonal S_t não esteja presente, o modelo será:

$$Z_t = T_t + a_t \quad (\text{C.3})$$

onde a_t é ruído branco, com variância σ_a^2 .

Há vários métodos para estimar T_t , sendo os mais utilizados em:

- (i) ajustar uma função do tempo, como um polinômio, uma exponencial ou outra função suave de t ;
- (ii) suavizar (ou filtrar) os valores da série ao redor de um ponto, para estimar a tendência naquele ponto;
- (iii) suavizar os valores da série através de sucessivos ajustes de retas de mínimos quadrados ponderados (*lowess*).

Estimando-se a tendência através de \hat{T}_t , pode-se obter a série ajustada para tendência ou livre tendência,

$$Y_t = Z_t - \hat{T}_t$$

Um procedimento que é também utilizado para eliminar a tendência de uma série é aquele de tomar diferenças. Normalmente, para séries econômicas, por exemplo, a primeira diferença

$$\Delta Z_t = Z_t - Z_{t-1}$$

já é estacionária.

C.2.1 Tendência polinomial

Um procedimento muitas vezes utilizado é ajustar uma curva aos valores observados da série para estimar T_t e fazer previsões. Tradicionalmente são utilizados vários tipos de funções, como a exponencial e a logística, mas aqui limita-se a descrever brevemente o ajuste de um polinômio. O problema mais sério que se encontra ao se

estimar T_t através de um polinômio é que, embora ele possa ajustar-se bem ao conjunto de valores observados, extrapolações futuras podem ser bastante ruins.

Supondo-se, então que

$$T_t = \beta_0 + \beta_1 t + \dots + \beta_m t^m, \quad (\text{C.4})$$

onde o grau m do polinômio é bem menor que o número de observações N . Para estimar os parâmetros β_j utiliza-se o método dos mínimos quadrados, ou seja, minimiza-se

$$f(\beta_0, \dots, \beta_m) = \sum_{t=1}^N (Z_t - \beta_0 - \beta_1 t - \dots - \beta_m t^m)^2 \quad (\text{C.5})$$

obtendo-se os estimadores de mínimos quadrados usuais $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_m$. Este é um assunto conhecido (ver, por exemplo, Draper e Smith [57]) e não será discutido neste trabalho. Assim, como no problema usual de regressão, também aqui é possível transformar as “variáveis independentes” $1, t, t^2, \dots, t^m$ em “variáveis independentes ortogonais”. Detalhes podem ser obtidos em [58].

C.2.2 Suavização

Quando se supõe que a tendência possa ser representada por um polinômio de baixo grau, isto implica em usar todas as observações Z_t , $t=1, \dots, N$ para estimar o polinômio, que representará T_t sobre todo o intervalo de tempo considerado.

A idéia de aplicar algum tipo de suavização é que a tendência num instante t será estimada a partir das observações Z_s , com s ao redor de t . Por exemplo, utilizam-se as observações $Z_{t-n}, Z_{t-n+1}, \dots, Z_{t+n}$ para estimar T_t .

Existem vários métodos de suavização dos quais serão apresentados três.

C.2.2.1 Médias Móveis

Neste método utiliza-se um filtro linear, ou seja, uma operação que transforma a série Z_t na série Z_t^* :

Dado o modelo (C.3), transformando-o através de F , obtém-se

$$Z_t^* = T_t^* + a_t^*, \quad (\text{C.6})$$

onde $T_t^* = F[T_t]$, $a_t^* = F[a_t]$. Deseja-se que F seja tal que $T_t^* \approx T_t$ e $E(a_t^*) = 0$, de modo que, suavizando-se as observações Z_t , obtenha-se $F[Z_t] = Z_t^* \approx T_t$.

Dadas as observações Z_1, \dots, Z_n , o filtro F comumente utilizado é da forma

$$Z_t^* = \sum_{j=-n}^n c_j Z_{t+j}, \quad t = n+1, \dots, N-n, \quad (\text{C.7})$$

onde $\sum_{j=-n}^n c_j = 1$. Note-se que há perda de n observações no início e n no final da série original. Z_t^* será uma estimativa da tendência no instante t podendo-se dizer que (C.7) é um filtro de médias móveis. O caso mais simples é aquele em que $c_j = 1/(2n+1)$, para todo j , de modo que

$$Z_t^* = \frac{1}{2n+1} \sum_{j=-n}^n Z_{t+j} \quad (\text{C.8})$$

De (C.7), tem-se

$$Z_t^* = \sum_{j=-n}^n c_j Z_{t+j} = \sum_{j=-n}^n c_j [T_{t+j} + a_{t+j}] = \sum_{j=-n}^n c_j T_{t+j} + \sum_{j=-n}^n c_j a_{t+j},$$

ou seja,

$$Z_t^* = \sum_{j=-n}^n c_j T_{t+j} + a_t^*, \quad (\text{C.9})$$

onde

$$a_t^* = \sum_{j=-n}^n c_j a_{t+j}. \quad (\text{C.10})$$

Como $E(a_t^*) = 0$, pois $E(a_t) = 0$, para todo t , segue-se que

$$E(Z_t^*) = \sum_{j=-n}^n c_j T_{t+j} \approx \sum_{j=-n}^n c_j T_t = T_t = E(Z_t), \quad (\text{C.11})$$

dado que $T_{t+j} \approx T_t$, supondo-se a tendência “suave”. Assim, a série original e a série suavizada têm praticamente a mesma média, para casa t . Por outro lado,

$$\text{Var}(a_t^*) = \sigma_a^2 \sum_{j=-n}^n c_j^2, \quad (\text{C.12})$$

dado que $\text{Var}(a_t) = \sigma_a^2$, constante. Como $\text{Var}(Z_t) = \sigma_a^2$ e $\text{Var}(Z_t^*) = \text{Var}(a_t^*)$, segue-se de (C.12) que a série suavizada terá uma variância menor se $\sum_{j=-n}^n c_j^2 < 1$, $\forall j$. Mas o filtro (C.7) introduz uma correlação nos resíduos. De fato, a suposição inicial era que $E(a_t, a_s) = 0$, $s \neq t$. É fácil ver que

$$E(a_t^* a_{t+h}^*) = \begin{cases} \sigma_a^2 \sum_{j=-n+j}^n c_j c_{j-h}, & h = 0, 1, \dots, 2n \\ 0, & h = 2n+1, \dots \end{cases} \quad (\text{C.13})$$

A série livre de tendência será $Z_t - Z_t^*$ e $E(Z_t - Z_t^*)$ representa o viés de estimação, dado por

$$v(t) = T(t) - \sum_{j=-n}^n c_j T_{t+j}, \quad (\text{C.14})$$

e como já visto, $v(t) \approx 0$, já que $E(Z_t^*) \approx E(Z_t)$.

O erro quadrático médio do estimador é

$$E(T_t - Z_t^*)^2 = \text{Var}(Z_t^*) + v^2(t) = \sigma_a^2 \sum_{j=-n}^n c_j^2 + v^2(t). \quad (\text{C.15})$$

A idéia é escolher n (dado que σ_a^2 é conhecido) de modo que (C.15) seja o menor possível; todavia, o vício $v(t)$ e a variância Z_t^* variam de modo oposto em relação a n . Desta maneira, cabe ao usuário selecionar o valor de n adequado; para dados mensais usualmente tomamos médias móveis de 12 observações sucessivas.

Segundo Anderson [58] há três desvantagens principais neste processo de suavização:

- (i) inferências estatísticas derivadas do método são limitadas, dado que ele não é baseado em nenhum modelo probabilístico;
- (ii) não podemos obter as estimativas nos instantes $t = 1, \dots, n$ e $t = N - n + 1, \dots, N$;
- (iii) não fornece um meio de fazer previsões.

C.3 Sazonalidade

Nesta seção o objetivo será o estudo de uma série para a componente sazonal, ou seja, estimar S_t e subtrair a série estimada de Z_t no modelo (C.1),

$$Z_t = T_t + S_t + a_t, \quad t = 1, 2, \dots, N.$$

Desta maneira, um procedimento de ajustamento sazonal consiste em:

- (a) obter estimativas \hat{S}_t de S_t ;
- (b) calcular

$$Z_t^{SA} = Z_t - \hat{S}_t. \quad (\text{C.16})$$

Se o modelo for multiplicativo, da forma

$$Z_t = T_t S_t a_t, \quad (C.17)$$

a série sazonalmente ajustada será

$$S_t^{SA} = Z_t | \hat{S}_t. \quad (C.18)$$

Como já salientado, o modelo (C.17) é muitas vezes adequado para séries econômicas, que apresentam um crescimento exponencial. Tomando-se logaritmos, obtemos o modelo aditivo (C.1) para os logaritmos.

Ao se estimar S_t comete-se em geral, um erro de ajustamento sazonal, dado por

$$\delta_t = S_t - \hat{S}_t.$$

Dize-se que um procedimento de ajustamento sazonal é ótimo se minimizar $E(\delta_t^2)$.

A importância de se considerar procedimentos de ajustamento sazonal pode melhor ser ilustrada pelo seguinte trecho, extraído de Pierce [59]:

“tem havido no passado um interesse em se ter dados disponíveis sobre fenômenos importantes, sociais e econômicos, para os quais a variação sazonal foi removida. As razões relacionam-se, geralmente, com a idéia que a nossa habilidade em reconhecer, interpretar ou reagir a movimentos importantes não-sazonais numa série (tais como pontos de mudança e outros eventos cíclicos, novos padrões emergentes, ocorrências não esperadas para as quais causas possíveis são procuradas) é perturbada pela presença dos movimentos sazonais”.

Além disso, é difícil definir, tanto do ponto de vista conceitual como estatístico, o que seja sazonalidade.

Empiricamente, considera-se como sazonais os fenômenos que ocorrem regularmente de ano para ano, como um aumento de vendas de passagens aéreas no verão, aumento da produção de leite no Brasil nos meses de novembro, dezembro e janeiro, aumento de vendas no comércio na época do Natal etc [60].

Tomando-se como exemplo a Série A_1 : temperatura na Cananéia. Observa-se que a série tem um comportamento aproximadamente periódico, havendo semelhança a cada $s = 12$ meses. Denomina-se s de período, mesmo que o padrão não seja exatamente periódico. Aqui, pode-se formar uma tabela com $s (=12)$ colunas, uma para cada mês, e p linhas, sendo p o número de anos. Há dois intervalos de tempo que são importantes, mês e ano. O que se constata em séries sazonais é que ocorrem relações:

- (a) entre observações para meses sucessivos em um ano particular;
- (b) entre as observações para o mesmo mês em anos sucessivos.

Assim, a observação Z_t correspondente a janeiro de 1980 é relacionada com os demais meses de 1980, bem como com os demais meses de janeiro de 1979, 1981 etc.

Note-se a semelhança com Análise de Variância, os meses representando “tratamentos” e os anos representando as “réplicas”. Assim, Z_t é relacionada com Z_{t-1}, Z_{t-2}, \dots , mas também com Z_{t-s}, Z_{t-2s}, \dots . Isto implica que séries sazonais são caracterizadas por apresentarem correlação alta em *lags* sazonais, isto é, *lags* que são múltiplos do período s . Um procedimento de ajustamento sazonal será tal que esta correlação será destruída (ou pelo menos removida em grande parte).

Sem perda de generalidade, consideraremos o caso que temos dados mensais e o número total de observações, N , é um múltiplo de 12, isto é, $N=12p$, p = número de anos, de modo que os dados podem ser representados como na Tabela C.1.

A notação da tabela C.1 é padrão, com

$$\bar{Z}_i = \frac{1}{12} \sum_{j=1}^{12} Z_{ij}, \quad i = 1, \dots, p, \quad (C.19)$$

$$\bar{Z}_{.j} = \frac{1}{p} \sum_{i=1}^p Z_{ij}, \quad j = 1, \dots, 12, \quad (C.20)$$

$$\bar{Z} = \frac{1}{12p} \sum_{i=1}^p \sum_{j=1}^{12} Z_{ij} = \frac{1}{N} \sum_{t=1}^N Z_t \quad (C.21)$$

Vê-se, pois, que é conveniente reescrever o modelo (C.1) na forma

$$Z_{ij} = T_{ij} + S_j + a_{ij}, \quad i = 1, \dots, p, \quad j = 1, \dots, 12. \quad (C.22)$$

No modelo (C.22) tem-se que o padrão sazonal não varia muito de ano para ano, e pode ser representado por doze constantes .

No caso de sazonalidade não-constante, o modelo ficaria

$$Z_{ij} = T_{ij} + S_{ij} + a_{ij}, \quad i = 1, \dots, p, \quad j = 1, \dots, 12. \quad (C.23)$$

Existem vários procedimentos para se estimar S_t , sendo que os mais usuais são:

- (a) método de regressão e (b) método de médias móveis.

Um outro enfoque é incorporar a variação sazonal e a tendência em um modelo ARIMA, ou a um modelos de espaço de estados.

Tabela C.1: Observações mensais de uma série temporal com p anos.

Anos	Meses					Médias
	jan	fev	mar	...	dez	
	1	2	3		12	
1	Z_{11}	Z_{12}	Z_{13}	...	$Z_{1,12}$	\bar{Z}_1
2	Z_{21}	Z_{22}	Z_{23}	...	$Z_{2,12}$	\bar{Z}_2
\vdots	\vdots	\vdots	\vdots		\vdots	\vdots
p	Z_{p1}	Z_{p2}	Z_{p3}	...	$Z_{p,12}$	\bar{Z}_3
Médias	$\bar{Z}_{.1}$	$\bar{Z}_{.2}$	$\bar{Z}_{.3}$...	$\bar{Z}_{.12}$	\bar{Z}

C.3.1 Sazonalidade Determinística – Método de Regressão

Os métodos de regressão são ótimos para séries que apresentam sazonalidade determinística, ou seja, que podem ser previstas perfeitamente a partir de meses anteriores.

Então, no modelo (C.1), temos que

$$T_t = \sum_{j=0}^m \beta_j t^j ; \quad (C.24)$$

$$S_t = \sum_{j=1}^{12} \alpha_j d_{jt} , \quad (C.25)$$

onde $\{d_{jt}\}$ são variáveis periódicas (senos, cosenos, ou variáveis sazonais *dummies*) e a_t é ruído branco, com média zero e variância σ_a^2 .

Supondo-se sazonalidade constante, α_j não depende de t , podendo-se ter como exemplo,

$$d_{jt} = \begin{cases} 1, & \text{se o período } t \text{ correspondente ao mês } j, \quad j = 1, \dots, 12, \\ 0, & \text{caso contrário} \end{cases} \quad (C.26)$$

Neste caso,

$$d_{1t} + d_{2t} + \dots + d_{12,t} = 1, \quad t = 1, \dots, N, \quad (C.27)$$

de forma que a matriz de regressão não é posto completo, mas de posto $m+12$ (observe que existem $m+13$ parâmetros: $\alpha_1, \dots, \alpha_{12}, \beta_0, \beta_1, \dots, \beta_m$).

Impondo-se-se a restrição adicional

$$\sum_{j=1}^{12} \alpha_j = 0, \quad (\text{C.28})$$

obtém-se um modelo de posto completo

$$Z_t = \sum_{j=0}^m \beta_j t^j + \sum_{j=1}^{11} \alpha_j D_{jt} + \alpha_t, \quad (\text{C.29})$$

onde agora

$$D_{jt} = \begin{cases} 1, & \text{se o período } t \text{ corresponde ao mês } j, \\ -1 & \text{se o período } t \text{ corresponde ao mês } 12, \\ 0, & \text{caso contrário, } j = 1, \dots, 11. \end{cases} \quad (\text{C.30})$$

Deste modo pode-se utilizar a teoria usual de mínimos quadrados e obter os estimadores de α_j e β_j , ou seja, para uma amostra Z_1, \dots, Z_N , chega-se ao modelo

$$Z = C\beta + D\alpha + a, \quad (\text{C.31})$$

onde

$$\begin{aligned} Z_{N \times 1} &= \begin{bmatrix} Z_1 \\ \vdots \\ Z_N \end{bmatrix}, \quad C_{N \times (m+1)} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & 2^m \\ \vdots & \vdots & \dots & \vdots \\ 1 & N & \dots & N^m \end{bmatrix}, \\ \beta_{(m+1) \times 1} &= \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}, \quad D_{N \times 11} = \begin{bmatrix} D_{11} & D_{21} & \dots & D_{11,1} \\ D_{12} & D_{22} & \dots & D_{11,2} \\ \vdots & \vdots & \dots & \vdots \\ D_{1N} & D_{2N} & \dots & D_{11,N} \end{bmatrix}, \\ \alpha_{11 \times 1} &= \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{11} \end{bmatrix}, \quad a_{N \times 1} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}. \end{aligned} \quad (\text{C.32})$$

A equação (C.31) pode ser escrita na forma

$$Z = X\gamma + a, \quad (\text{C.33})$$

onde

$$X = [C : D] \text{ e } \gamma = \begin{bmatrix} \beta \\ \alpha \end{bmatrix},$$

de modo que

$$\hat{\gamma} = [X'X]^{-1} X'Z \quad (C.34)$$

são os estimadores usuais de mínimos quadrados.

C.3.2 Sazonalidade Estocástica – Método de Médias Móveis

Na seção C.2.2 estimamos-se T_t no modelo

$$Z_t = T_t + a_t$$

por meio de um filtro linear, ou seja,

$$\hat{T}_t = \sum_{j=-n}^n c_j Z_{t+j}, \quad t = n + \dots N - n. \quad (C.35)$$

A série $Z_t - \hat{T}_t$ estimará, então, a série residual a_t .

Agora, além de T_t , temos a componente S_t , que pretende-se estimar. O procedimento a ser utilizado é semelhante: dado o modelo (C.1), estima-se T_t através de (C.35) e considera-se

$$Y_t = Z_t - \hat{T}_t. \quad (C.36)$$

Esta série fornecerá meios para estimar S_t .

O método de médias móveis é apropriado quando se tem uma série temporal cuja componente sazonal varia com o tempo, ou seja, para séries cuja sazonalidade é estocástica.

Todavia, este procedimento é aplicado usualmente mesmo para padrão sazonal constante. Pode-se demonstrar [59] que este procedimento é ótimo para classe dos modelos ARIMA e como já salientado, embora este procedimento seja mais apropriado para o caso em que há um padrão sazonal constante.

Dado que a tendência é estimada por (C.35), as componentes sazonais supostas constantes, são estimadas num segundo estágio.

A partir de (C.36), toma-se as médias dos meses: janeiro, fevereiro etc.:

$$\bar{Y}_{.j} = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij}, \quad j = 1, \dots, 12, \quad (C.37)$$

usando a notação referente à Tabela C.1.

Como a soma dos $\hat{Y}_{.j}$ em geral não é zero, toma-se como estimativas das constantes sazonais

$$\hat{S}_j = \bar{Y}_{.j} - \bar{Y}, \quad (C.38)$$

onde

$$\bar{Y} = \frac{1}{12} \sum_{j=1}^{12} \bar{Y}_{.j}. \quad (\text{C.39})$$

O modelo pode ser escrito como

$$Z_t = T_t + S_j + a_t, \quad (\text{C.40})$$

com $t = 12i + j$, $i = 0, 1, \dots, p-1$, $j = 1, \dots, 12$, havendo p anos. Então, (C.37) pode ser escrita como

$$\begin{aligned} \bar{Y}_{.j} &= \frac{1}{p-1} \sum_{i=1}^{p-1} Y_{12i+j}, \quad j = 1, \dots, 6, \\ &= \frac{1}{p-1} \sum_{i=0}^{p-2} Y_{12i+j}, \quad j = 7, \dots, 12. \end{aligned} \quad (\text{C.41})$$

A série livre de sazonalidade é

$$Z_t^{SA} = Z_t - \hat{S}_t. \quad (\text{C.42})$$

É possível demonstrar que, para o modelo aditivo, não é necessário estimar T_t para obter os \hat{S}_j . Ou seja, é possível obter as médias mensais $\bar{Y}_{.j}$ diretamente dos desvios das médias mensais original em relação à média geral da série, mais um termo de correção dependendo somente do primeiro e últimos doze termos da série, ver Durbin [62].

Se o modelo for multiplicativo, obté-se $Y_t = Z_t / \hat{T}_t$, $\bar{Y}_{.j}$, \bar{Y} como antes e estima-se S_j por

$$\hat{S} = \bar{Y}_{.j} / \bar{Y}. \quad (\text{C.43})$$

Também,

$$Z_t^{SA} = Z_t / \hat{S}_t. \quad (\text{C.44})$$

ANEXO D

INTRODUÇÃO À TEORIA DE WAVELETS

D.1 Introdução

Neste Anexo é realizada uma introdução à Teoria de *Wavelets* ressaltando as potencialidades e vantagens quando comparada a outras técnicas de análise de sinais disponíveis na literatura. São ressaltadas principalmente a capacidade de detecção de anormalidades, erros e principalmente a possibilidade de identificação e extração de ruído no sinal.

D.2 Análise de Fourier

A análise de sinais sempre teve a disposição uma ampla gama de ferramentas, dentro das quais a análise de Fourier sempre foi a mais conhecida. Fourier desagrega um sinal temporal qualquer em sinais senóides com diferentes frequências. Uma outra forma de entender Fourier é como uma técnica matemática que transforma um ponto de vista temporal em outro ponto de vista baseado na frequência. Em muitos tipos de sinais, Fourier se configura como uma técnica muito útil porque a informação no espectro de frequências proporciona importantes informações[50]. Assim sendo, surge a seguinte pergunta: por que são necessárias outras técnicas?

Uma das razões é que Fourier possui um sério problema: ao levar o sinal para o domínio da frequência, todas as informações no domínio temporal são perdidas[50]. Quando se observa a transformada de Fourier de um sinal é impossível saber quando um determinado evento aconteceu. Se as propriedades do sinal não mudam muito num determinado espaço de tempo, isto é, se o sinal é estacionário, esta limitação passa a não ser tão importante. No entanto, a maioria dos sinais de interesse possuem características não estacionárias ou transitórias, que correspondem às partes mais importantes do sinal, que fogem à detecção da análise de Fourier.

D.2.1 Análise de Fourier de Curta Duração

Num esforço de corrigir estas deficiências, Gabor, em 1946 [51], adaptou a transformada de Fourier para a análise de somente uma pequena secção, definida como “janelas de tempo” do sinal. A adaptação de Gabor STFT (*Short Time Fourier Transform*) mapeia o sinal numa função de duas dimensões (tempo e frequência).

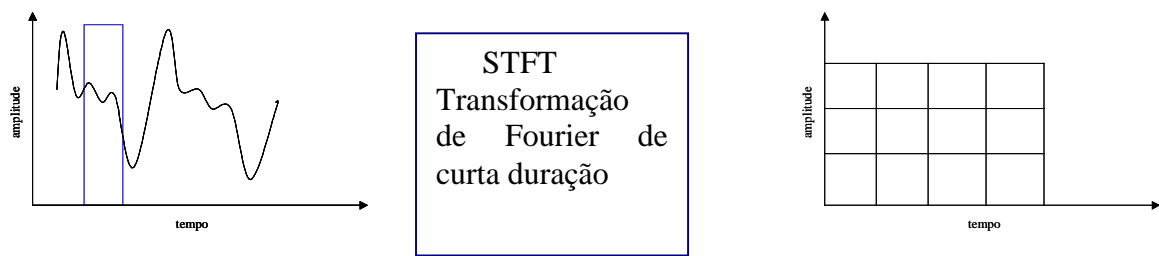


Figura D.1. Representação esquemática da STFT

O STFT representa uma visão do sinal considerando uma espécie de compromisso entre tempo e frequência, provendo informações de quando e com que frequência um determinado evento ocorre. No entanto, a precisão da informação depende do tamanho da janela de tempo utilizada. O problema é que uma vez escolhida uma determinada janela de tempo, o tamanho da mesma deve ser utilizada para toda frequência. Muitos sinais precisam de uma metodologia de análise mais flexível onde a janela de tempo possa ser modificada em função das necessidades.

D.3 Análise de *Wavelets*

Análise de *wavelets* é o passo natural na evolução da metodologia de Gabor: uma técnica que usa janelas de tamanho variáveis. A metodologia de *wavelets* permite a utilização de grandes intervalos de tempo quando se quer maior precisão na informação de baixa frequência e curtos intervalos quando se deseja informação de alta frequência.



Figura D.2. Representação esquemática da transformada de *wavelets*

A seguir, na Figura D.3, é possível analisar e comparar a análise de *wavelets* de um sinal com a análise no domínio do tempo apresentada por Shannon [52], no domínio da frequência (Fourier), e usando a transformada de Gabor. Percebe-se que a análise de *wavelets* não utiliza uma região tempo-frequência e sim uma região tempo-escala.

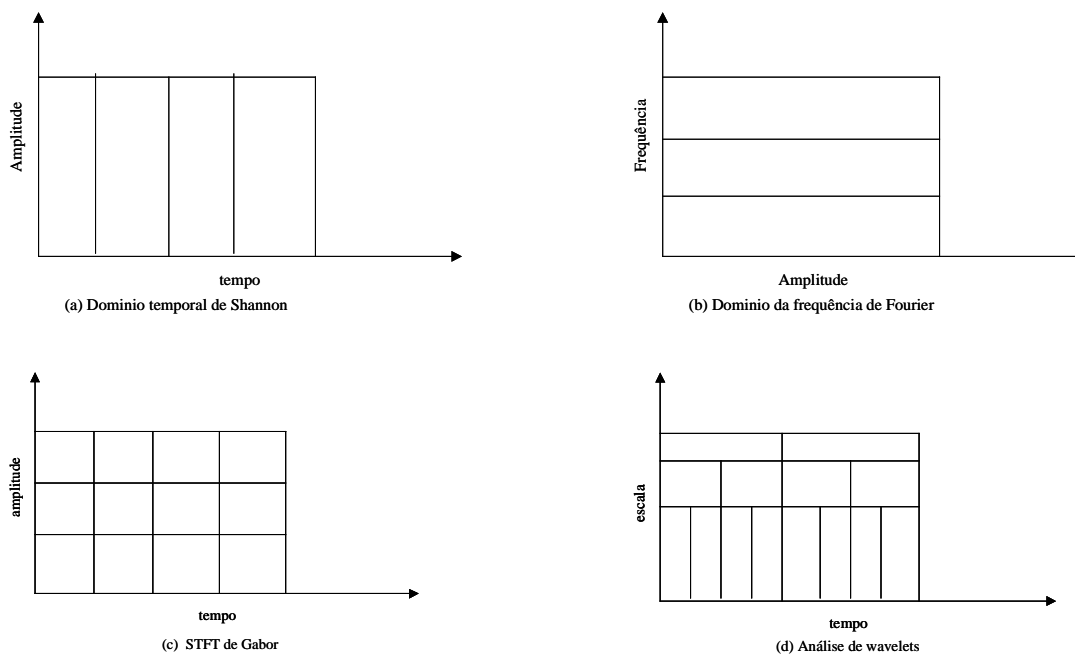


Figura D.3. Comparação das técnicas disponíveis em análise de sinais

D.3.1 O que é possível realizar com *wavelets*?

Uma das maiores vantagens que a análise de *wavelets* proporciona é a possibilidade de realizar análises locais num sinal de longa duração. Considere como exemplo um sinal senoidal com uma descontinuidade como mostrado na Figura D.4 .

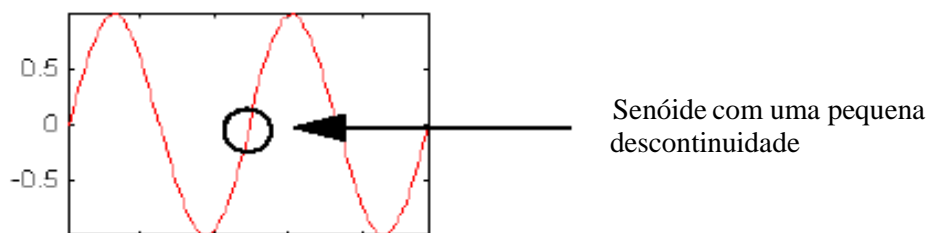


Figura D.4. Sinal senoidal com descontinuidade

Tal descontinuidade pode ter sido gerada por uma variação de potência ou pelo ruído causado por uma manobra num sistema elétrico. O gráfico de Fourier não mostra nada particularmente interessante: um comportamento *flat* com dois picos representando uma única frequência. Por outro lado, o gráfico dos coeficientes de *wavelets* mostra a exata localização temporal das descontinuidades. A análise de *wavelets* é capaz de

revelar características que outros tipos de análise ignoram, tais como: tendências, pontos de quebra, descontinuidades em derivadas superiores, auto similaridade.

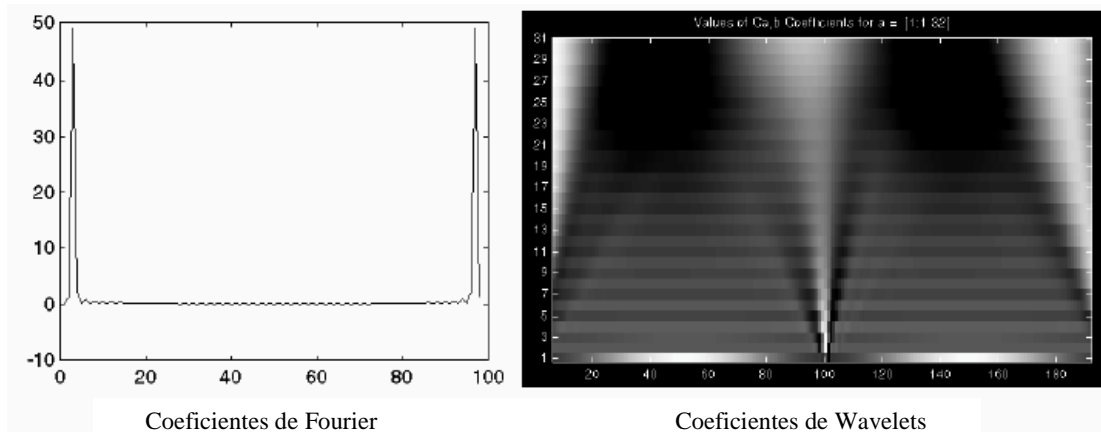


Figura D.5. (a) Análise de Fourier, (b) análise de *wavelets*

Além de oferecer um novo ponto de vista do comportamento de um sinal, *wavelets* é capaz de realizar a compressão de um sinal ou a eliminação do ruído de um sinal sem provocar sua excessiva degradação.

D.3.2 Definição de *wavelets*

A rigor, uma *wavelet* pode ser definida como uma forma de onda com duração limitada que tem uma média com valor zero. Ao comparar *wavelets* com senóides (base do análise de Fourier) observa-se que estas últimas são infinitas. Além disso, as funções seno são “comportadas” e previsíveis, já as *wavelets* tendem a ser irregulares e assimétricas .

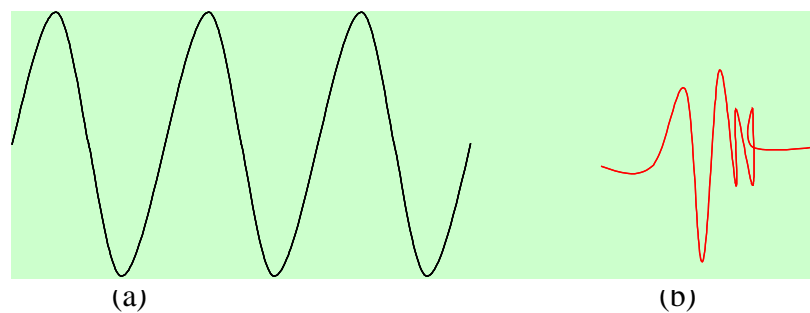


Figura D.6. (a) sinal senoidal, (b) *wavelet* de Daubechies (db10)

Como já anteriormente mencionado, a análise de Fourier consiste em desagregar um sinal em vários sinais senoidais de diferentes frequências. De forma similar, a análise de *wavelets* consiste na desagregação de um sinal temporal em versões defasadas e “escaladas” da *wavelet* original ou *wavelet* mãe. Por este motivo justifica-se

que sinais com mudanças bruscas sejam melhor analisadas usando uma onda irregular (*wavelet*) que uma suave senoidal [53].

D.4 A Transformada Contínua de *Wavelets* (CWT)

Matematicamente o processo da transformada de Fourier é representado pela seguinte expressão:

$$F(w) = \int_{-\infty}^{\infty} f(t) \cdot e^{-j\omega t} dt$$

(D.1)

que é o somatório de todos os intervalos de tempo do sinal, multiplicado pelo exponencial complexo. Ressaltando que o exponencial complexo pode ser separado numa componente senoidal imaginária e em outra real.

Os resultados da transformada são os coeficientes de Fourier $F(w)$, os quais quando multiplicado por uma frequência senoidal w , produzem os componentes senoidais do sinal original. Graficamente o processo é apresentado na Figura D.7

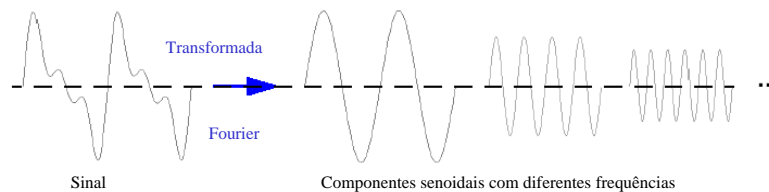


Figura D.7 Sinal temporal e seus componentes senoidais

Similarmente, a transformada de *wavelets* contínua (CWT) é definida como a soma no tempo do sinal, multiplicada pelas versões escaladas e defasadas da função *wavelets* ψ :

$$C(escala, posição) = \int_{-\infty}^{\infty} f(t) \cdot \psi(escala, posição, t) dt$$

(D.2)

O resultado de CWT são muitos coeficientes *wavelets* (C) em função da escala e posição. Multiplicando cada coeficiente pela apropriada *wavelets* escalada e defasada é produzida a *wavelets* componente do sinal original. A seguir na Figura D.8 uma representação gráfica do processo:

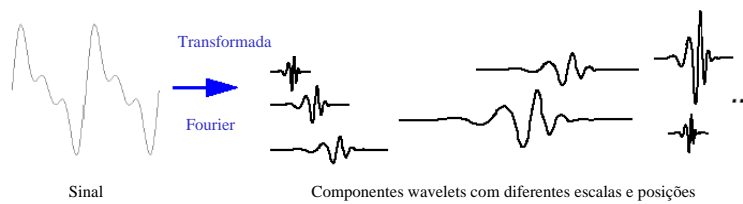


Figura D.8 Sinal temporal e seus componentes *wavelets*.

D.4.1 Escalamento

Temos aludido ao fato que a análise de *wavelets* fornece uma visão do sinal numa base escala-tempo, mas qual o significado de escala neste contexto?. Escalamento de uma *wavelet* significa simplesmente esticar ou comprimir uma *wavelet*. Para ir além de uma descrição coloquial como “esticamento” pode-se definir um *fator de escala* a . Se tomarmos como exemplo senóides, o fator de escala é muito fácil de ser observado (vide Figura D.9)

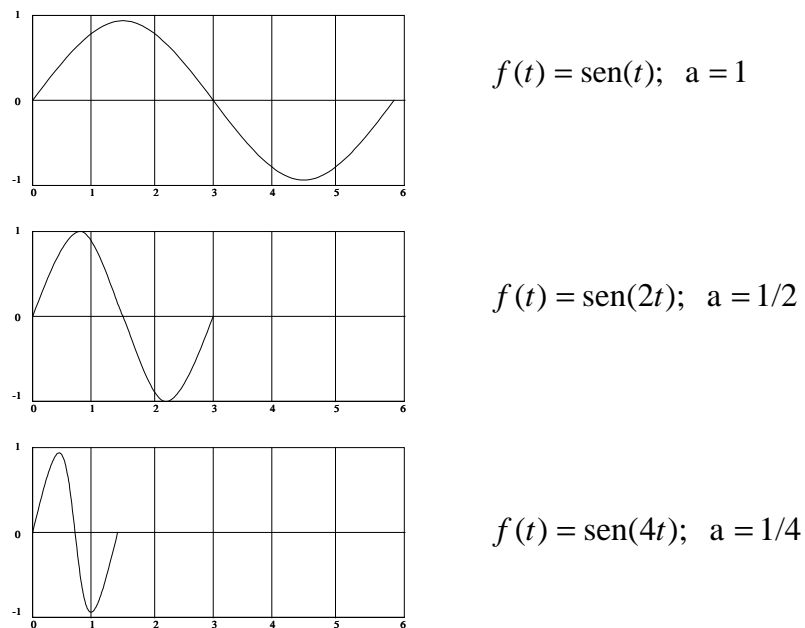


Figura D.9 Conceito de fator de escala.

O fator de escala trabalha da mesma forma em *wavelets*. Quanto menor o fator de escala, maior o grau de compressão da *wavelet* (vide Figura D.10).

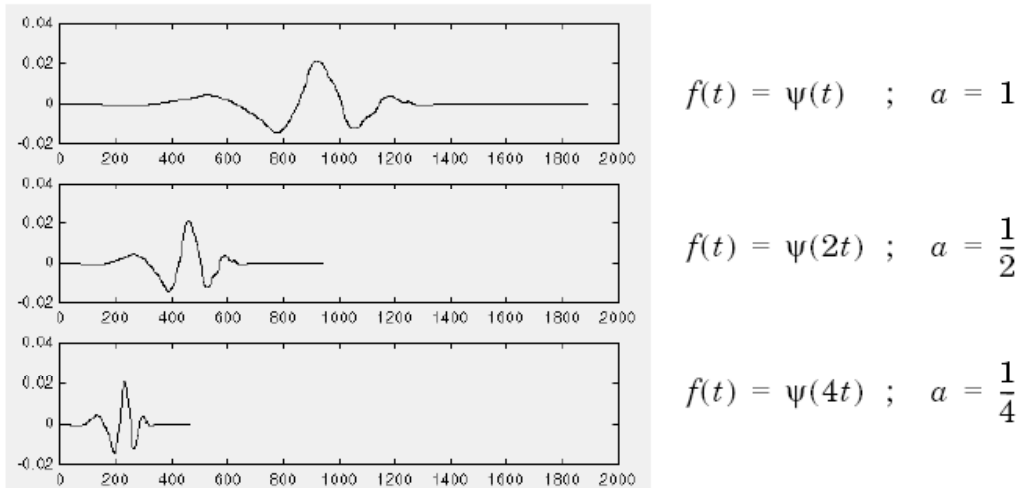


Figura D.10 Exemplo usando fator de escala.

Pode-se observar claramente na Figura D.10 que , para uma senóide $\text{sen}(wt)$, o *fator de escala a* é inversamente relacionado com a freqüência em radianos. Da mesma forma, na análise de *wavelet*, a escala é relacionada com freqüência do sinal.

D.4.2 Defasagem

Defasagem, na análise de *wavelets*, significa atrasar ou adiantar uma determinada *wavelet*. Matematicamente, atrasar uma função $f(t)$ em k pode ser representada por $f(t-k)$:

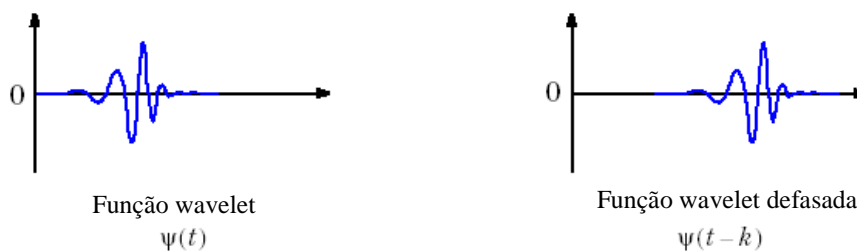


Figura D.11 Exemplo de defasagem numa *wavelet*

D.4.3 Procedimento para Transformação de *Wavelets* Contínua

A transformada contínua de *wavelet* é o somatório de todos os valores do sinal, multiplicado pelas versões escaladas e defasadas da *wavelet*. Este processo produz coeficientes *wavelets* em função da escala e posição.

Este processo é realmente muito simples, podendo ser realizado em cinco passos:

1. Escolha uma *wavelet* e compare sua secção com o início do sinal original;

2. calcule o número C , que representa o grau de correlação que existe entre a *wavelet* e a seção do sinal. Quanto maior o valor de C , maior o grau de similaridade. Mais precisamente, se a energia de C e a energia da *wavelet* é igual a uma unidade, C pode ser interpretado como um coeficiente de correlação. Obviamente os resultados desta etapa dependem da forma de onda da *wavelet* escolhida.

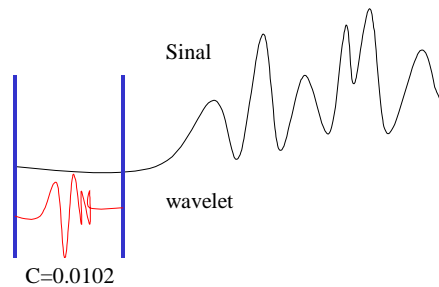


Figura D.12 Exemplo de cálculo de C numa *wavelet*

3. deslocar a *wavelet* para direita e repetir os passos 1 e 2 até alcançar o final do sinal;

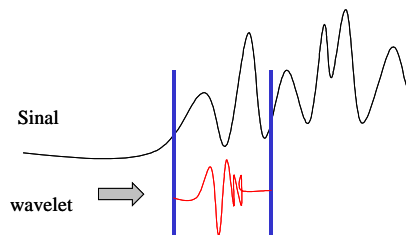


Figura D.13 Exemplo de cálculo de C numa *wavelet*. Deslocamento da *wavelet*

4. realizar o escalamento (esticamento) da wavelet e repetir os passos 1, 2 e 3

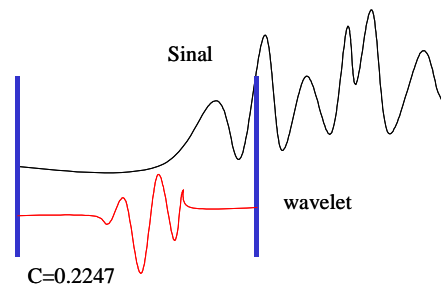


Figura D.14 Exemplo de cálculo de C numa *wavelet*. Esticamento da *wavelet*

5. repetir os passos 1 até 4 para todas as escalas

Realizado o processo, serão produzidos coeficientes a diferentes escalas usando diferentes secções do sinal. Os coeficientes constituem os resultados de uma regressão do sinal original baseado na utilização de uma *wavelet*.

D.4.4 Escala e Frequência

Pode-se observar na Figura D.15 que escalas superiores correspondem a *wavelets* mais esticadas. A análise de *wavelet* revela que existe uma correspondência entre escala de uma *wavelet* e frequência:

- Escala baixa a \rightarrow *wavelet* comprimida \rightarrow detalhes mudando rapidamente \rightarrow alta frequência w .
- Escala alta a \rightarrow *wavelet* esticada \rightarrow mudanças lentas \rightarrow baixa frequência w .

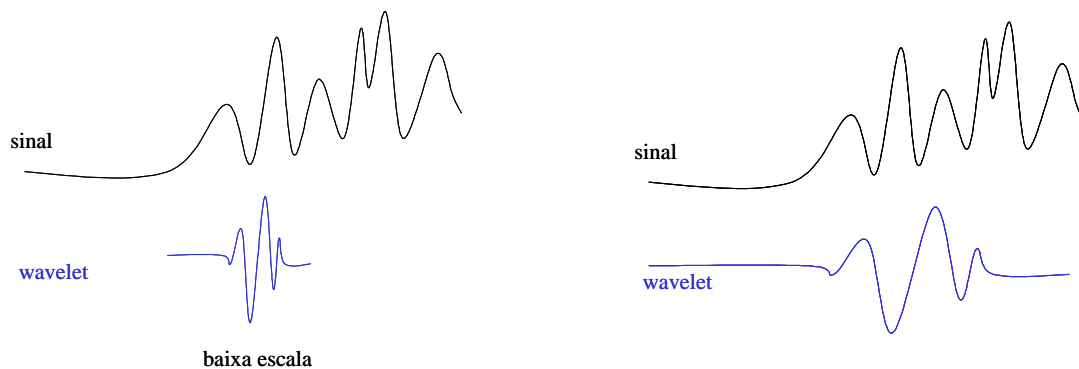


Figura D.15. Comparação de *wavelet* segundo a escala.

D.4.5 O que é contínuo na Transformada Contínua de Wavelet CWT ?

Qualquer processamento de um sinal real realizado por um computador é realizado usando dados discretos do sinal, isto é, dados obtidos em tempo discreto. Então, o que é exatamente “contínuo” no CWT?

As características “contínuas” do CWT, e que a distinguem da Transformada Discreta de *Wavelets* (DWT) são o conjunto de escalas e posições no qual opera.

Diferentemente da DWT, a CWT pode operar em diferentes escalas, que variam do sinal original até uma escala máxima definida pelo usuário em função da necessidade de detalhes e da disponibilidade computacional. A CWT é também contínua do ponto de vista da defasagem: durante o cálculo, a *wavelet* escolhida é deslocada sobre todo o domínio da função analisada.

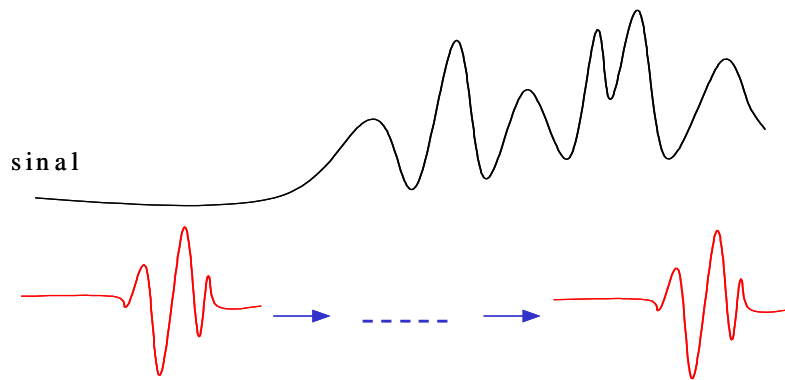


Figura D.16. processo de deslocamento de uma *wavelet*

D.5 A Transformada Discreta de *Wavelet* - DWT

Calcular os coeficientes de *wavelets* para cada uma das escalas possíveis é uma tarefa considerável que gera uma enorme quantidade de dados. O que acontece se ao invés de usar todas as possibilidades, fossem escolhidas somente algumas escalas e posições para realizar o cálculo?

Foi demonstrado em [54] que quando escolhidas escalas e posições baseados em potência de dois, conhecidas como *dyadic* a análise tende a ser muito mais eficiente e precisa. Esta análise é possível de ser realizado usando Transformada Discreta de *Wavelet* (DWT).

Uma eficiente forma de implementar esta metodologia foi desenvolvida em [55] por Mallat. O algoritmo de Mallat é conhecido na comunidade de processamento de sinais como “código de sub-bandas com dois canais”. Este prático algoritmo de filtragem produz uma transformada rápida de *wavelet*, que é uma caixa na qual um sinal entra e ao sair, emergem rapidamente os coeficientes de *wavelets*.

D.5.1 Filtragem de um Estágio: Aproximação e Detalhes

Para muitos sinais o conteúdo de baixa frequência é a parte mais importante, já que fornece a identidade do sinal. O conteúdo de alta frequência, por outro lado, fornece o matiz ou nuança do sinal. Considere como exemplo a voz humana, se eliminadas as componentes de alta frequência, o som da voz muda, mas ainda é possível compreender o que está sendo dito. No entanto, se removidas as componentes de baixa frequência já não é possível entender o que está sendo transmitido.

Na análise de *wavelet* são definidos dois conceitos novos: aproximação e detalhe. Aproximação corresponde a altas escalas ou componentes de baixa frequência

enquanto que detalhe está relacionado com baixas escalas ou componentes de alta frequência. Como pode ser observado na Figura D.17, o sinal original S passa através de dois filtros complementários e emerge como dois sinais.

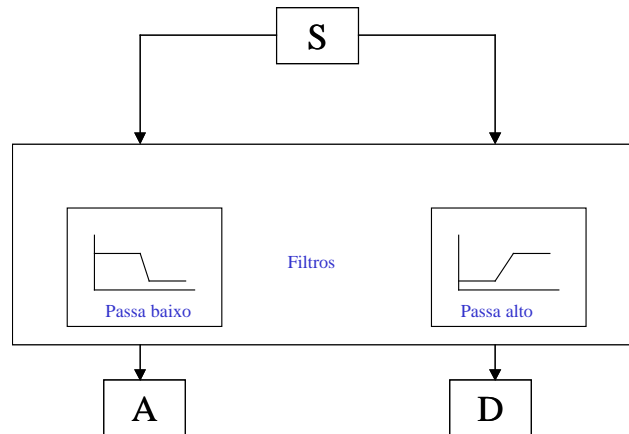


Figura D.17. Processo de filtragem de um sinal. Aproximação e detalhes.

Desafortunadamente, esta operação duplica a quantidade de dados. Assim, se no exemplo anterior o sinal S possuía 1000 dados no final do processamento terá 2000 dados.

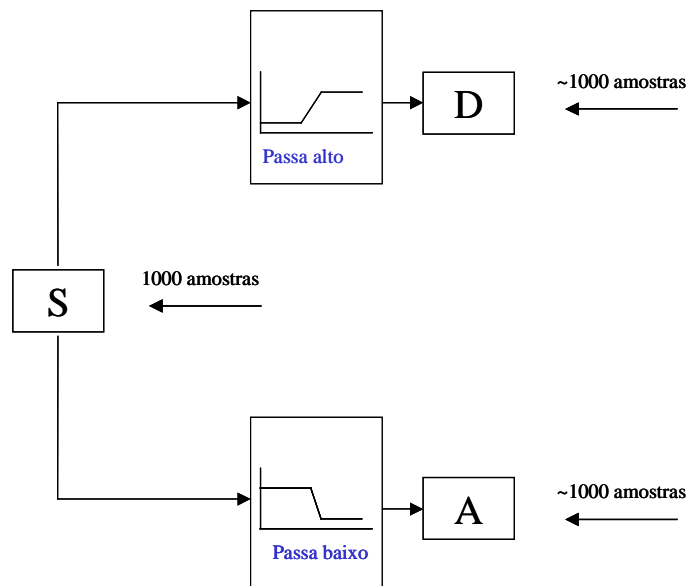


Figura D.18. Processo de filtragem de um sinal. Aproximação e detalhes.

No entanto, existe uma forma mais sutil de decomposição usando *wavelets*, que consiste em manter apenas um em cada dois dados da amostragem completa. Este conceito é conhecido como *downsampling* e produz duas seqüências de dados cA e cD .

onde A está relacionada com aproximação e D com detalhes. Este processo está representado na Figura D.19 e no final do mesmo é possível obter os coeficientes DWT.

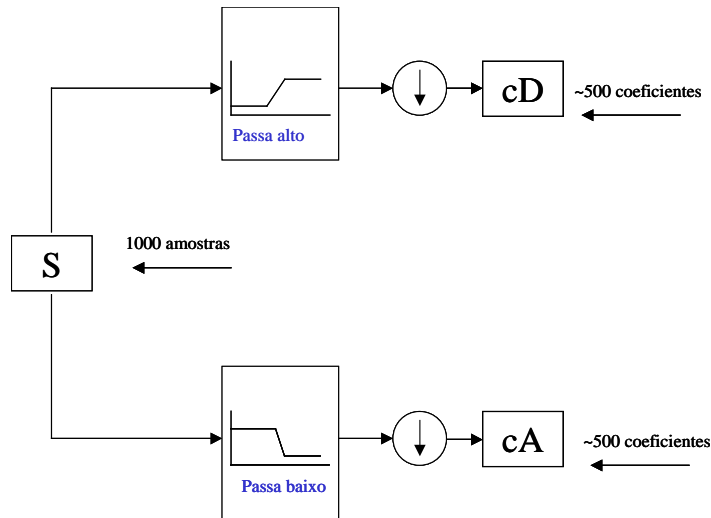


Figura D.19. Processo que produz os coeficientes DWT

Visando ganhar uma melhor compreensão do processo descrito acima, é apresentado na Figura D.20 um exemplo de um processo de DWT de um estágio usando um sinal senoidal ao qual foi adicionado um ruído de alta frequência

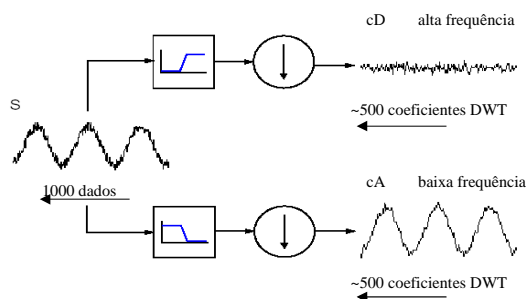


Figura D.20. Processo que produz os coeficientes DWT: exemplo usando sinal senoidal

É possível observar que os coeficientes Detalhe cD são pequenos e consistem principalmente de um ruído de alta frequência, enquanto que os coeficientes Aproximação cA contém uma quantidade de ruído muito inferior ao sinal original. Na prática tem sido observado que o comprimento dos vetores coeficientes (Aproximação e Detalhes) tem uma quantidade de dados levemente maior que a metade de dados do sinal original. Este fato é atribuído ao processo de filtragem, o qual é implementado

aplicando convolução usando um filtro. A convolução “mancha” o sinal original, introduzindo várias amostras extras no resultado.

D.5.2 Decomposição de vários níveis

O processo de decomposição pode ser iterativo, com aproximações sucessivas tal como apresentado na Figura D.21, onde o sinal é aberto em vários componentes com resolução inferior. Este processo é conhecido como árvore de decomposição de *wavelet*.

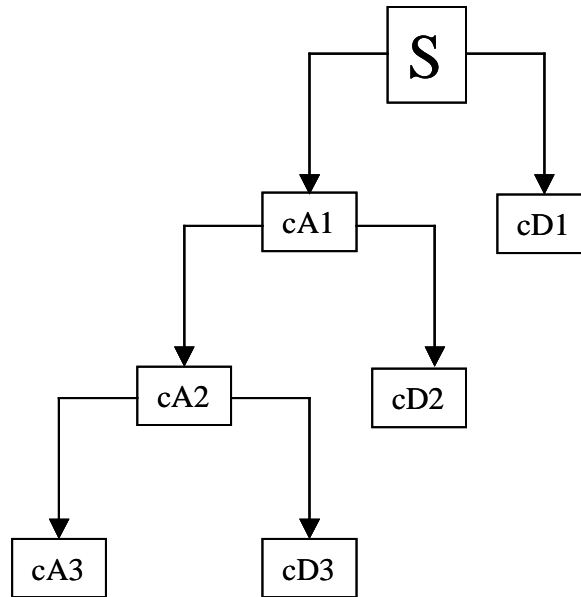


Figura D.21. Árvore de decomposição de *wavelet*

D.5.3 Reconstrução de *wavelets*

Até agora foi mostrado como utilizar a DWT para analisar ou decompor sinais. Este processo é denominado decomposição ou análise. A outra parte da história relaciona-se com a forma de recomposição do sinal original, sem perda de informação, a partir destes componentes. Este processo é conhecido como reconstrução, sendo que a manipulação matemática do processo é conhecida como Inversa da Transformada Discreta de *Wavelet* (IDWT). Na Figura D.22 é apresentado um diagrama do processo de reconstrução a partir dos coeficientes *wavelets*.

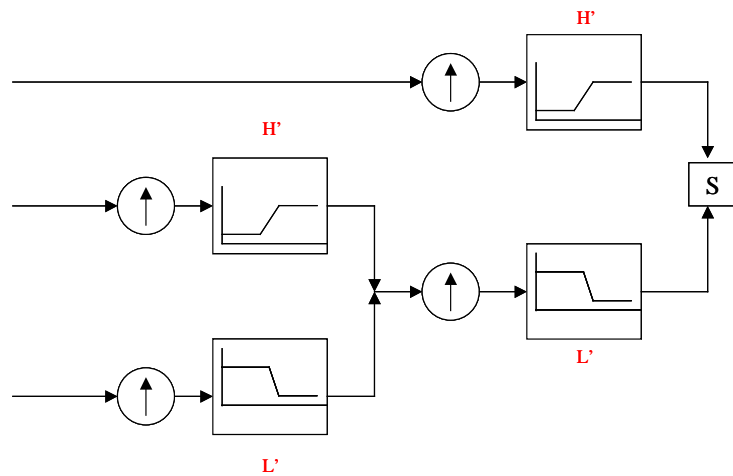


Figura D.22. Reconstrução a partir dos coeficientes *wavelets*

Da mesma forma que o processo de decomposição de *wavelet* envolve filtragem e *downsampling*, o processo de reconstrução segue um caminho contrário que envolve além da filtragem um processo conhecido como *upsampling*. Como pode ser observado na Figura D.23, o processo de *upsampling* consiste em incrementar os componentes de um sinal inserindo zeros entre cada valor amostral.

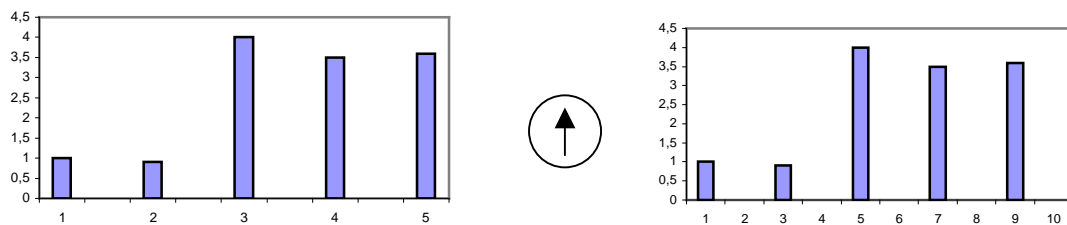


Figura D.23. Processo de *upsampling*

D.5.4 Reconstrução de Filtros

A etapa de filtragem no processo de reconstrução merece alguma discussão devido principalmente à escolha do filtro, elemento crucial para alcançar uma perfeita reconstrução do sinal original. O processo de *downsampling* dos componentes do sinal realizado durante a etapa de decomposição introduz uma distorção conhecida como *aliasing*. Uma correta escolha dos filtros nas etapas de decomposição e reconstrução pode cancelar o efeito citado. A decomposição com filtros passa alto e passa baixos (H e L) em conjunto com os filtros associados ao processo de reconstrução (H' e L')

mostrados na Figura D.24, formam um sistema conhecido como espelho de quadratura de filtros.

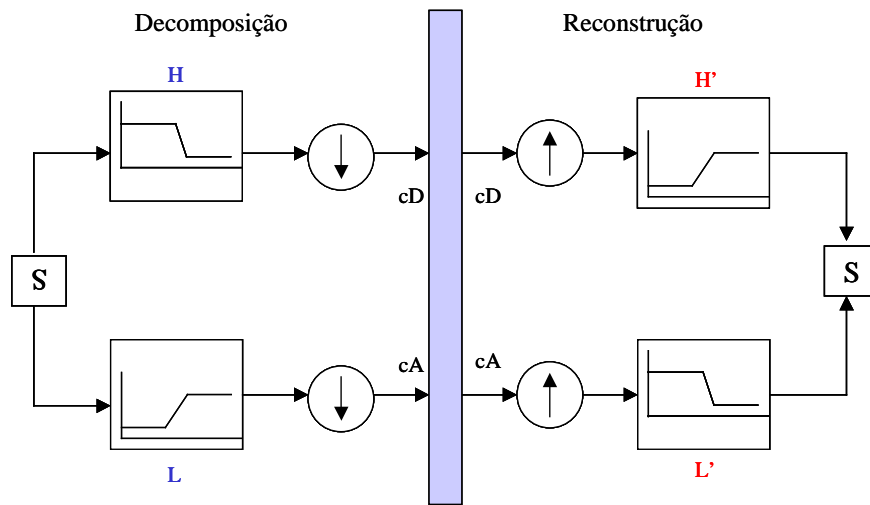


Figura D.24. Espelho de quadratura de filtros

D.5.5 Reconstrução de Aproximação e Detalhes

Nas seções anteriores demonstrou-se ser possível reconstruir o sinal original a partir dos coeficientes de aproximação e detalhes, como também foi visto que é possível reconstruir as aproximações e detalhes a partir de seus vetores coeficientes.

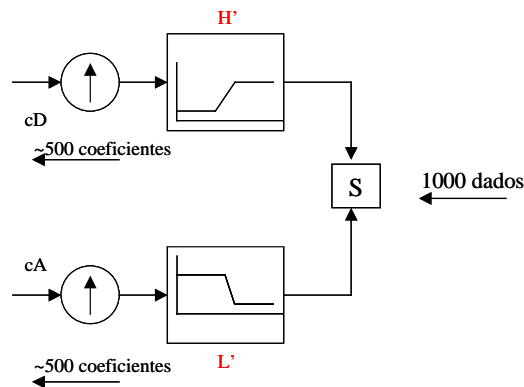


Figura D.25. Reconstrução do sinal original

Como exemplo considere a reconstrução do primeiro nível de aproximação A1 a partir do vetor de coeficientes $cA1$. Será realizado o mesmo processo usado para reconstrução do sinal original, apresentado na Figura D.25, no entanto ao invés de combinar com o primeiro nível de detalhes $cD1$, é combinado com um vetor de zeros, como apresentado na Figura D.26.

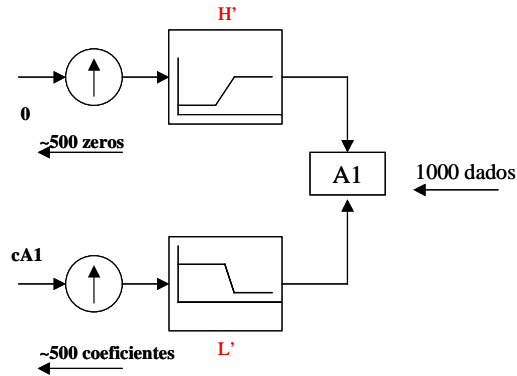


Figura D.26. Reconstrução do primeiro nível de aproximação A1

Similarmente pode ser reconstruído o primeiro nível de detalhes D1 utilizando-se processo análogo tal como apresentado na Figura D.27

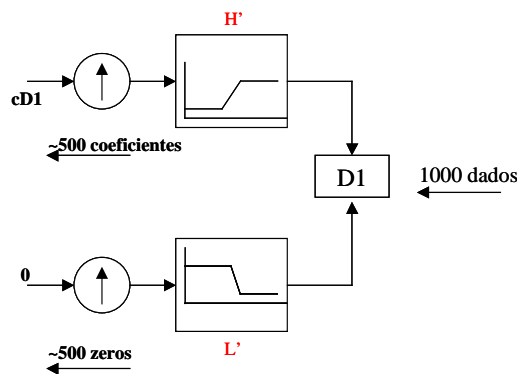


Figura D.27. Reconstrução do primeiro nível de detalhes D1

Os detalhes e aproximações reconstruídos são componentes verdadeiros do sinal original. De fato, ao serem combinadas obtém-se a seguinte equação: $A1+D1=S$. Observa-se que os vetores coeficientes $cA1$ e $cD1$, por terem sido produzidos por *downsampling* tem somente a metade do comprimento do sinal original, não sendo possível combiná-los eles diretamente para obter o sinal original. Desta forma, é necessário reconstruir as aproximações e detalhes antes de reconstruir o sinal.

Expandindo esta técnica aos componentes numa análise de vários níveis se encontram relações similares para todos os componentes do sinal reconstruído. Por este motivo pode-se afirmar que existem, tal como representado Figura D.28.

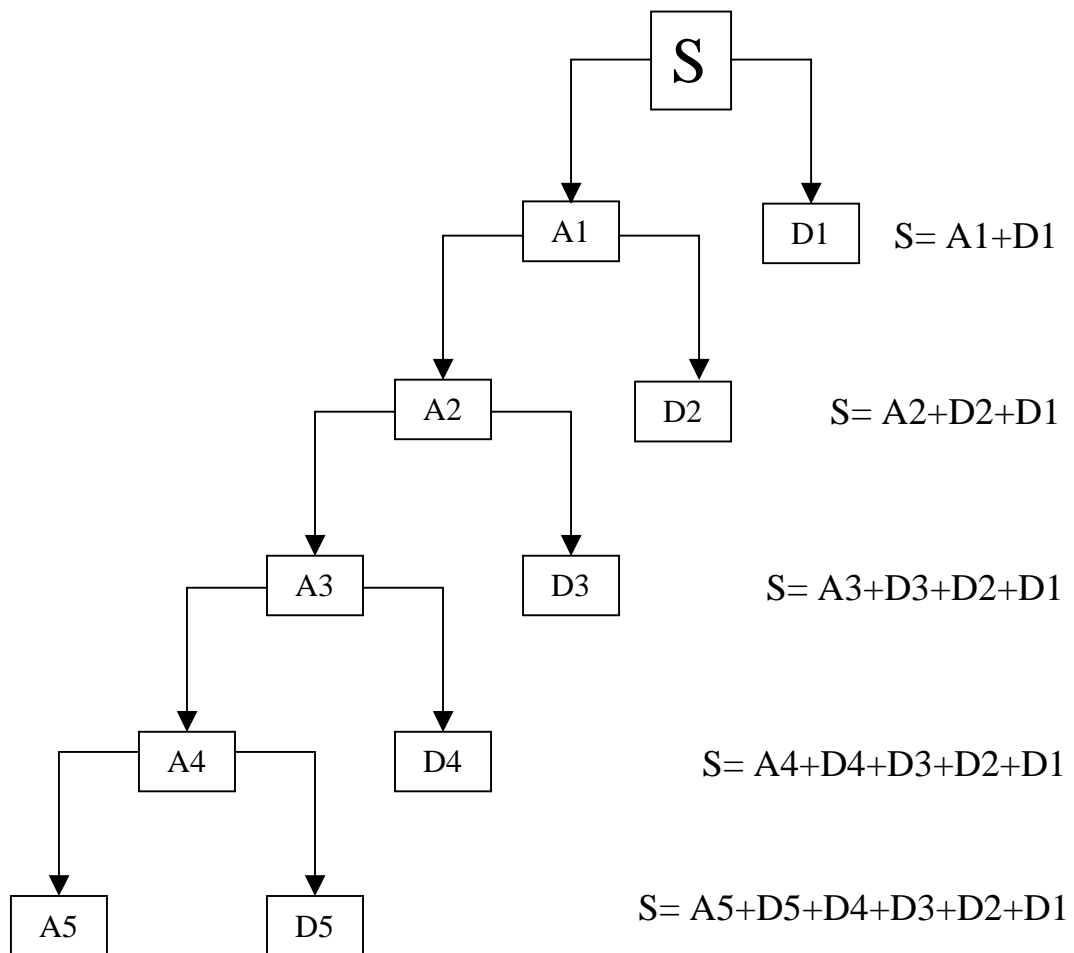


Figura D.28. Componentes do sinal reconstruído

BIBLIOGRAFIA

- [1] HIPPERT H.S, PEDREIRA A.C, SOUZA R.C. “Neural Networks for Short-Term Load Forecasting: A Review and Evaluation”. *IEEE Transactions on Power Systems*, Vol 16 No.1, February 2001.
- [2] LOTUFO A.D.P, MINUSSI C.R. “Electric Power Systems Load Forecasting: A Survey”.In: *IEEE Power Tech 99 Conference*, Budapest, Hungary, 1999.
- [3] ISLAM S.M, AL-ALAWI S.M, “Principles of Eletricity Demand Forecastig, Part 2 *Power Engineering Journal* April 1997.
- [4] MEDEIROS M.C, TERASVIRTA T, RECH G, “Building Neural Networ Models for Time series: A Statistical Approach”. November 14, 2004
- [5] HAYKIN S, *Redes neurais: princípios e prática*.2^a Edição, Editora Bookman, 2001.
- [6] HOCKING R, “The analysis and selection of variables in linear regression”, *Biometrics*, vol.32, pp 1-49, Mar, 1976
- [7] GOUTTE C. “Extraction of the Relevant Delays for Temporal Modeling”. *IEEE Transactions on Signal Processing*, Vol 48 No.6, June 2000.
- [8] NARENDRA P.M, FUKUNAGA K. “ A branch and bound algorithm for feature subset selection”. *IEEE Transactions Comput.*, Vol 26, pp 917-922, Sept. 1977.
- [9] LERAY P, GALLINARI P, “Feature selection with neural networks”, Tech.Rep.LIP6 1998/012, 1998.
- [10] BATITTI R, “Using Mutual Information for Selecting Features in Supervised Neural Net Learning”. *IEEE transactions on neural networks*, vol 5, No.4, july 1994.
- [11] GOUTTE C. “Lags space estimation in times series modeling”. In Proc. *ICASSP*, 1997, pp. 3313-3316.
- [12] DREZGA I, RAHMAN S, “Input Variable Selection for ANN-Based Short-Term Load Forecasting”. *IEEE Transaction Power Systems*, vol 13, n^o 4, pp.1238-1244, 1998.

- [13] GUJARATI D.N, *Econometria Básica*. 3ª Edição. Editora Makron Books. 2000.
- [14] JOLLIFFE I.T. *Principal Component Analysis*. Springer, New York. 1986.
- [15] ALVES DA SILVA A, ROCHA REIS A, “Seleção de variáveis de entrada para previsores neurais-Um estudo comparativo”. *In: IX Sepope*, Rio de Janeiro, maio de 2004.
- [16] RECH G, TERASVIRTA T, TSCHERNIG R, “A Simple Variable Selection Technique for Nonlinear Models”. *Communications in Statistics, Theory and Methods* , vol 30, pp 1227-1241.2001.
- [17] SAVI M.A, “Caos em sistema mecânicos”. *In: Anais do I congresso de dinâmica, controle e aplicações* vol 1. pág 1-27.Sociedade Brasileira de Matemática Aplicada e Computacional- SBMAC. 2002.
- [18] TAKENS F, “Detecting strange attractors in turbulence, in Dynamical systems and turbulence”, *Lectures Notes in mathematics*, vol 898, pp 366-381.1981.
- [19] SAVIT R, GREEN M, “Time series and dependent variables”, *Physica D*, vol.50, pp. 95-116, 1991.
- [20] Pi H, Peterson C, “Finding the embedding dimension and variable dependences in time series”, *Neural Comput.* , vol 6, no 3, pp, 509-520, May 1994.
- [21] He X, Asada H, “A new method for identifying orders of input-output models for non linear dynamic systems, ” in *Proc.Amer.Conf.Contr.*, San Francisco, CA, 1993.
- [22] MOLINA C, SAMPSON N, FITZGERALD W.J et. al. “Geometrical techniques for finding the embedding dimension of time series”, . In *Proc. IEEE Workshop Neural Networks Signal Process. VI*, 1996, pp. 161-169.
- [23] FRASER A.M, SWINNEY H.L, “Independent coordinates for strange attractors from mutual information”. *Phys Review A*, vol 33, pp 1134.1986.
- [24] KENNEL M.B, BROWN R, ABARBANEL H.D.I, “Determining embedding dimension for phase space reconstruction using a geometrical construction”. *Phys Review A*, vol 45, pp 3403. 1992.
- [25] GRASSBERGER P, PROCACCIA I, “Measuring the strangeness of strange attractors”. *Physica D*, 8,189. 1983.

- [26] ALVES MONTEIRO L.H, *Sistemas Dinâmicos*. 3ª Edição, Editora Livraria da Física. 2002.
- [27] LORENZ E.N, “Deterministic Nonperiodic Flow”, *Journal of Atmospheric Science*, v.20, pp 130-141. 1963.
- [28] GLEICK.*Caos*. Editora Campus. 1987.
- [29] MAY R.M, “Simple Mathematical Models with Very Complicated Dynamics”, *Nature*,v.261, pp 459-267. 1976.
- [30] RUELLE D, TAKENS F, “On the nature of turbulence”,*Commun. Math. Phys.*20, 167 (1971).
- [31] GOLLUP J.P, BENSON S.V, “Many Routes to Turbulence Convection”, *J.Fluid Mech.* 100, pp 449-470. 1980.
- [32] ABARBANEL H.D.I, *Analysis of Observed Chaotic Data*, Ed. Wiley, New York.1996.
- [33] WOLF A, SWIFT J.B, SWINNEY H.L, et.al, “Determining Lyapunov Exponents from a Times Series”, *Physica 16D*, pp. 285-317. 1985
- [34] PARKER T.S, CHUA L.O. *Practical Numerical Algorithms for Chaotics Systems*, Springer, New York.1989.
- [35] KOLMOGOROV A.N, “A new invariant of transitive dynamical systems. Dokl. Akad. Nauk. SSSR, 119, 861.
- [36] BELATONI, RENYI ?
- [37] PACKARD N.H.,CRUTCHFIELD J.P, FARMER J.D, and SHAW R.S, “Geometry from a time series”, *Physics.Review.Letters* ,vol 4, pp 712-716, 1980.
- [38] MALDENBROT B,“Comments on computer rendering of fractal stochastic models”.*Communications of the Association for Computing Machinery*: 25, pp. 581-584.1982.
- [39] SHANNON C.E, “A mathematical theory of communication”, *Bell System Technical Journal*, vol 27, pp 379-423, 623-656. 1948.
- [40] FRASER A.M, SWINNEY H.L, “Independent coordinates for strange attractors from mutual information”. *Phys Review A*, vol 33, pp 1134.1986.
- [41] LINSKER R, “How to generate ordered maps by maximizing the mutual information between input and output signals”, *Neural Computations* vol 1, No3, 1989.
- [42] VRA-Visual Recurrence Analysis. Vdemo4.7.

- [43] HEGGER R, KANTZ H, SCHREIBER T, “Practical implementation of nonlinear time series methods: The TISEAN package”. *Chaos*, 9, 413. The TISEAN routines can be downloaded from: <http://www.mpiiks-dresden.mpg.de/~tisean>.
- [44] KANTZ H, SCHREIBER T, “Nonlinear Time series analysis”, 2 ed.,Cambridge university press, cambridge, 1990.
- [45] AITOKALLIO T, GYLLEMBERG M, et al. “Improving the false nearest neighbors method with graphical analysis” ArXiv: chao-dyn/9812019 v1 13 december 1998.
- [46] KENNEL M.B, BUHL M. “Estimating good discrete partitions from observed data: symbolic false nearest neighbors”. ArXiv:nlin.CD/0304054 v1 26 Apr 2003.
- [47] CARVAJAL R, VALLVERDÚ M, et al. “Detección de determinismo no lineal en la variabilidad de la frecuencia cardiaca”. *Memorias II Congreso Latinoamericano de Ingenieria Biomédica, Habana 2001*. La Habana, Cuba, mayo 23-25 de 2001.
- [48] EGUILUZ L.I, MAÑANA M. “Clasificación de perturbaciones mediante la representación en el espacio fase”. *Memorias II Congreso Latinoamericano de Ingenieria Biomédica, Habana 2001*. La Habana, Cuba, mayo 23-25 de 2001.
- [49] Correlation Dimension Calculator (version 1.0). Peter A. Noble. <http://stahl.ce.Washington.edu/readmes/CorrelDimReadme.htm>
- [50] MITISI M, MITISI Y, OPENHEIM G, POGGI G, “Wavelet Toolbox for Use with Matlab-Users guide version 2”, The Math Works Inc, MA, 2000.
- [51] GABOR D, “Theory of Communication,” *Journ. Inst. Elect. Eng.*, 93(3): 891-906, September 1946.
- [52] SHANNON C.E, WEAVER W, “The Mathematical Theory of Communication”, The University of Illinois Press, Urbana, IL, 1964.
- [53] HUBBARD B.B, “The World According to Wavelets: The Story of a Mathematical Technique in the Making”, second edition, AK Peters, MA, 1998.
- [54] MALLAT S, “A Compact Multiresolution Representation: The Wavelet Model”, In Proceedings of the IEEE Computer Society Workshop on Computer Vision, pp.2-7, IEEE Computer Society Press, Washington, D.C., 1987.
- [55] MALLAT S, “Multiresolution Aproximation and Wavelets”, *Trans. Amer. Math.Soc.*, 315: 69-88, 1989.

- [56] PIERCE D.A, “Seasonal adjustment when both deterministic and stochastic seasonality are present”, *In Seasonal Analysis of Economic Time Series*. Arnold Zellner Ed. 242-269. Washington, D.C., U.S. Dept. of Commerce, Bureau of the Census. 1979.
- [57] DRAPER N.R, “Applied Regression Analysis”. Third Edition. New York: Wiley. 1998.
- [58] ANDERSON T.W, “The Statistical Analysis of Time Series”. Wiley, New York 1971.
- [59] PIERCE D.A, “A survey of recent developments in seasonal adjustment”, *The American Statistician*. 34, 125-134. 1980.
- [60] MORETTIN P, TOLOI C, *Análise de Séries Temporais*. Editora Edgard Blücher LTDA. São Paulo. 2004.
- [61] DURBIN J, “Trend elimination by moving-average and variate-difference filters”, *Bulletin of the International Statistical Institute*. 139, 130-141. 1962.
- [62] MISITI M, MISITI Y, OPPENHEIM G, POGGI J.M, “Ondelettes en statistique et traitement du signal”, *Revue de Statistique Appliquée*. Vol. XLI, no 4, pp.33-43.
- [63] STRANG G, NGUYEN T, “Wavelets and Filter Banks”. Wellesley, Cambridge Press 1996.
- [64] CHEN B, CHANG M, LIN C. “Load Forecasting Using Support Vector Machines: A Study on EUNITE Competition 2001”. *IEEE Transactions on Power Systems*, Vol 19 No.4, November 2004.