Predição Não-Linear de Séries Temporais Usando Redes Neurais RBF por Decomposição em Componentes Principais

por
Maria Cristina Felippetto de Castro

Tese de Doutorado apresentada à
Universidade Estadual de Campinas - UNICAMP,
Faculdade de Engenharia Elétrica e de Computação,
como requisito parcial para a obtenção do grau de
Doutor em Engenharia Elétrica.

Orientador:
Prof. Dr. Dalton Soares Arantes

Banca Examinadora:
Prof. Dr. Dario Francisco Guimarães Azevedo
Prof. Dr. João Marcos Travassos Romano
Prof. Dr. Max Machado Costa
Prof. Dr. Paulo Roberto Girardello Franco

Campinas, São Paulo
Março de 2001

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

C279p

Castro, Maria Cristina Felippetto de
Predição não-linear de séries temporais por redes neurais utilizando decomposição em componentes principais / Maria Cristina Felippetto de Castro. -- Campinas, SP: [s.n.], 2001.

Orientador: Dalton Soares Arantes.
Tese (doutorado) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Teoria da previsão. 2. Análise de séries temporais. 3. Redes neurais (Computação). 4. Telecomunicações. 5. Análise de componentes principais. I. Arantes, Dalton Soares. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.

Resumo

Esta tese apresenta uma nova técnica de predição não-linear de séries temporais através de redes neurais artificiais do tipo *Radial Basis Function*, com atribuição dos centros Gaussianos das funções de base radial por decomposição do espaço de dados em sub-espaços. A decomposição em sub-espaços – ou decomposição em componentes principais – é baseada na Transformada Karhunen-Loève. A predição obtida através da parametrização da rede neural via decomposição em sub-espaços resulta em um menor erro de predição e requer o conhecimento de um menor número de amostras prévias do que as técnicas de predição convencionais. Adicionalmente é apresentada uma possível solução para o problema de adaptar dinamicamente a arquitetura da rede neural às não-estacionariedades presentes em muitas séries temporais.

Abstract

This thesis proposes a new technique for non-linear time series forecasting based upon Radial Basis Function Neural Networks and the Karhunen-Loève Transform. A significant performance improvement is obtained with the novel technique in comparison with usual prediction methods. By obtaining the neural network centers from the data set sub-spaces – or data set principal components – the new method yields lower prediction error and requires less previous known samples than the usual technique that applies the own training set vectors to the centers. Additionally we present a possible solution to the problem of dynamically adapting the neural network architecture to the time-varying series statistics.

Agradecimentos

A meu orientador, Professor Dalton Soares Arantes, por seu apoio e por sua orientação segura, tranqüila e paciente. A sua encantadora família, em especial a sua esposa Auxiliadora, pela acolhida carinhosa que nos dedicou.

Aos demais professores, em especial aos que generosamente nos acolheram como ouvintes.

A meus queridos familiares, em especial minha mãe Carminha, minha sobrinha Paula, meu sogro Ney e minha sogra Rosalina, que aceitaram de forma altruísta nossa ausência nestes últimos anos, tão difíceis em suas vidas.

Aos amigos do sul que nos visitaram, diminuindo a saudade de casa. Aos amigos do sul que não puderam vir, mas permaneceram presentes.

A Fernando, que tem sido meu mais querido colega desde a graduação, por seu inestimável auxílio e por permanecer norteando sua vida pelos ideais e princípios que nos aproximaram vinte anos atrás, permitindo que estes sejam, ainda, nossos ideais comuns.

Não posso deixar de dizer, enfim, o quanto considero difícil nominar apenas algumas pessoas a quem gostaria de agradecer. Acredito que, se pararmos para refletir e quisermos ser justos, teremos que agradecer a todos que cruzam nosso caminho, pois há um propósito maior em tudo, mesmo que, de imediato, não o reconheçamos. Veio em meu auxílio, no entanto, um provérbio que encontrei em uma das teses lidas recentemente e que expressa com a máxima fidelidade meu sentimento:

When one is blessed with too many people to thank, one should thank God.

Cristina.

Índice

Lista de Figuras

Lista de Tabelas

Lista de Acrônimos

APC	Atribuição Padrão de Centros
DSE	Decomposição em Sub-Espaços
FIR	*Finite Impulse Response*
JPS	Janela de Predição Seletiva
KLT	*Karhunen-Loève Transform*
LAN	*Local Area Network*
LMS	*Least Mean Square*
LSE	*Least Square Error*
MF	Modelo Fixo
MLP	*Multilayer Perceptron*
MPEG	*Moving Pictures Expert Group*
MSE	*Mean Square Error*
NMSE	*Normalized Mean Square Error*
NMSEA	*Normalized Mean Square Error* de Aproximação
PCA	*Principal Components Analysis*
RBF	*Radial Basis Function*
SVD	*Singular Value Decomposition*
VLSI	*Very Large Scale Integration*
WSS	*Wide Sense Stationary*

Capítulo 1

Introdução

" ... Gurnemanz estava sentado na cadeira real, olhando para o cálice no centro da mesa. A delicada irradiação que emanava do cálice atuava sobre ele como bálsamo, devolvendo-lhe, ao mesmo tempo, sua tranqüilidade inabalável. A tristeza que sentira nos últimos dias havia-o oprimido pesadamente ... Olhando para os druídas que se haviam sentado na grande mesa redonda, a falta dos cinco, de repente era-lhe indiferente, pois lembrou-se de como eram, no fundo, insignificantes e secundários os seres humanos ... Tinham, aliás, seu destino nas próprias mãos, podiam organizar sua vida como bem entendessem, escolhendo a direção de seus caminhos, não obstante não eram livres, pois sua vida futura estava ligada às decisões tomadas por eles no presente.

- Nós todos juramos, outrora, fidelidade ao nosso Senhor e Rei de todos os espíritos, Parsefal! Começou Gurnemanz. Através desse juramento estamos ligados com a fonte de Luz que irradia dele. Sabeis que finas irradiações partem de cima até nós, embaixo. Podemos chamá-las, também, de caminhos. Todas as revelações, até agora, puderam vir por esses caminhos, através de espíritos superiores, até nós. Também a profecia chegou por esses caminhos e agora nos foi dado conhecer a data em que o cataclismo do nosso país se tornará realidade! Nossa melhor proteção jaz no saber do que virá!... "

(Gurnemanz, sábio mentor espiritual do reino de Atlântida, reunido com os druídas no salão real - extraído de "Atlântida! Princípio e Fim da Grande Tragédia", de Roselis von Sass.)

Apesar de distantes dos sábios druídas, em muitas situações ainda reconhecemos como verdade que "nossa melhor proteção jaz no saber do que virá". Imensamente menos afortunados que os atlantes, para nos beneficiarmos do saber do futuro, precisamos contar com nossa capacidade lógica para inferir e extrapolar modelos a partir da observação, no passado e no presente, dos diversos processos de nosso interesse (pelo menos até que nossa intuição nos permita voltar a ter acesso a métodos hoje esquecidos, como aqueles adotados com naturalidade pelos sábios druídas...).

A predição de séries temporais é um estudo de extrema relevância, já que exemplos de séries temporais são abundantemente encontrados na natureza (em campos tais como geofísica, astrofísica e meteorologia), nas ciências sociais (em campos como a demografia), nas ciências médicas (em estudos de processos fisiológicos involuntários), nas ciências econômicas (no acompanhamento das taxas de câmbio de moedas e mercado de ações) e nas diversas engenharias (em tratamento e transmissão de sinais, sistemas dinâmicos, etc.), entre muitos outros.

Dentro de um contexto histórico, desde o começo deste século são conhecidos métodos matemáticos empregados para a predição de séries temporais (devo referir-me aqui ao século XX, pois encontro-me na rara situação de estar escrevendo esta tese no final do século XX, enquanto que o leitor estará, provavelmente, lendo-a já no século XXI).

Até 1920, a predição de séries temporais era feita simplesmente extrapolando os dados através de um ajuste global no domínio tempo. O começo da moderna predição de séries é atribuído a 1926, quando a técnica auto-regressiva foi proposta por Yule [3][54]. O intento de Yule era prever o número anual de manchas solares. A técnica de predição usada por Yule consistia em determinar o valor a ser predito através de uma soma ponderada das observações prévias da série. Uma operação puramente linear.

Nos cinqüenta anos seguintes tais modelos lineares, acrescidos de ruído, foram objeto de pesquisas. Ao redor de 1980, com o advento dos grandes computadores e do estudo de redes neurais artificiais, puderam ser estudadas séries mais longas e aplicados a estas, um universo de modelos mais complexos. Assim, com o auxílio das técnicas de redes

neurais artificiais foi possível explorar, de forma adaptativa, grandes conjuntos de dados. E, a partir das relações extraídas dos dados, inferir o processo gerador [3][7].

As técnicas anteriores consistiam em procurar dentro de um universo limitado de modelos, aqueles que melhor representassem os processos geradores das séries (uma boa análise estatística requer assumir uma certa forma para os dados e testar sua validade, repetidas vezes, até que a forma correta seja encontrada – uma tarefa custosa e muitas vezes nem mesmo realizável).

As redes neurais artificiais têm a capacidade de aprender padrões subjacentes presentes nos conjuntos de dados, apresentando melhor desempenho que os métodos estatísticos tradicionais quando o processo regente dos dados é desconhecido, não-linear e/ou não-estacionário – como o são a maior parte dos processos encontrados no mundo real. Por isso, representam uma grande contribuição ao estudo das séries temporais resultantes de tais processos [51].

A análise de séries temporais sempre foi uma pesquisa multi-disciplinar. Apesar de partilharem da inspiração científica de construir modelos a partir de observações e, não obstante os avanços alcançados após 1980, os pesquisadores das diversas áreas sempre encontraram dificuldade em sintetizar os progressos obtidos e até mesmo em reunir uma bibliografia comum a todos [3].

Em 1990, um novo impulso foi dado à pesquisa no âmbito da predição de séries temporais, através de uma estranha idéia. Neil Gershenfeld (pesquisador do *MIT Media Laboratory*) e Andrea Weigend (pesquisador da *University of Colorado* e da *Xerox PARC*) naquela ocasião cursavam juntos o programa *Complex Systems Summer School* do *Santa Fe Institute*[1.1] [3]. As pesquisas de Gershenfeld e Weigend envolviam temas que suscitavam a análise de séries temporais. Os dois pesquisadores esbarraram em grande dificuldade ao procurar literatura consistente, na qual fossem compiladas as técnicas utilizadas nas diversas áreas do conhecimento envolvidas na análise de séries temporais. E, mesmo dentre

[1.1] O *Santa Fe Institute* é uma instituição de ensino e pesquisa multi-disciplinar, a nível de pós-graduação, formada para estimular a pesquisa em sistemas complexos. É uma instituição privada e independente, fundada em 1984. O *Santa Fe Institute* está localizado em Santa Fe, New Mexico, USA.

a escassa literatura encontrada, ficaram surpresos com o pequeno esforço demonstrado em explicar como as técnicas conhecidas se relacionavam a outras ou entre si, e qual a confiabilidade de tais técnicas. Instigados por estas dificuldades, Gershenfeld e Weigend decidiram desafiar a comunidade científica propondo uma competição.

A idéia, aparentemente pouco científica, tinha o claro intuito de tentar entender e organizar questões pertinentes à análise de séries temporais e difundir novas técnicas para além dos domínios restritos da área do conhecimento específica, dentro da qual evoluíram. Gershenfeld e Weigend pretendiam também que as séries adotadas na disputa se tornassem *benchmarks* para avaliação de futuros resultados de novas pesquisas.

Para surpresa dos dois pesquisadores, a idéia foi bem recebida pela comunidade científica. A competição foi patrocinada pelo *Santa Fe Institute* e contou com um grupo de consultores, representativo da maior parte das disciplinas relevantes envolvidas com análise de séries temporais. Os consultores foram escolhidos entre pesquisadores das áreas de biologia, economia, física pura, física experimental, astrofísica, análise numérica, estatística e sistemas dinâmicos.

Na primavera de 1992 foi realizado um encontro para explorar os resultados da competição. O *NATO Advanced Research Workshop*[1.2] objetivou reunir os participantes do desafio, membros dos grupos que coletaram dados, consultores e demais interessados.

O livro *Time Series Prediction: Forecasting the Future and Understanding the Past*, referenciado em [3], compila alguns dos interessantes trabalhos apresentados à competição. Como pode ser constatado em [3], o maior interesse demonstrado pelos grupos participantes da competição foi em predição de séries temporais e a maior parte da discussão, centrada em modelos não-lineares. O número dominante de contribuições e também as melhores predições foram devidas aos métodos coneccionistas, conhecidos como redes neurais artificiais [3].

Na verdade, o assunto predição de séries temporais sempre ocupou o interesse dos pesquisadores em redes neurais artificiais. Em 1964, Hu aplicou a rede linear adaptativa de

[1.2] *NATO Advanced Research Workshop* - evento patrocinado pela OTAN através do *NATO Science and Technology*.

Widrow a estudos de previsão climática. Em 1987, Lapedes e Farber treinaram uma rede neural não-linear para emular a relação entre pontos sucessivos para séries temporais geradas computacionalmente e Weigend, Huberman e Rumelhart (1990, 1992) desenvolveram redes de complexidade adequada para séries temporais observadas, colhidas no mundo real [3][7].

Em nossas pesquisas em Telecomunicações, assim como Gershenfeld e Weigend, deparamo-nos repetidamente com a necessidade de um aprofundamento na análise de séries temporais, principalmente no estudo de métodos robustos para predição de séries temporais.

Trabalhávamos em um projeto conjunto do Departamento de Comunicações da Faculdade de Engenharia Elétrica e Computação da Universidade Estadual de Campinas (UNICAMP) com o Centro de Pesquisas e Desenvolvimento da Telebrás (CPqD). O projeto tinha por objetivo desenvolver um controlador de taxa de bits para um codificador MPEG, baseado em predição não-linear, através de redes neurais artificiais. A taxa de bits predita na saída da rede neural artificial era, então, aplicada a um mapeamento não-linear para determinar o passo de quantização do codificador de vídeo [21][33] .

Na pesquisa bibliográfica realizada na ocasião, encontramos os resultados da competição proposta por Gershenfeld e Weigend. Da mesma forma que muitos, nos sentimos instigados a "participar da competição" – o que nos conduziu à concepção de uma nova heurística de predição.

Esta tese apresenta a nova técnica de predição não-linear de séries temporais em que a predição é obtida através de redes neurais artificiais do tipo *Radial Basis Function* (RBF), com atribuição dos centros Gaussianos das funções de base radial por decomposição do conjunto de dados em sub-espaços, técnica que denominamos atribuição de centros por Decomposição em Sub-Espaços (DSE). A decomposição em sub-espaços – ou componentes principais – do espaço de dados é baseada na Transformada Karhunen-Loève (KLT). A predição obtida através da parametrização da rede RBF via DSE resulta em um menor erro de predição e requer menos amostras prévias conhecidas do que as técnicas convencionais [32]. Adicionalmente é apresentada uma possível solução para o problema

de encontrar dinamicamente a arquitetura da rede neural mais apta a acompanhar as não-estacionariedades presentes em muitas séries temporais, heurística que denominamos Janela de Predição Seletiva (JPS).

O Capítulo 2 caracteriza as séries temporais que serão utilizadas como exemplos ilustrativos das técnicas apresentadas ao longo de toda a tese. No Capítulo 3 é descrito o problema da predição linear, a motivação para considerar técnicas de predição não-linear e o critério convencionado para avaliação da performance da predição. O Capítulo 4 descreve como é realizada a predição de séries temporais através de redes neurais artificiais do tipo RBF – aborda as redes neurais artificiais RBF propriamente ditas no contexto de aproximação de funções e redes RBF vistas como filtros preditivos não-lineares, utilizadas nas técnicas convencionais de predição. O Capítulo 5 apresenta a nova técnica de predição por DSE. No Capítulo 6 é descrita a heurística JPS. O Capítulo 7 apresenta comentários e conclusões e, na seqüência, são apresentadas as Referências Bibliográficas e os Apêndices.

Capítulo 2

Caracterização das Séries Temporais Utilizadas nesta Tese

Neste capítulo são descritas as séries temporais que serão utilizadas ao longo desta tese. Para cada série temporal é apresentada uma descrição do conjunto de dados que a originou e a motivação em adotá-la como exemplo. O título de cada seção refere-se ao mnemônico que designa a respectiva série.

As séries apresentadas provêm de eventos do mundo real e, seguindo os critérios propostos por Gershenfeld e Weigend [3], procurou-se utilizar um conjunto de séries que, por suas características, representem o núcleo do problema de análise de séries temporais que emerge de várias áreas do conhecimento.

O conjunto de séries utilizadas compreende três séries provenientes da *Santa Fe Competition* descrita no livro *Time Series Prediction: Forecasting the Future and Understanding the Past*, referenciado em [3], e encontradas em http://www.stern.nyu.edu/~aweigend/TimeSeries/SantaFe.html [29]; uma série proveniente do pioneiro trabalho de Rumelhart *et al.* referenciado em [4], obtida de ftp://ftp.santafe.edu/pub/Time-Series [24]; cinco séries provenientes do mercado de ações brasileiro encontradas em http://www.nettrade.com.br/ [28] e uma série proveniente do volume de tráfego de dados em uma LAN (*Local Area Network*) Ethernet, descrita por Leland, Taqqu, Willinger e Wilson em [57] e por Leland e Wilson em [56], e obtida de [27].

2.1 Série *Sunspots*

Descrição : Cada um dos 280 pontos que constituem esta série corresponde ao número normalizado da ocorrência anual de manchas solares, no período de 1700 a 1979. A representação gráfica da série é mostrada na Figura 2.1.

Fonte: Origem, normalização e definição de regiões de teste e treinamento descritas por Weigend, Huberman e Rumelhart em [4], série obtida de [24].

Motivação: Esta série é clássica no contexto de predição, tendo sido historicamente uma das primeiras séries temporais estudadas.

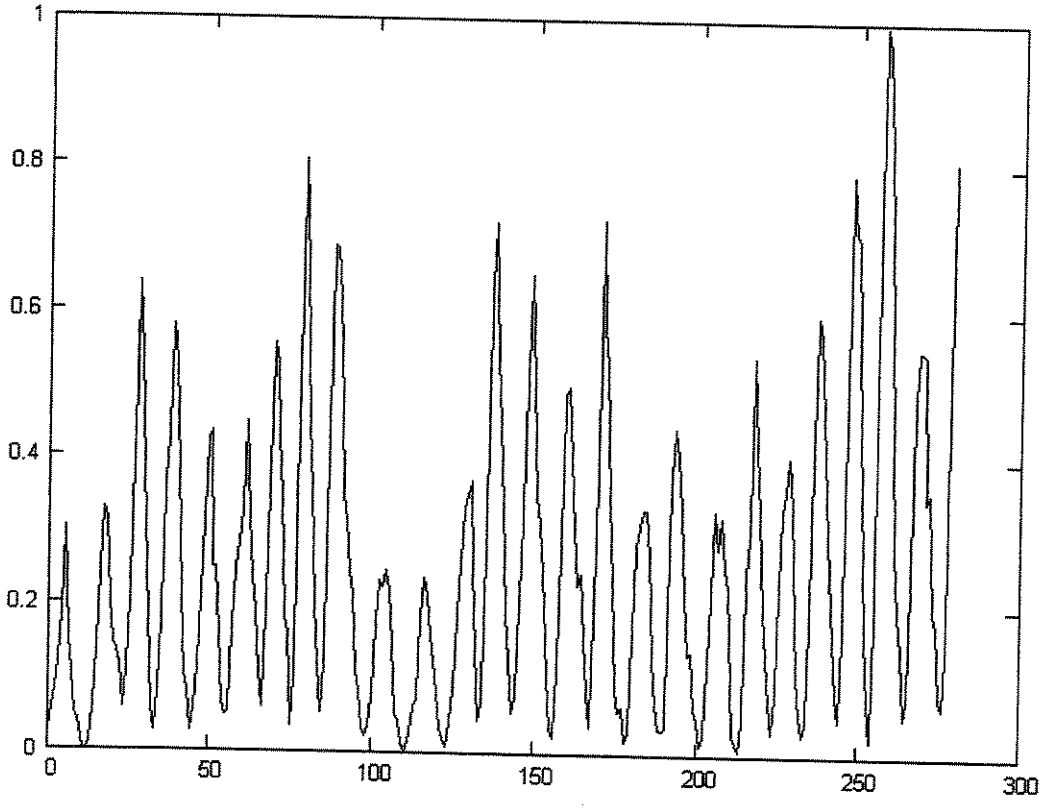

Figura 2.1: Representação gráfica da série *Sunspots*. Ordenada: Número normalizado da ocorrência anual de manchas solares. Abscissa: Índice da observação.

2.2 Série *Chaotic_LASER*

Descrição : Os 1000 pontos que constituem esta série correspondem à intensidade de um *Far-Infrared-LASER* em estado caótico. A série foi obtida por Udo Huebner do *Phys.-Techn. Bundesanstalt*, Braunschweig, Germany. As medidas foram feitas a partir de um LASER NH3 não-pulsado [55]. A representação gráfica da série é mostrada na Figura 2.2.

Fonte: Série descrita por Weigend e Gershenfeld em [3] e obtida de [29].

Motivação: Esta série é um típico exemplo de séries que resultam de um fenômeno físico de comportamento complicado, porém bem representado pela série temporal que descreve uma de suas variáveis, no caso, a intensidade.

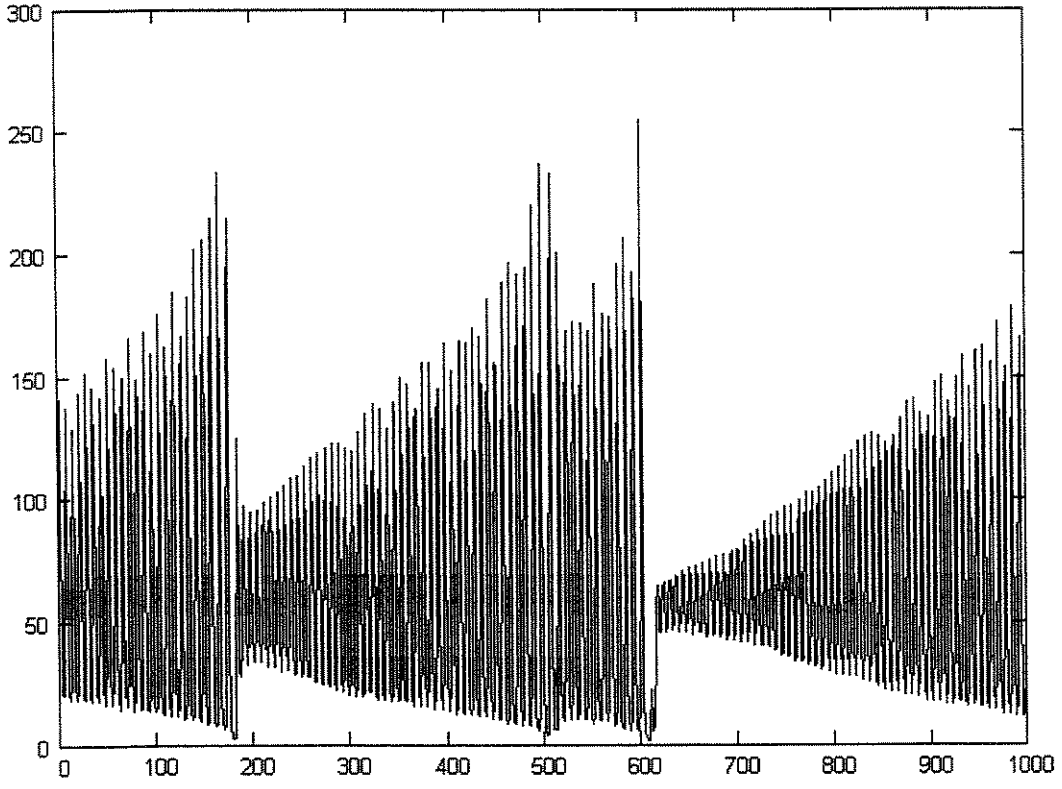

Figura 2.2: Representação gráfica da série *Chaotic_LASER*. Ordenada: Intensidade de um *Far-Infrared-LASER* em estado caótico. Abscissa: Índice da medição.

2.3 Série *Starlight*

Descrição : Esta série provém da observação da curva de intensidade luminosa $I(t)$ da estrela anã-branca PG1159-035 durante o mês de Março de 1989, com um intervalo de $10s$ entre observações consecutivas. $I(t)$ foi dividida em 17 intervalos, e a série *Starlight* , composta por 2605 amostras, corresponde ao intervalo 11 desta curva. A curva $I(t)$ da estrela PG1159-035 foi obtida pelo *Whole Earth Telescope* e foi disponibilizada por J. Dixson e D. Winget do *McDonald Observatory of the University of Texas at Austin*. A representação gráfica da série é mostrada na Figura 2.3.

Fonte: Série descrita por Weigend e Gershenfeld em [3] e obtida de [29].

Motivação: Esta série é um típico exemplo de séries que resultam de um fenômeno físico oscilatório superposto a ruído inerente ao processo de observação – ao longo do mês de observação a absorção de luz pela atmosfera varia significativamente e de maneira aleatória.

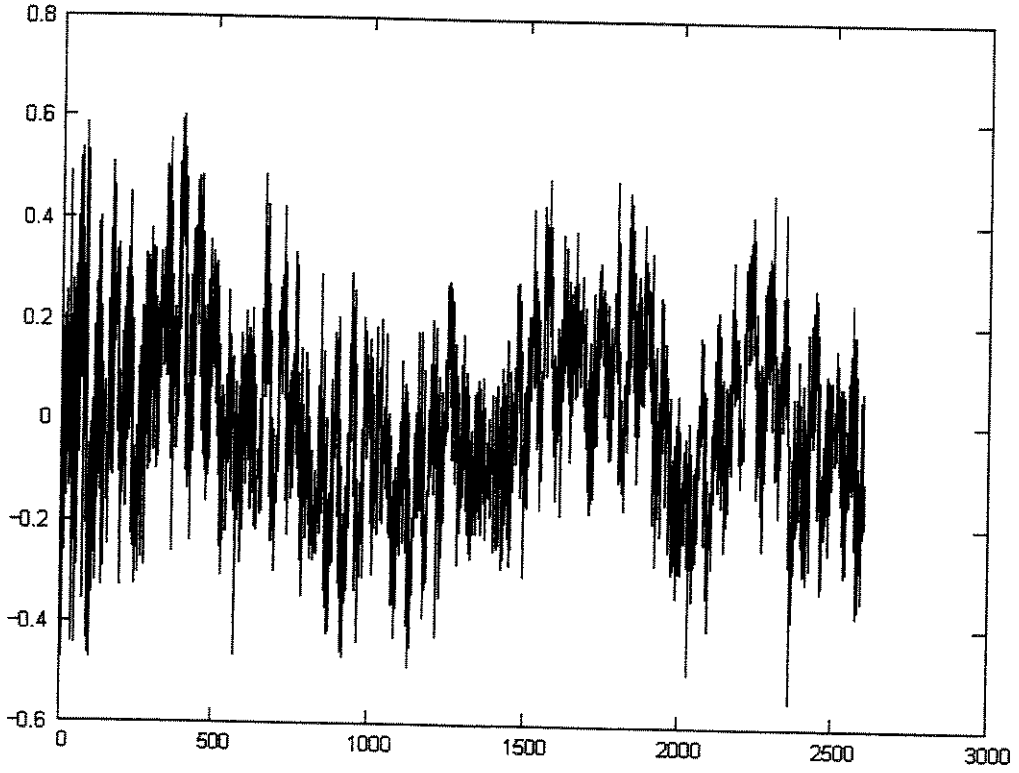

Figura 2.3: Representação gráfica da série *Starlight*. Ordenada: Intensidade luminosa $I(t)$ da estrela anã-branca PG1159-035. Abscissa: Índice da observação.

2.4 Série *Dollar2Franc*

Descrição : Cada um dos 3000 pontos desta série corresponde ao valor da taxa de câmbio US$ / Franco Suíço constante no respectivo boletim informativo emitido pelo *Union Bank of Switzerland* de 7 de agosto 1990 – 2h21min16s até 22 de agosto de 1990 – 9h47min08s, ambos horários contados a partir da abertura das transações do respectivo dia. A representação gráfica da série é mostrada na Figura 2.4.

Fonte: Série descrita por Weigend e Gershenfeld em [3] e obtida de [27].

Motivação: Predizer taxas de câmbio é um problema clássico de grande interesse acadêmico e financeiro. Especialmente na atual conjuntura econômica mundial em que o capital especulativo transformou-se no senhor do destino de nosso triste mundo globalizado.

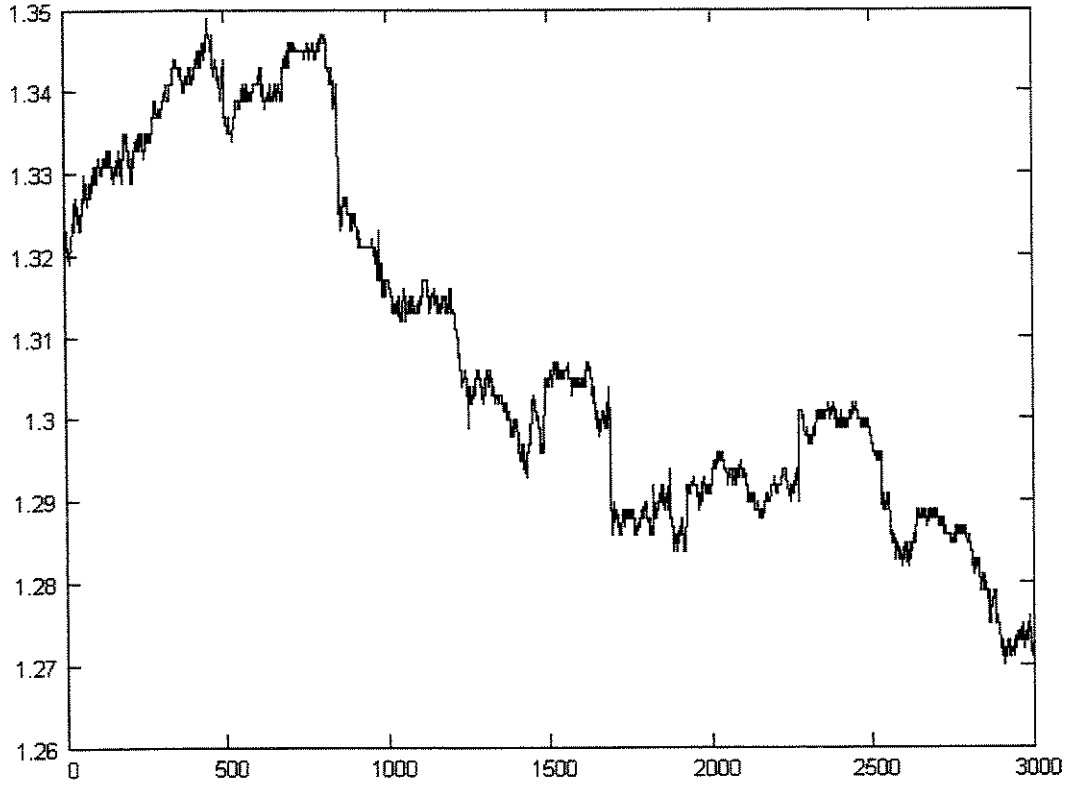

Figura 2.4: Representação gráfica da série *Dollar2Franc*. Ordenada: Taxa de câmbio US$ / Franco Suíço. Abscissa: Índice da cotação.

2.5 Série *Bellco*

Descrição : Os 1000 pontos desta série correspondem à média das 1000 amostras de cada um dos 1000 intervalos da série de 1.000.000 de amostras que representa o monitoramento por 122.797,83 segundos ($\approx$ 35 horas) do número de pacotes Ethernet externos que chegam à rede local do Bellcore Morris Research and Engineering Center, contados a partir das 23h46min de 3 de outubro de 1989, em Morristown, New Jersey, USA. A representação gráfica da série é mostrada na Figura 2.5.

Fonte: Série descrita por Leland *et al.* em [57] e obtida de [29].

Motivação: A habilidade de modelar, e portanto a capacidade de predizer tráfego em redes, é de fundamental importância para que o sistema de gerenciamento possa evitar perdas e atraso de pacotes.

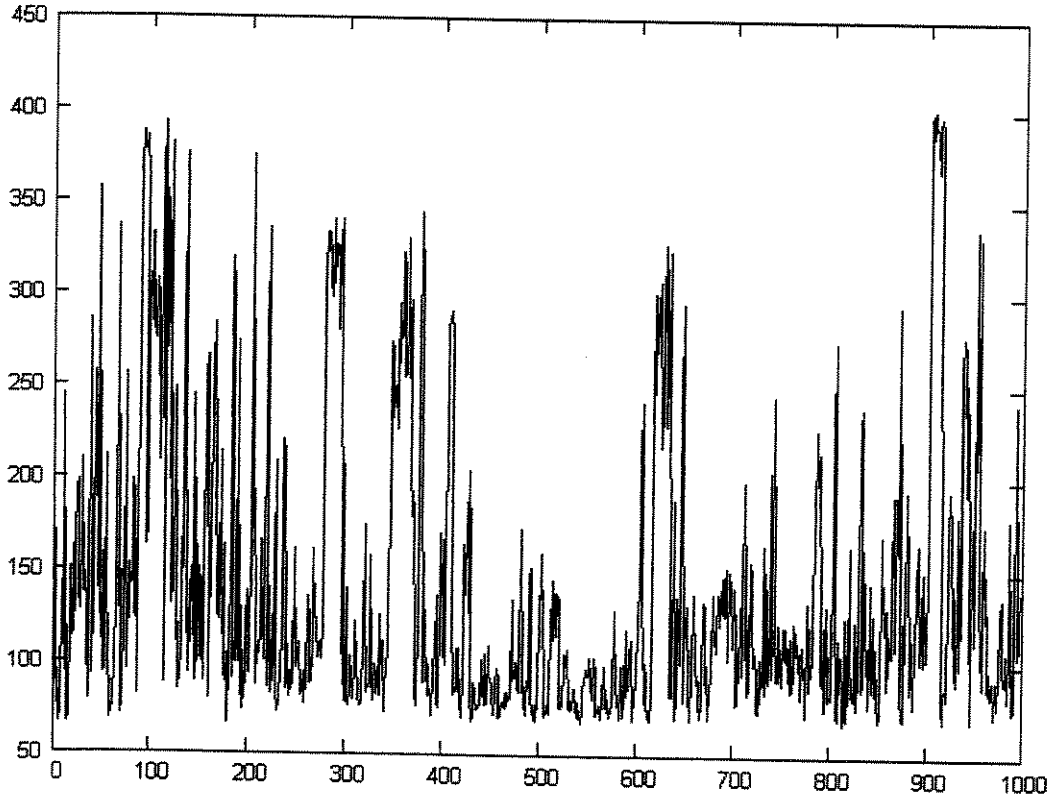

Figura 2.5: Representação gráfica da série *Bellco*. Ordenada: Número de pacotes Ethernet. Abscissa: Índice da amostra.

2.6 Série *Tspc6f*

Descrição : Cada um dos 338 pontos desta série corresponde ao preço de fechamento diário do lote de 1000 ações da Telesp Celular PNB no pregão da BOVESPA no período de 01 de Julho de 1998 a 16 de Novembro de 1999. A representação gráfica da série é mostrada na Figura 2.6.

Fonte: Série descrita pela Nettrade e obtida de [28].

Motivação: Predizer séries temporais obtidas do mercado de ações é um problema tão clássico e de igual interesse acadêmico e financeiro quanto o são as séries de taxas de câmbio.

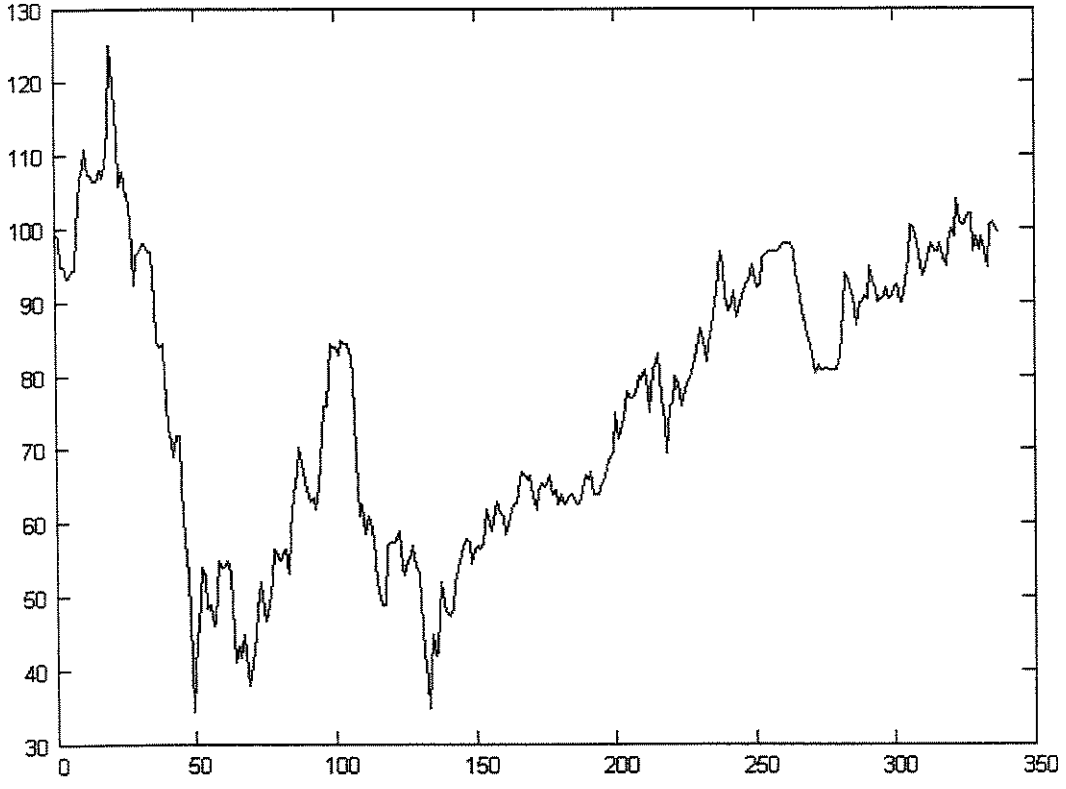

Figura 2.6: Representação gráfica da série *Tspc6f*. Ordenada: Preço de fechamento diário do lote de 1000 ações da Telesp Celular PNB. Abscissa: Índice da cotação.

2.7 Série *Elet3f*

Descrição : Cada um dos 338 pontos desta série corresponde ao preço de fechamento diário do lote de 1000 ações da Eletrobrás ON no pregão da BOVESPA no período de 01 de Julho de 1998 a 16 de Novembro de 1999. A representação gráfica da série é mostrada na Figura 2.7.

Fonte: Série descrita pela Nettrade e obtida de [28].

Motivação: Conforme Seção 2.6.

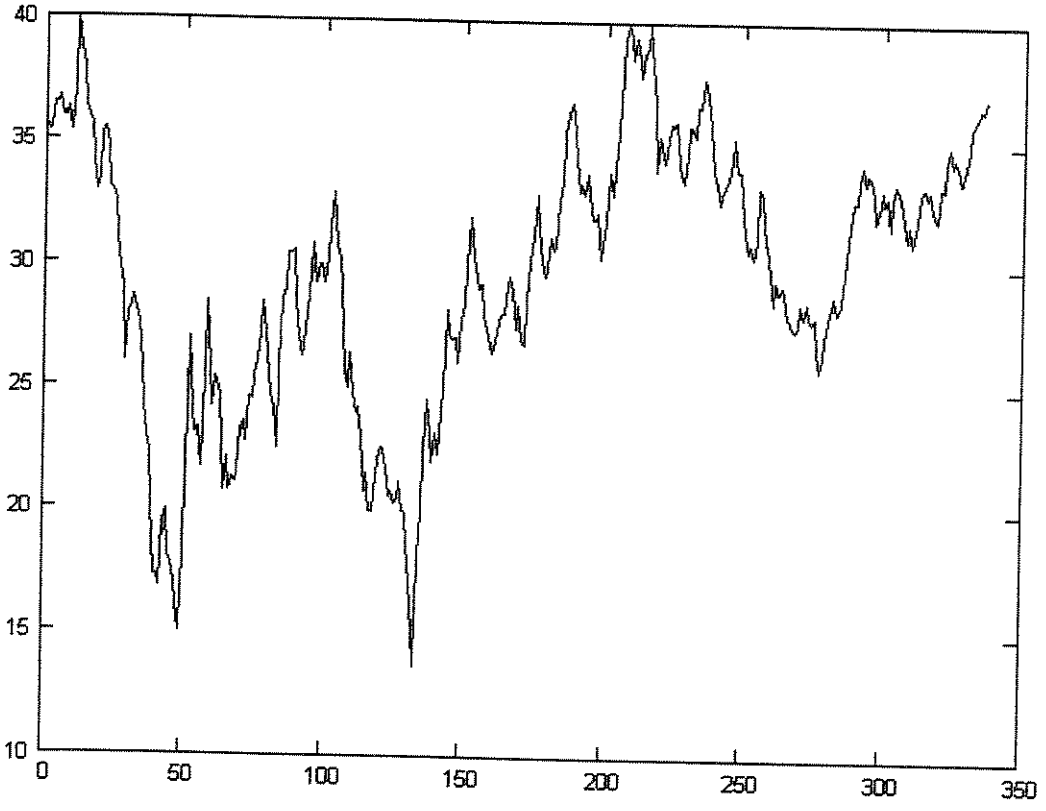

Figura 2.7: Representação gráfica da série *Elet3f* .Ordenada: Preço de fechamento diário do lote de 1000 ações da Eletrobrás ON. Abscissa: Índice da cotação.

2.8 Série *Cmig4f*

Descrição : Cada um dos 337 pontos desta série corresponde ao preço de fechamento diário do lote de 1000 ações da Cemig PN no pregão da BOVESPA no período de 01 de Julho de 1998 a 16 de Novembro de 1999. A representação gráfica da série é mostrada na Figura 2.8.

Fonte: Série descrita pela Nettrade e obtida de [28].

Motivação: Conforme Seção 2.6.

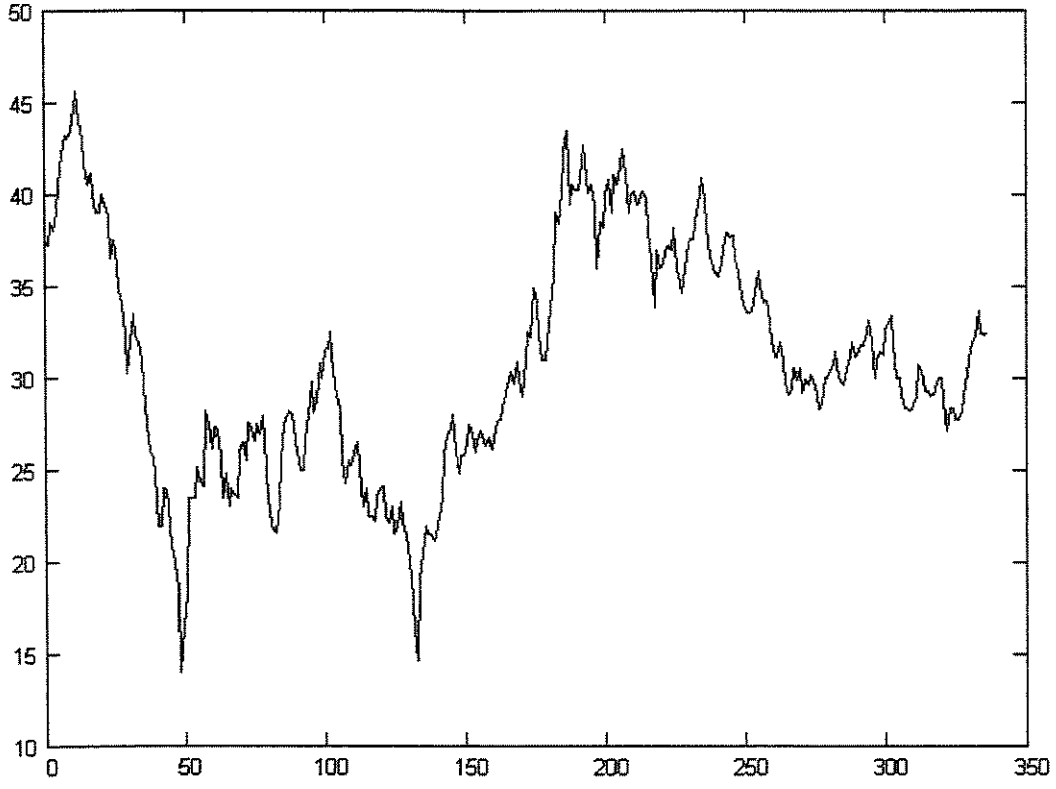

Figura 2.8: Representação gráfica da série *Cmig4f*. Ordenada: Preço de fechamento diário do lote de 1000 ações da Cemig PN . Abscissa: Índice da cotação.

2.9 Série *Petr4f*

Descrição : Cada um dos 337 pontos desta série corresponde ao preço de fechamento diário do lote de 1000 ações da Petrobrás PN no pregão da BOVESPA no período de 01 de Julho de 1998 a 16 de Novembro de 1999. A representação gráfica da série é mostrada na Figura 2.9.

Fonte: Série descrita pela Nettrade e obtida de [28].

Motivação: Conforme Seção 2.6.

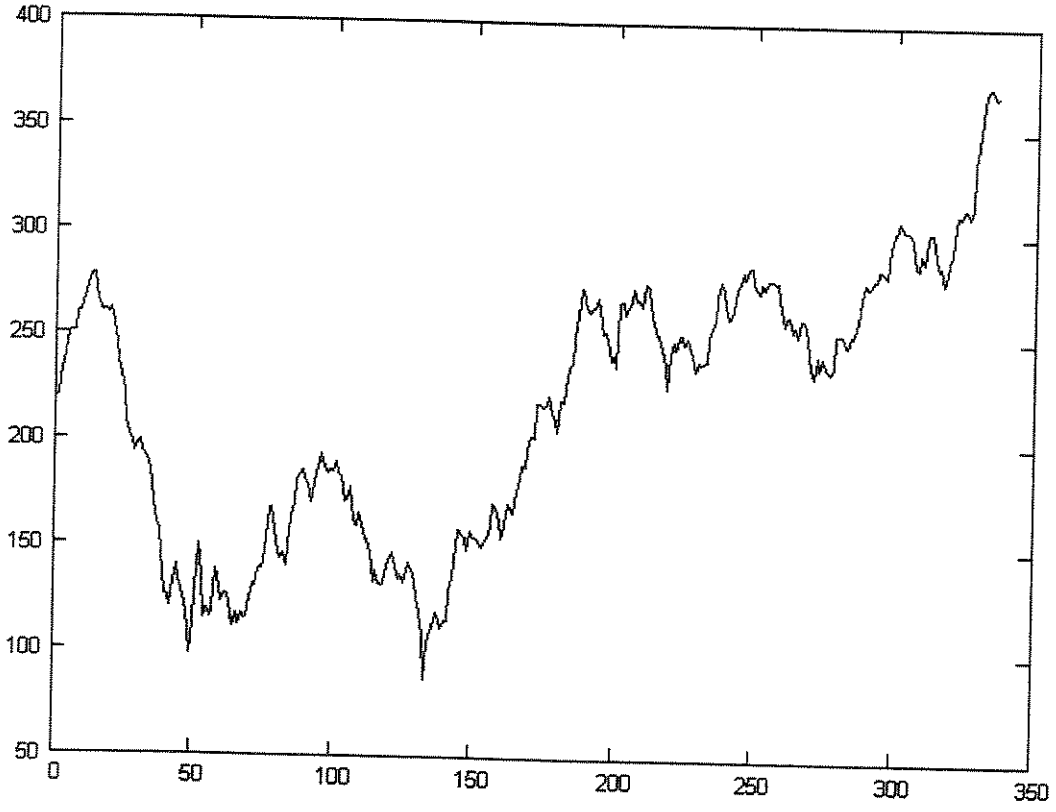

Figura 2.9: Representação gráfica da série *Petr4f*. Ordenada: Preço de fechamento diário do lote de 1000 ações da Petrobrás PN. Abscissa: Índice da cotação.

2.10 Série *Elet6f*

Descrição : Cada um dos 338 pontos desta série corresponde ao preço de fechamento diário do lote de 1000 ações da Eletrobrás PNB no pregão da BOVESPA no período de 01 de Julho de 1998 a 16 de Novembro de 1999. A representação gráfica da série é mostrada na Figura 2.10.

Fonte: Série descrita pela Nettrade e obtida de [28].

Motivação: Conforme Seção 2.6.

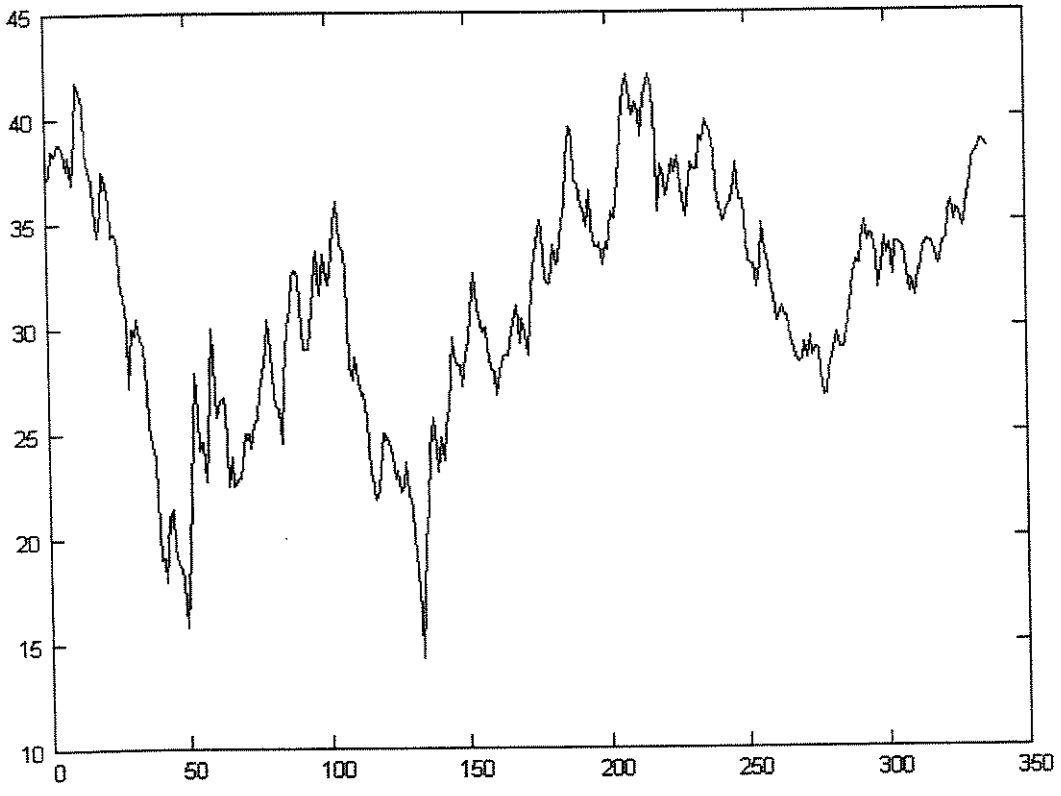

Figura 2.10: Representação gráfica da série *Elet6f*. Ordenada: Preço de fechamento diário do lote de 1000 ações da Eletrobrás PNB. Abscissa: Índice da cotação.

Capítulo 3

Predição de Séries Temporais Descrição do Problema

Este capítulo aborda o problema da predição de séries temporais através da predição linear, apresenta o critério de avaliação de predição a ser utilizado ao longo do desenvolvimento desta tese e introduz o conceito das redes neurais artificiais utilizadas na predição não-linear como uma possível solução aos problemas inerentes à predição linear.

3.1 Predição Linear

Esta tese considera o problema de predição da amostra $u(n+1)$, subseqüente a um conjunto conhecido de amostras consecutivas prévias $\{u(n), u(n-1), \cdots\}$ pertencentes a uma série temporal discreta, problema este conhecido como predição a um passo [3].

Uma das mais celebradas abordagens para a solução deste problema é a técnica denominada predição linear [50]. Por esta razão, o método da predição linear será um dos termos de comparação adotados nesta tese.

Em predição linear, a estimativa da amostra predita, $\hat{u}(n+1)$, é expressa como uma combinação linear de M amostras prévias $\{u(n), u(n-1), \cdots, u(n-M+1)\}$. Os coeficientes W_k, $k = 0, 1, \cdots, M-1$ que ponderam tal combinação linear definem um filtro FIR [47] transversal. A Figura 3.1 detalha um preditor FIR de ordem M, o qual é mostrado no instante n naquela figura [50]. Portanto, como o instante é definido, visando tornar compactas as equações no desenvolvimento que segue, não será explicitado o indexador n

para as variáveis envolvidas, a menos que n não seja inequivocamente definido pelo contexto. Um preditor linear de ordem M utiliza M amostras prévias conhecidas da série temporal para estimar $u(n+1)$, no entanto, necessita do conhecimento de todas as amostras que compõem a série para emular a matriz de correlação associada.

A função de custo J mede o erro médio quadrático entre a estimativa da predição $y(n)=\hat{u}(n+1)$ e o valor efetivamente obtido para a amostra em questão, $u(n+1)$. O vetor $\underline{W}$ que define o filtro FIR tem seus coeficientes determinados de forma a minimizar a função de custo J.

Conforme pode ser observado na Figura 3.1, a amostra predita $\hat{u}(n+1)$ é dada por

$$\hat{u}(n+1)=y(n)=\sum_{k=0}^{M-1} W_k\, u(n-k)=\underline{W}^T\,\underline{u} \tag{3.1}$$

e o erro de predição *e(n)* pode ser expresso por

$$e(n)=d(n)-y(n) \tag{3.2}$$

O operador gradiente é aplicado com o intuito de obter os valores para os pesos W_i do filtro transversal que minimizem a função de custo J, resolvendo-se a equação $\nabla J=0$. Assim, tomando a derivada parcial da função de custo J com relação a cada peso W_i,

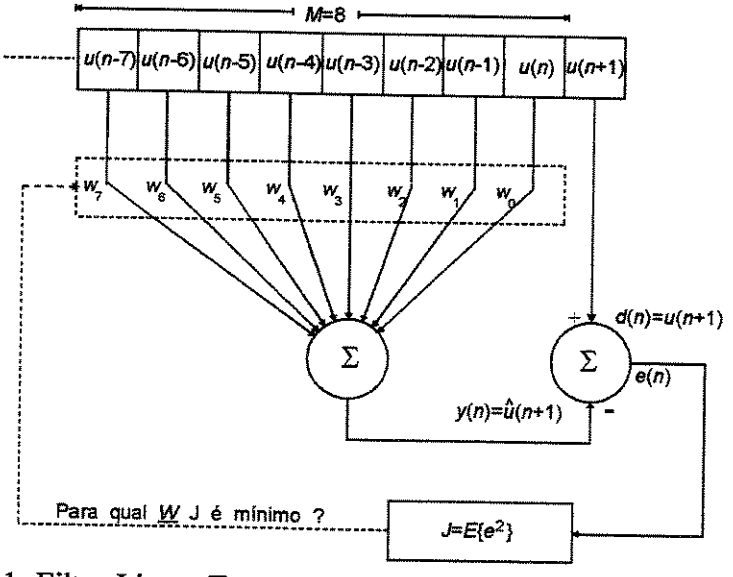

Figura 3.1: Filtro Linear Transversal utilizado como preditor de $u(n+1)$.

$$\nabla_i J = \frac{\partial J}{\partial W_i} = \frac{\partial}{\partial W_i} E\{e^2\} = E\left\{2e\frac{\partial e}{\partial W_i}\right\} = E\left\{2e\frac{\partial}{\partial W_i}(d-y)\right\} = E\left\{-2e\frac{\partial y}{\partial W_i}\right\}; \tag{3.3}$$

$$i = 0, 1, \cdots, M-1$$

E, considerando o gradiente a partir do instante n, teremos

$$\nabla_i J(n) = E\left\{-2e(n)\frac{\partial y(n)}{\partial W_i}\right\} = E\left\{-2e(n)\frac{\partial}{\partial W_i}\sum_{k=0}^{M-1} W_k\, u(n-k)\right\} = -2E\{e(n)u(n-i)\} \tag{3.4}$$

Como a função de custo J é uma função quadrática, J será globalmente mínimo para $\nabla J = 0$. Assim, a partir da Equação (3.4) podemos escrever que

$$\nabla_i J(n) = -2E\{e(n)u(n-i)\} = 0 \tag{3.5}$$

Substituindo as Equações (3.1) e (3.2) na Equação (3.5), obteremos

$$E\left\{\left[d(n) - \sum_{k=0}^{M-1} W_k\, u(n-k)\right] u(n-i)\right\} = 0 \tag{3.6}$$

Distribuindo os produtos e rearranjando a Equação (3.6),

$$\sum_{k=0}^{M-1} W_k E\{u(n-k)u(n-i)\} = E\{d(n)u(n-i)\} \tag{3.7}$$

Observa-se no lado esquerdo da igualdade expressa na Equação (3.7), que

$$E\{u(n-k)u(n-i)\} = R_{uu}(k-i) \tag{3.8}$$

onde R_{uu} é a função de auto-correlação do processo estocástico u para um atraso $k-i$ entre as amostras, com $k, i = 0, 1, \cdots, M-1$. Da mesma forma, observando o lado direito da igualdade expressa na Equação (3.7),

$$E\{d(n)u(n-i)\} = R_{du}(-i) \tag{3.9}$$

onde R_{du} é a função de correlação cruzada entre o processo estocástico que descreve a saída desejada $d = u(n+1)$ e o processo u.

Considerando as Equações (3.8) e (3.9), a Equação (3.7) pode ser reescrita como

$$\sum_{k=0}^{M-1} W_k R_{uu}(k-i) = R_{du}(-i); \qquad i = 0, 1, \cdots, M\text{-}1 \tag{3.10}$$

Para escrever a Equação (3.10) sob a forma matricial, consideremos que seja $\underline{u}(n) = [u(n)\ u(n-1) \cdots u(n-M+1)]^T$, tal que

$$\mathbf{R} = E\left\{\underline{u}(n)\ \underline{u}^T(n)\right\} \tag{3.11}$$

isto é

$$\mathbf{R} = \begin{bmatrix} E\{u(n)u(n)\} & E\{u(n)u(n-1)\} & \cdots & E\{u(n)u(n-M+1)\} \\ E\{u(n-1)u(n)\} & E\{u(n-1)u(n-1)\} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ E\{u(n-M+1)u(n)\} & E\{u(n-M+1)u(n-1)\} & \cdots & E\{u(n-M+1)u(n-M+1)\} \end{bmatrix} \tag{3.12}$$

ou

$$\mathbf{R} = \begin{bmatrix} R_{uu}(0) & R_{uu}(1) & \cdots & R_{uu}(M-1) \\ R_{uu}(1) & R_{uu}(0) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ R_{uu}(M-1) & R_{uu}(M-2) & \cdots & R_{uu}(0) \end{bmatrix} \tag{3.13}$$

Para melhor ilustrar as Equações (3.11), (3.12) e (3.13), consideremos um exemplo. Para o caso em que M=3, teremos $\underline{u}(n) = [u(n)\ u(n-1)\ u(n-2)]^T$ e $\mathbf{R}$ será dado por

$$\mathbf{R} = E\left[\underline{u}(n)\underline{u}^T(n)\right] = \tag{3.14}$$

$$= E\left\{\begin{bmatrix} u(n) \\ u(n-1) \\ u(n-2) \end{bmatrix}\begin{bmatrix} u(n) & u(n-1) & u(n-2) \end{bmatrix}\right\} =$$

$$= \begin{bmatrix} R_{uu}(0) & R_{uu}(-1) & R_{uu}(-2) \\ R_{uu}(1) & R_{uu}(0) & R_{uu}(-1) \\ R_{uu}(2) & R_{uu}(1) & R_{uu}(0) \end{bmatrix}$$

Mas, como $R_{uu}(x) = R_{uu}(-x)$, $\mathbf{R}$ poderá, por fim, ser expresso como

$$\mathbf{R} = \begin{bmatrix} R_{uu}(0) & R_{uu}(1) & R_{uu}(2) \\ R_{uu}(1) & R_{uu}(0) & R_{uu}(1) \\ R_{uu}(2) & R_{uu}(1) & R_{uu}(0) \end{bmatrix} \tag{3.15}$$

Seja, agora, o vetor $\underline{P}$ definido por

$$\underline{P} = E\left\{d(n)\underline{u}(n)\right\} = \tag{3.16}$$

$$= \left[E\{d(n)u(n)\}\ E\{d(n)u(n-1)\} \cdots E\{d(n)u(n-M+1)\}\right]^T =$$

$$= \left[P(0)\ P(-1) \cdots P(1-M)\right]^T$$

e seja também o vetor de pesos dado por

$$\underline{W} = \left[W_0\ W_1 \cdots W_{M-1}\right]^T \tag{3.17}$$

Assim, partindo das Equações (3.10), (3.11), (3.13), (3.16) e (3.17), teremos

$$\mathbf{R}\,\underline{W} = \underline{P} \tag{3.18}$$

A Equação (3.18) é denominada Equação de Wiener-Hopf [50]. A solução de (3.18) para $\underline{W}$ define os coeficientes do filtro linear transversal mostrado na Figura 3.1. O filtro prediz com o mínimo erro quadrático médio a amostra $u(n+1)$ de uma série temporal que apresenta correlação entre as M prévias amostras. Se a matriz de correlação $\mathbf{R}$ da série temporal é não-singular para M definido, então $\underline{W}$ pode ser obtido por

$$\underline{W} = \mathbf{R}^{-1}\underline{P} \tag{3.19}$$

sendo $\underline{P}$ o vetor que define a correlação cruzada entre o vetor $\underline{u}(n)$ de entrada e a saída desejada $d(n)=u(n+1)$.

É importante observar que, para uma dada série temporal com N_t amostras totais, apresentando correlação entre M amostras consecutivas prévias ao instante a ser predito, a precisão com que $\mathbf{R}$ e $\underline{P}$ representam as correlações envolvidas será tanto maior quanto maior for N_t com relação a M.

Isto ocorre porque, na prática, não se conhece o processo estocástico subjacente que determina a série temporal em questão[3.1]. Portanto, não são conhecidas as funções correlações que são realmente envolvidas no processo. Assim, o operador $E\{\cdot\}$ nas Equações (3.11) e (3.16) é substituído pela média dos vetores de M componentes envolvidos no cômputo de $\mathbf{R}$ e $\underline{P}$, média esta realizada sobre o intervalo de N_t amostras totais conhecidas da série temporal.

Desta maneira, a predição linear só tem sentido quando o processo estocástico subjacente é estacionário, pois, em caso contrário, $\mathbf{R}$ e $\underline{P}$ não são univocamente definidas, mesmo para N_t suficientemente grande. Ou seja, se a série temporal resulta de um processo estocástico não-estacionário, $\mathbf{R}$ e $\underline{P}$ variam ao longo da série, invalidando o uso da Equação (3.19) para a obtenção do vetor de pesos $\underline{W}$. A solução algumas vezes adotada é assumir que a série temporal é estacionária em intervalos e adaptar $\mathbf{R}$ e $\underline{P}$ para cada

[3.1] Entenda-se por processo estocástico subjacente o processo estocástico que representa o fenômeno que rege a série temporal em questão, i.é, um processo físico subjacente, um processo econômico subjacente...

intervalo. No entanto, o número de amostras em cada intervalo nem sempre é suficiente para expressar com fidelidade a operação $E\{\cdot\}$.

Como será visto, o método de predição não-linear proposto nesta tese contorna a necessidade de um número grande de amostras conhecidas, suficientes para que o operador $E\{\cdot\}$ seja aproximado com fidelidade.

Objetivando reduzir a complexidade computacional envolvida no cômputo da Equação (3.19), como $\mathbf{R}$ resulta em uma matriz Töeptliz [23], a sua inversão é, em geral, realizada pelo método de Durbin-Levinson [50], muito embora a pseudo-inversão de Moore-Penrose via Decomposição em Valores Singulares [50][59] seja freqüentemente utilizada para contornar os problemas resultantes de uma matriz $\mathbf{R}$ quase singular.

3.1.1 Critério de Avaliação do Erro de Predição

O critério de avaliação adotado nesta tese para o desempenho de predição é o critério sugerido por Gershenfeld e Weigend em [3]. A partir da realização da competição referida no Capítulo 1 desta tese, o critério de avaliação lá utilizado passou a ser considerado referência pela comunidade de pesquisadores, da mesma forma que as séries temporais utilizadas e os melhores resultados obtidos também vieram a se tornar *benchmarks*. Nesta tese foi adotado o mesmo critério de avaliação, por coerência científica.

A qualidade da predição será expressa em termos da razão entre as somas de erros quadráticos mostrada em (3.20).

$$\frac{\sum_t (observação_t - predição_t)^2}{\sum_t (observação_t - observação_{t-1})^2} \tag{3.20}$$

Em (3.20) o denominador expressa o erro médio quadrático (MSE) de predição obtido para a chamada predição pela última amostra. Tal método de predição considera que a melhor predição possível para a próxima amostra consiste simplesmente em repetir o

valor efetivamente observado para a amostra atual. O valor obtido por tal critério para o MSE é tomado como normalizador para o MSE resultante das diferenças entre os valores efetivamente obtidos, após a observação da amostra em questão, e os respectivos valores obtidos pelo preditor que está sendo avaliado. Uma razão inferior a 1 corresponde a uma predição melhor do que aquela obtida pela simples repetição do valor efetivamente observado para a amostra anterior àquela a ser predita – limiar que qualifica um preditor que pretenda ser útil.

O erro obtido através do procedimento expresso em (3.20) é chamado Erro Médio Quadrático Normalizado (*Normalized Mean Squared Error*) e é referido na literatura por NMSE.

Expressando (3.20) em forma de equação, teremos

$$\mathrm{NMSE}(n)=\frac{\sum_{i=1}^{n}(o(i)-p(i))^2}{\sum_{i=1}^{n}(o(i)-o(i-1))^2}=\frac{\sum_{i=0}^{n}(u(i+1)-\hat{u}(i+1))^2}{\sum_{i=0}^{n}(u(i+1)-u(i))^2} \tag{3.21}$$

onde $o(i)$ e $p(i)$ são respectivamente a observação (o valor efetivamente observado) e a predição no instante i. Para uma dada série temporal S com N_t amostras totais o erro ao final do processo de predição de S é dado por $\mathrm{NMSE}(N_t-1)$, onde N_t-1 é o índice do último elemento da série.

3.1.2 Resultados Experimentais

Nesta seção apresenta-se os resultados obtidos para a predição linear da série *Sunspots*, com $N_t=280$ amostras totais, descrita na Seção 2.1. A predição estimada $\hat{u}(n+1)$ é, neste caso, expressa como uma combinação linear de 4 amostras prévias, equivalendo a dizer que a ordem da predição linear adotada é $M=4$. Os valores para os coeficientes que ponderam tal combinação linear foram obtidos através de (3.19) e são

$W_0 = 1.35782$, $W_1 = -0.442764$, $W_2 = -0.186194$ e $W_3 = 0.175639$, de tal forma que a Equação (3.1) resulta em:

$$\hat{u}(n+1) = 1.35782\,u(n) - 0.442764\,u(n-1) - 0.186194\,u(n-2) + 0.175639\,u(n-3)$$

O valor obtido para o Erro Médio Quadrático Normalizado final é $\mathrm{NMSE}(N_t - 1) = 0.608$, conforme Equação (3.21).

É importante salientar que, apesar de a ordem de predição ser $M = 4$, a predição linear necessita do conhecimento prévio de todos os $N_t = 280$ elementos da série temporal para montar a matriz de correlação $\mathbf{R}$, expressa na Equação (3.13).

A Figura 3.2 apresenta as representações gráficas da série *Sunspots*, observada e predita.

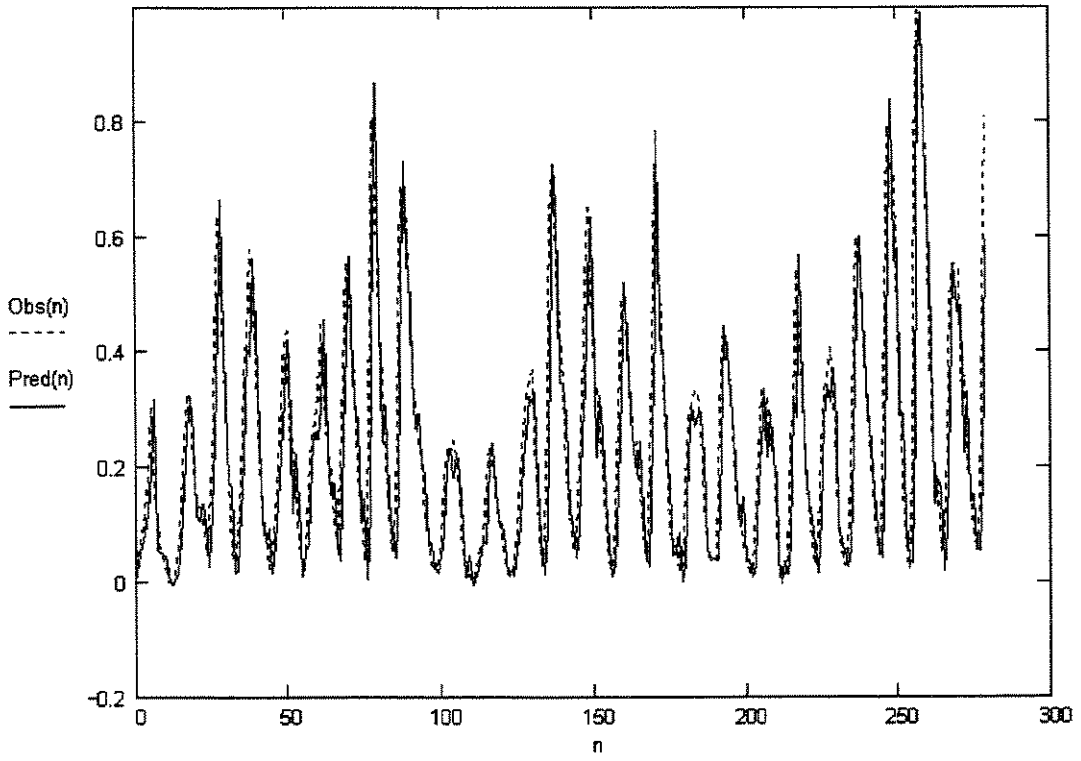

Figura 3.2: Série *Sunspots* – Predição Linear – $M = 4$, $\mathrm{NMSE}(N_t - 1) = 0.608$.

3.2 Predição Não-Linear – Redes Neurais Artificiais

Dentre as redes neurais artificiais existentes há uma classe particular que tem a capacidade de mimetizar processos estocásticos associados a conjuntos de dados através de um aprendizado feito de forma supervisionada. As redes neurais artificiais supervisionadas, assim como na forma convencional de um filtro linear adaptativo, têm a capacidade de, através da informação de uma resposta desejada tentar aproximar um sinal alvo durante o processo de aprendizado. Esta aproximação é obtida através do ajuste, de forma sistemática, de um conjunto de parâmetros livres, característico de cada rede neural. Na verdade, o conjunto de parâmetros livres provê um mecanismo para armazenar o conteúdo de informação subjacente presente nos dados que são apresentados à rede na fase de treinamento [51][11].

Diferentemente da análise estatística tradicional, portanto, as redes neurais não requerem prévio conhecimento sobre a distribuição dos dados, para analisá-los. Desde que haja uma relação subjacente entre os dados, mesmo que desconhecida sua representação analítica e/ou estatística, as redes neurais artificiais podem apresentar um melhor desempenho do que os métodos estatísticos tradicionais.

As redes neurais artificiais são ferramentas extremamente flexíveis em um ambiente dinâmico. Elas têm a capacidade de aprender rapidamente padrões complexos e tendências presentes nos dados e de se adaptar rapidamente às mudanças, características estas que são extremamente desejáveis em se tratando de predição de séries temporais.

Além disso, no caso de predição de séries temporais, quando as séries são governadas por processos subjacentes não-lineares, as técnicas lineares de modelamento têm sucesso apenas limitado em seu desempenho. Especialmente nestes casos, a idéia de empregar redes neurais artificiais é intuitivamente atrativa [40].

Algumas características relevantes das redes neurais artificiais são descritas em [50] e aqui citadas. Segundo S. Haykin, tais atributos justificam e fundamentam a adequação da aplicação de redes neurais artificiais ao processamento de sinais e processos estocásticos.

As qualidades igualmente justificam o uso das redes neurais artificiais no estudo de nosso particular interesse, a predição de séries temporais:

- Possibilidade de considerar o comportamento não-linear dos fenômenos físicos responsáveis pela geração dos dados de entrada;
- Habilidade de aproximar qualquer mapeamento entrada/saída de natureza contínua;
- Necessidade de pouco conhecimento estatístico sobre o ambiente no qual a rede está inserida;
- Capacidade de aprendizado, a qual é atingida através de uma sessão de treinamento com exemplos entrada/saída que sejam representativos do ambiente;
- Capacidade de generalização, a qual permite à rede ter um desempenho satisfatório em resposta a dados não pertencentes ao conjunto de treino;
- Tolerância a falhas, o que permite à rede continuar a apresentar resultados aceitáveis no caso de falha de alguns neurônios – unidades computacionais básicas das redes neurais artificiais;
- Possibilidade da implementação em VLSI, permitindo considerar elevado grau de paralelismo no projeto da rede.

Os dois principais tipos de redes neurais artificiais supervisionadas são as redes MLP (*Multilayer Perceptrons*) treinadas pelo algoritmo *back-propagation* e as redes RBF (*Radial Basis Function*). Ambas as redes são aproximadoras universais. A principal diferença entre elas é que as redes RBF tendem a produzir aproximações locais, enquanto que as redes MLP tendem a resultar em aproximações globais [51][11]. Quando se trata de aprendizado continuado, no entanto, como no caso da predição de séries temporais, as redes MLP se mostram menos adequadas porque o custo computacional de treino de uma rede MLP é muito superior ao de uma rede RBF, o que impossibilta a operação de forma dinâmica [50][51].

A técnica de predição de séries temporais apresentada nesta tese utiliza as redes neurais artificiais do tipo *Radial Basis Function* por serem especialmente adequadas para a predição de séries temporais, mesmo aquelas séries regidas por processos não-lineares e/ou não-estacionários. As redes neurais artificiais do tipo RBF serão detalhadamente abordadas no próximo capítulo desta tese.

Capítulo 4

Predição de Séries Temporais através de Redes Neurais Artificiais RBF

Este capítulo aborda as técnicas convencionais utilizadas na predição de séries temporais através de redes neurais artificiais do tipo RBF.

A série temporal é considerada como um processo com matriz de estados $\mathbf{\Phi}$. Esta matriz é representada por uma rede neural RBF, a qual é considerada como um filtro não-linear com matriz de interpolação definida por $\mathbf{\Phi}$. A matriz de estados $\mathbf{\Phi}$ armazena informação sobre os estados básicos do processo a ser predito com base no conjunto de vetores centro das funções de base radial, vetores estes que constituem os vetores de estado do processo associado à série [50]. Uma janela definida sobre a série temporal é particionada em vetores de entrada da rede RBF e em vetores de estado do processo. Estes vetores, após uma transformação não-linear, são usados para formar a matriz $\mathbf{\Phi}$, a qual armazena os estados do processo – agora visto como um processo não-linear. A cada estado armazenado em $\mathbf{\Phi}$ é associada uma saída desejada da rede RBF, de tal forma a se definir um vetor de saídas desejadas $\underline{d}$. Cada elemento em $\underline{d}$ é definido pelo elemento que está uma posição à frente na série temporal com respeito ao vetor de entrada que gerou o correspondente vetor de estados em $\mathbf{\Phi}$. Assim, $\underline{d} = \mathbf{\Phi} \cdot \underline{w}$, onde $\underline{w}$ é o vetor de pesos sinápticos da rede RBF, o qual define a transição entre os estados prévios do processo associado à série e o próximo e imediato elemento da mesma, isto é, $\underline{w}$ armazena informação sobre o modo como o próximo elemento na série é gerado a partir de seus estados prévios. Por inversão matricial, caso $\mathbf{\Phi}$ não seja singular, pode-se determinar $\underline{w} = \mathbf{\Phi}^{-1} . \underline{d}$. Deslizando a matriz $\mathbf{\Phi}$ uma posição à frente na janela e usando a informação de transição contida em $\underline{w}$ pode-se estimar o próximo elemento na série.

Como introdução ao tema, a primeira seção deste capítulo discorre sobre as redes neurais RBF no contexto da aproximação de funções. A segunda seção descreve as redes RBF sob o enfoque da predição não-linear.

4.1 Redes Neurais Artificiais RBF no Contexto de Aproximação de Funções

As redes neurais artificiais do tipo *Radial Basis Function* (RBF) compõem uma classe de redes neurais artificiais particularmente adequadas à aproximação de funções.

A arquitetura das redes neurais RBF é tal que apresenta uma camada escondida definida por um conjunto de funções de base radial, das quais a rede deriva seu nome.

O aprendizado de uma rede RBF é equivalente a ajustar uma superfície não-linear ao conjunto de dados, em um espaço muti-dimensional, de acordo com algum critério estatístico. O processo de generalização equivale a usar esta superfície multi-dimensional para interpolar outros pontos que não pertençam ao conjunto de treino, mas estejam em sua vizinhança.

Os neurônios da camada escondida de uma rede neural RBF são um conjunto de funções que constitui uma base arbitrária no espaço por eles formado, em cujo espaço o conjunto de entrada pode ser expandido. Os dados representados através de redes neurais RBF são, portanto, expandidos com referência a um conjunto finito de funções de base radial (algumas vezes chamadas de funções de ativação neural [51][50]), cada uma delas centrada em uma particular coordenada do espaço multi-dimensional dos pontos que compõem o espaço de dados de entrada [50][11]. Cada uma destas coordenadas particulares caracteriza-se por definir o centro de uma – entre várias possíveis – região de maior aglomeração de pontos, ou *cluster* [11], do espaço de dados de entrada.

As redes neurais RBF foram originalmente desenvolvidas para interpolação de dados em espaços multi-dimensionais. Segundo B. Mulgrew [5], o problema da interpolação de dados pode ser assim formulado: dado um conjunto de vetores $\{\underline{u}_i\}$ e um conjunto de escalares $\{y_i\}$, busca-se uma função $F(.)$, tal que,

$$y_i = F(\underline{u}_i), \ \forall i \tag{4.1}$$

Desde que definida analiticamente, a função $F(.)$ pode ser usada para mapear vetores $\underline{u}$ que não pertençam ao conjunto original, no conjunto de pontos y associados. Uma possível solução para o mapeamento analítico é escolher $F(\underline{u})$, tal que:

$$F(\underline{u}) = \sum_i w_i \, \phi\left(\|\underline{u} - \underline{u}_i\|^2 \right) \tag{4.2}$$

onde $\phi\left(\|\underline{u} - \underline{u}_i\|^2 \right)$ é uma função escalar radialmente simétrica, tendo $\underline{u}_i$ como centro. Os vetores $\underline{u}_i$ são, por esta razão, referidos como centros no contexto de redes neurais RBF. O operador $\|\cdot\|$ é usualmente a norma Euclidiana – ou Norma L2 – e mede o módulo do vetor argumento, isto é, a distância Euclidiana da ponta do vetor à sua origem [8][51].

Em 1986 Micchelli indicou a existência de um conjunto de funções (tanto limitadas quanto ilimitadas) que são adequadas para interpolação por resultarem em um conjunto de equações lineares para as incógnitas w_i para as quais existe uma única solução [5][50]. A Tabela 4.1 apresenta exemplos destas funções, mais comumente utilizadas como funções de base radial.

O parâmetro σ controla o raio de influência de cada função. Este fator é particularmente evidente no caso da função multi-quadrática inversa e da função Gaussiana, em que ambas as funções são localizadas e monotonicamente decrescentes ($\phi(\zeta) \to 0$ à medida que $\zeta \to \infty$). O parâmetro σ determina o quão rapidamente o valor da função de base radial cai a zero à medida que $\underline{u}$ se afasta do centro $\underline{u}_i$.

Lâmina *spline* fina	$\phi(\zeta)=\frac{\zeta}{\sigma^2}\log\left(\frac{\zeta}{\sigma}\right)$
Multi-Quadrática	$\phi(\zeta)=\sqrt{\zeta^2+\sigma^2}$
Multi-quadrática inversa	$\phi(\zeta)=\frac{1}{\sqrt{\zeta^2+\sigma^2}}$
Gaussiana	$\phi(\zeta)=\exp\left(-\frac{\zeta^2}{2\sigma^2}\right)$

Tabela 4.1: Algumas funções de base radial comumente utilizadas.

A função de base radial do tipo Gaussiana é a mais comumente utilizada em aplicações práticas, e será adotada nas redes neurais RBF utilizadas nesta tese. Neste caso, o parâmetro σ é o desvio-padrão da função Gaussiana. Assim, σ define a distância Euclidiana média (raio médio) que mede o espalhamento dos dados representados pela função de base radial em torno de seu centro.

O raio de cada uma das funções de base radial de uma rede RBF pode assumir diferentes valores, no entanto, para redes RBF reais, o mesmo raio utilizado para cada neurônio não-linear já permite que a rede uniformemente aproxime qualquer função contínua, desde que haja número suficiente de funções de base radial. Na prática, o valor do raio das funções de base radial afeta as propriedades numéricas dos algoritmos de aprendizado, mas não afeta a capacidade geral de aproximação das redes RBF [5][11].

Originalmente, nas primeiras tentativas de aproximação de funções com redes RBF eram utilizadas tantas funções de base radial quantos fossem os padrões do conjunto de dados representativo da função a ser aproximada, objetivando a exatidão da aproximação. No entanto, esta abordagem não só era computacionalmente custosa, como também gerava o problema de *overfitting* quando o objetivo era não só a aproximação como também a interpolação dos pontos que geravam uma determinada função [5][11].

A solução para estes problemas foi apresentada por Broomhead e Lowe (1988) que sugeriram modificações ao procedimento de interpolação exata (que utiliza tantas funções de base radial quantos forem os padrões presentes nos dados). Uma delas é permitir que nem todos os vetores de entrada (do conjunto de dados) tenham uma função de base radial associada. A outra modificação sugerida exclui a necessidade de que a escolha dos centros das funções de base radial seja restrita ao conjunto original de vetores. Para tanto, Broomhead e Lowe reinterpretaram a rede RBF como um estimador de mínimos quadrados (*Least Squares Estimator*) [5][11][50].

Consideradas estas duas generalizações, o sistema de equações lineares cujas incógnitas são os pesos w_i será sobre-determinado. A solução de tal sistema é obtida através do uso da operação de pseudo-inversão matricial de Moore-Penrose para a matriz $\mathbf{\Phi}$ [50][59], e, em conseqüência, será uma solução de mínimo erro médio quadrático.

A abordagem de Broomhead e Lowe resultou em uma redução do custo computacional e no aumento da capacidade de generalização das redes RBF, o que possibilitou a sua aplicação a uma vasta gama de problemas em processamento digital de sinais, tais como predição de séries temporais, modelamento de sistemas, rejeição de interferência e equalização/desconvolução de canal.

A Figura 4.1 apresenta a arquitetura da rede neural RBF que é habitualmente usada em tais aplicações. A rede é composta de uma camada de nós fonte (que conectam a rede a seu ambiente externo), à qual é apresentado o vetor de entrada $\underline{u}(n) \in \Re^M$. Uma única camada intermediária de neurônios não-lineares, cada um deles computando uma função distância entre o vetor de entrada e o centro da função de base radial associada, constitui a chamada camada escondida. Na Figura 4.1 o mapeamento não-linear é expresso por funções de ativação Gaussianas, da forma

$$\varphi_k(n) = \varphi_k\left(\underline{u}(n), \underline{t}_k(n), \sigma_k^2(n)\right) = \exp\left[-\frac{1}{\sigma_k^2(n)}\left\|\underline{u}(n) - \underline{t}_k(n)\right\|^2\right], \tag{4.3}$$

onde $\underline{u}(n) \in \Re^M$ representa o vetor de entrada u no instante n, $\underline{t}_k(n) \in \Re^M$ representa o vetor centro da k–ésima função de base radial $k = 0, 1, \cdots, K-1$, K é o número de funções de base radial, e $\sigma_k^2(n) \in \Re$ é a variância associada a cada uma das funções no instante n.

A camada de saída da rede neural é formada por um único neurônio linear. O neurônio que compõe a camada de saída é definido como um combinador linear das funções de base radial. A saída y da rede RBF é, portanto, a soma das saídas de cada Gaussiana, ponderadas pelos respectivos pesos sinápticos w_k, de tal forma que a combinação linear é expressa por

$$y = \sum_{k=0}^{K-1} w_k \varphi_k\left(\underline{u}, \underline{t}_k, \sigma_k^2\right) \tag{4.4}$$

Nesta equação, o termo $\varphi_k\left(\underline{u}, \underline{t}_k, \sigma_k^2\right)$ é a k–ésima função de base radial. Note que φ_k computa o quadrado da distância Euclidiana $D_k^2 = \| \underline{u} - \underline{t}_k \|^2$ entre um vetor de entrada $\underline{u}$ e o centro $\underline{t}_k$ da k–ésima função de base radial. O sinal de saída produzido pelo k–ésimo neurônio escondido é, portanto, devido à função $\exp(\cdot)$ e ao operador $(\cdot)^2$, uma função não-linear da distância D_k. O fator de escala w_k representa o peso do caminho que conecta o k–ésimo neurônio escondido ao nó de saída da rede. À Equação (4.4) pode, em alguns casos, ser ainda acrescido um termo constante de polarização ou *bias* [50].

A transformação não-linear acima referida é definida pelo conjunto de funções de base radial φ_k e a transformação linear é definida pelo conjunto de pesos w_k, $k = 0, 1, \cdots, K-1$.

O mapeamento entrada/saída de uma rede RBF Gaussiana é muito semelhante à técnica estatística chamada Mistura de Modelos (*mixture models*), que são misturas de distribuição de probabilidades [10][11]. Em particular, as misturas de distribuição de probabilidades Gaussianas têm sido usadas como modelos em uma grande variedade de aplicações onde os dados de interesse provêm de duas ou mais populações misturadas entre si, com parâmetros estatísticos distintos. A resposta φ_k do neurônio k da camada escondida

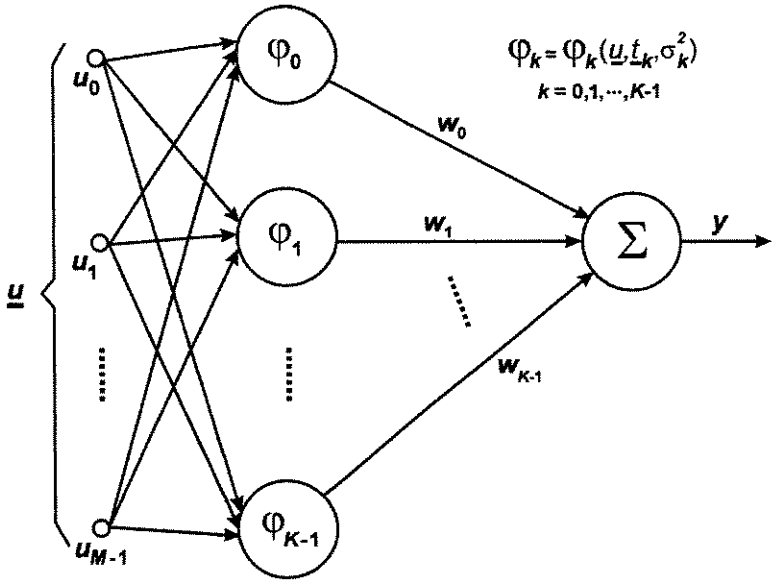

Figura 4.1: Rede neural do tipo *Radial Basis Function*.

de uma rede RBF representa a densidade probabilística condicional de u dado o centro $\underline{t}_k$, isto é, $\varphi_k(u \mid \underline{t}_k)$. O coeficiente w_k representa a probabilidade *a priori* de u no contexto da densidade condicional $\varphi_k(u \mid \underline{t}_k)$. Assim, o conjunto das K densidades probabilísticas condicionais modelam a função de densidade de probabilidade representativa do mecanismo estatístico subjacente que gerou os dados, sendo o modelamento definido através de $\sum_{k=0}^{K-1} w_k \varphi_k(u \mid \underline{t}_k)$. Neste sentido, a rede RBF é muitas vezes referida como um Estimador Bayesiano [10][11].

O procedimento para a implementação de uma rede neural RBF compreende a determinação, através de um processo de aprendizagem, dos valores adequados aos parâmetros livres da RBF, que são as variâncias σ_k^2, os centros $\underline{t}_k$ e os pesos sinápticos w_k. O aprendizado ou treinamento consiste em determinar estes parâmetros de tal forma que, dado um conjunto de estímulos $\underline{u}$ na entrada, as saídas y se aproximem o mais possível do conjunto de valores desejado.

Diferentes algoritmos podem ser utilizados para a adaptação dos parâmetros livres das redes RBF. Por exemplo, o algoritmo *k-means* [50][51] pode ser utilizado para a

inicialização e/ou atualização dos centros das funções de base radial, o algoritmo de Moore-Penrose para pseudo-inversão de matrizes [50] pode ser utilizado para a atualização dos pesos sinápticos, enquanto que o método do Gradiente Estocástico [6][50] pode ser aplicado na atualização dos pesos da rede RBF, das variâncias e dos centros das funções de base radial. A Tabela 4.2 apresenta alguns dos possíveis algoritmos de aprendizado empregados para ajuste dos parâmetros livres.

Possíveis Algoritmos de Aprendizado para Ajuste dos Parâmetros Livres		
Centros das RBF	Pesos Sinápticos	Variâncias dos centros
Constante: Por conhecimento prévio e inferência a partir do conjunto de vetores de treino.	Gradiente Estocástico (LMS). Supervisionado: usa $e(n) = d(n) - y(n)$	Constante: Por conhecimento prévio e inferência a partir do conjunto de vetores de treino.
"Clusterização" pelo algoritmo *k-means*. Não-supervisionado.	Pseudo Inversa por decomposição em valores singulares: $\underline{w}(n) = \mathbf{\Phi}^{-1}(n).\underline{d}(n)$	Gradiente Estocástico (LMS). Supervisionado: usa $e(n) = d(n) - y(n)$
Gradiente Estocástico (LMS). Supervisionado: usa $e(n) = d(n) - y(n)$		$\sigma_k^2(n) = \xi_k(n).\max_{a,b}\left\{\left\|\underline{t}_a(n) - \underline{t}_b(n)\right\|^2\right\}$ onde $\xi_k(n)$ é fixo ou ajustado pelo LMS.

Tabela 4.2: Possíveis Algoritmos de Aprendizado [50].

Diferentes modos de treinamento resultam da combinação dos algoritmos para atualização de centros, variâncias e pesos sinápticos presentes na Tabela 4.2. As redes RBF são, em geral, mais fáceis de treinar do que os *Multilayer Perceptrons* (MLPs), principalmente porque os processos de aprendizagem para os centros, as variâncias e os pesos sinápticos podem ser encadeados seqüencialmente [5], possibilitando que o aprendizado das redes RBF seja otimizado pela resultante divisão de tarefas. Por exemplo, um algoritmo não-supervisionado pode ser utilizado para estimar os centros, uma posterior estimativa da distância do vetor de entrada com respeito a cada centro pode ser usada para

especificar σ e, finalmente, já tendo sido definidos os centros e os raios σ, os pesos sinápticos podem ser calculados através do algoritmo LMS (*Least Mean Squares*)[50]. Após a estimativa inicial de parâmetros da rede, um ajuste mais fino pode ser dado a esta estimativa, utilizando técnicas de gradiente aplicadas a todos os parâmetros, ao invés de apenas aos pesos sinápticos.

Uma abordagem clássica das redes neurais RBF sob a ótica da interpolação de funções é tratada em [50]. Neste modo de treinamento, em que é utilizado o algoritmo Gradiente Estocástico, os centros das funções de base radial e todos os demais parâmetros livres são atualizados através de um processo de aprendizado supervisionado, baseado na minimização da função de custo dada pelo valor esperado do erro quadrático entre a saída fornecida pela rede RBF e a saída desejada para o processo.

Nesta abordagem, os centros das funções de base radial são primeiramente inicializados pelo algoritmo *k-means*. A cada iteração n, o algoritmo *k-means* determina as distâncias entre o vetor $\underline{u}(n)$ pertencente ao conjunto de entrada e cada um dos centros $\underline{t}_k(n)$. Ao centro $\tilde{k}$ que corresponder à menor distância – Equação (4.5) – é aplicada a atualização mostrada na Equação (4.6). Na Equação (4.6), η é a razão de atualização do algoritmo *k-means*.

As iterações prosseguem até que $\| \underline{t}_k(n) - \underline{t}_k(n+1) \| \to \varepsilon$, $\forall k$, onde ε é um número muito pequeno.

$$\tilde{k}(\underline{u}) = \arg\min_k \| \underline{u}(n) - \underline{t}_k(n) \|; \quad k = 0,1,\cdots,K-1 \tag{4.5}$$

$$\underline{t}(n+1) = \begin{cases} \underline{t}_k(n) + \eta [\underline{u}(n) - \underline{t}_k(n)]; & k = \tilde{k}(\underline{u}) \\ \underline{t}_k(n); & \text{outros casos} \end{cases} \tag{4.6}$$

Após a inicialização dos centros pelo algoritmo *k-means*, e definindo a variância inicial comum a todos os centros pela Equação (4.7), os pesos sinápticos são inicializados

com zero. Na Equação (4.7), $d^2_{\max}(\underline{t}_i - \underline{t}_j)$ expressa o quadrado da maior distância Euclidiana entre os centros , isto é, $\max\left\{\left\|\underline{t}_i - \underline{t}_j\right\|^2\right\}$.

$$\sigma^2 = d^2_{\max}(\underline{t}_i - \underline{t}_j); \quad \text{com} \quad i,j = 0,1,\cdots,K-1 \tag{4.7}$$

Para cada vetor $\underline{u}(n)$ do conjunto de treino apresentado à entrada da rede, a saída $y(n)$ da rede RBF é determinada e é avaliada a diferença entre este valor de saída e aquele desejado $d(n)$, conforme

$$e(n) = d(n) - y(n) \tag{4.8}$$

O erro assim obtido é utilizado para a posterior atualização até a convergência dos centros, pesos sinápticos e variância. As equações de atualização, derivadas do algoritmo Gradiente Estocástico [50], são expressas por

$$w_k(n+1) = w_k(n) + \mu_w e(n)\varphi_k(n) \tag{4.9}$$

$$\underline{t}_k(n+1) = \underline{t}_k(n) + 2\mu_t e(n) w_k(n)\varphi_k(n)\frac{\underline{u}(n) - \underline{t}_k(n)}{\sigma_k^2(n)} \tag{4.10}$$

$$\sigma_k^2(n+1) = \sigma_k^2(n) + \mu_\sigma e(n) w_k(n)\varphi_k(n)\frac{\left\|\underline{u}(n) - \underline{t}_k(n)\right\|^2}{\left(\sigma_k^2(n)\right)^2} \tag{4.11}$$

A apresentação de N vetores de dados à entrada $\underline{u}(n)$ da RBF, $n = 0,1,\cdots,N-1$, constitui uma época de treino. Ao final de cada época, o conjunto de vetores de dados é embaralhado, para evitar que a rede aprenda o padrão seqüencial de apresentação dos vetores de treino. Isto porque estamos interessados na capacidade de generalização da rede com relação ao conjunto de dados em si, e a mesma ordem de apresentação dos vetores a cada época de treino poderia prejudicar tal capacidade de generalização.

O treinamento de uma RBF através das Equações (4.9) a (4.11) é continuado até a sua convergência, situação em que o valor obtido para o erro de aproximação é menor que um valor máximo permitido ε. O erro de aproximação será definido na Seção 4.1.1.

4.1.1 Critério de Avaliação do Erro de Aproximação

Especificamente ao longo da Seção 4.1, em que é avaliada a capacidade de aproximação das redes RBF e não a sua capacidade preditiva, a medida de erro será definida de modo diferente do NMSE expresso pela Equação (3.21). Definiremos aqui o NMSEA, ou seja, o Erro Médio Quadrático Normalizado de Aproximação dado por

$$\text{NMSEA} = \frac{1}{N}\sum_{n=0}^{N-1}\frac{(d(n)-y(n))^2}{d^2(n)} \tag{4.12}$$

4.1.2 Resultados Experimentais

Este item destina-se à apresentação dos resultados experimentais obtidos no procedimento de treinamento em que os centros são inicializados pelo algoritmo *k-means* e a atualização até a convergência dos centros, variâncias e pesos sinápticos é efetuada através do Gradiente Estocástico. Este é o algoritmo clássico apresentado em [50] e foi descrito na Seção 4.1 deste capítulo.

Considere-se a situação hipotética em que se deseja estabelecer a relação analítica entre os oito primeiro algarismos representados em base binária e os valores que definem o quadrado de um décimo de sua representação em base decimal, conforme mostrado na Tabela 4.3.

$\underline{u} \in \Re^3$			$F(\underline{u})$
u_0	u_1	u_2	
0	0	0	0.0
0	0	1	0.01
0	1	0	0.04
0	1	1	0.09
1	0	0	0.16
1	0	1	0.25
1	1	0	0.36
1	1	1	0.49

Tabela 4.3: Mapeamento $F:\Re^3 \rightarrow \Re$ que se deseja aproximar.

Uma possível solução seria tentar expressar $F(\underline{u})$, $\underline{u}=[u_0 \;\; u_1 \;\; u_2]^T$, através do mapeamento linear $F(\underline{u})=\underline{w}^T\underline{u}=w_0u_0+w_1u_1+w_2u_2$, sendo $\underline{w}=[w_0 \;\; w_1 \;\; w_2]^T$ o vetor que define $F(\underline{u})$ obtido da solução do sistema de equações

$$\begin{aligned} 0w_0+0w_1+0w_2 &= 0.0 \\ 0w_0+0w_1+1w_2 &= 0.01 \\ 0w_0+1w_1+0w_2 &= 0.04 \\ 0w_0+1w_1+1w_2 &= 0.09 \\ 1w_0+0w_1+0w_2 &= 0.16 \\ 1w_0+0w_1+1w_2 &= 0.25 \\ 1w_0+1w_1+0w_2 &= 0.36 \\ 1w_0+1w_1+1w_2 &= 0.49 \end{aligned} \tag{4.13}$$

o qual é evidentemente sobre-determinado – sem solução – pois não existe $\underline{w}=[w_0 \;\; w_1 \;\; w_2]^T$ que atenda simultaneamente todas as Equações (4.13).

No entanto, o mapeamento $F:\Re^3 \rightarrow \Re$ pode ser feito um mapeamento não-linear quando expresso pela rede neural RBF da Figura 4.1. Especificamente, o conjunto de

$N=8$ vetores $\underline{u}(n)\in\Re^3$ da Tabela 4.3, $n=0,1,\cdots,N-1$, é considerado como o conjunto de treino da rede RBF, com saída desejada $d(n)$ dado pelo respectivo valor na coluna $F(\underline{u})$ da tabela, isto é, $d(n)=F(\underline{u}(n))$. Como $\underline{u}(n)\in\Re^3$, faz-se $M=3$. Adotou-se $K=3$ funções de base radial para formar a superfície de aproximação. Os centros das funções de base radial são inicializados pelo algoritmo *k-means*, através da Equação (4.6) com $\eta=0.1$, e as variâncias são inicializadas pelo valor dado pela Equação (4.7). A atualização dos pesos sinápticos, centros e variâncias, através das Equações (4.9) a (4.11), utiliza as razões de aprendizado $\mu_w=0.1$, $\mu_t=0.1$ e $\mu_\sigma=0.1$.

Para o treinamento da rede RBF, o conjunto de treino é normalizado de forma que os seus valores extremos situem-se no intervalo $[-1,1]$. Esta é uma precaução usual, que visa evitar *overflow* das variáveis de ponto flutuante ao longo da operação do algoritmo Gradiente Estocástico. Cada componente do conjunto de vetores $\underline{u}(n)\in\Re^3$ é normalizado através da transformação $\theta_u:\Re\to\Re$, $\theta_u(x)=2x-1$, e cada um dos valores do conjunto de N saídas desejadas $d(n), n=0,1,\cdots,N-1$, é normalizado através da transformação $\theta_d:\Re\to\Re$, $\theta_d(x)=4.08163x-1$.

A apresentação dos $N=8$ vetores do conjunto de treino definido pela Tabela 4.3 constitui uma época. A Figura 4.2 mostra a evolução do NMSEA, dado por (4.12), à medida que as épocas de treinamento se sucedem.

Após 2000 épocas de treino, a rede RBF apresenta a relação analítica dada pela Equação (4.14) como aproximação para o mapeamento da Tabela 4.3. Note que a Equação (4.14) inclui o efeito das normalizações θ_u e θ_d.

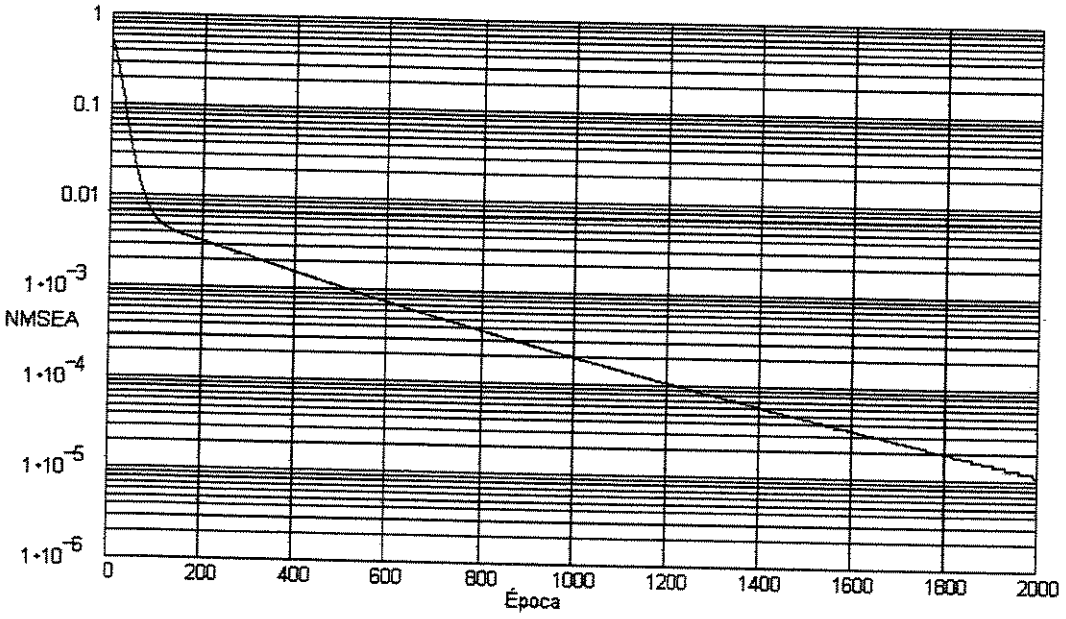

Figura 4.2: Evolução do NMSEA à medida que as épocas de treinamento se sucedem, com conjunto de treino definido pela Tabela 4.3.

$$F(\underline{u}) = \frac{\left\{ -1.948417\exp\left(\frac{-\left\| (2\underline{u}-\underline{1}) - \begin{bmatrix} -0.069259 \\ -0.962758 \\ 0.015931 \end{bmatrix} \right\|^2}{2.774243} \right) + \left\{ 2.111299\exp\left(\frac{-\left\| (2\underline{u}-\underline{1}) - \begin{bmatrix} 1.265731 \\ 0.180856 \\ 0.316373 \end{bmatrix} \right\|^2}{2.372715} \right) + \left\{ -1.045209\exp\left(\frac{-\left\| (2\underline{u}-\underline{1}) - \begin{bmatrix} -1.234718 \\ 1.019335 \\ -0.118491 \end{bmatrix} \right\|^2}{1.934487} \right) \right\} + 1 \right\}}{4.08163} \tag{4.14}$$

Verificando a consistência de (4.14) para representar o mapeamento da Tabela 4.3, temos que $F([0\,0\,0]^T) = -7.739\times10^{-6}$, $F([0\,0\,1]^T) = 0.01$, $F([0\,1\,0]^T) = 0.04$, $F([0\,1\,1]^T) = 0.09$, $F([1\,0\,0]^T) = 0.16$, $F([1\,0\,1]^T) = 0.25$, $F([1\,1\,0]^T) = 0.36$, $F([1\,1\,1]^T) = 0.49$. Portanto, a Equação (4.14) representa a Tabela 4.3 com boa aproximação.

4.2 Redes Neurais Artificiais RBF no Contexto de Filtragem Preditiva Não-Linear

A rede neural artificial RBF utilizada para predição não-linear de séries temporais é dita dinâmica, porque o aprendizado acontece de forma contínua com o desenrolar temporal da série [50]. A Figura 4.3 apresenta a arquitetura da rede RBF em questão.

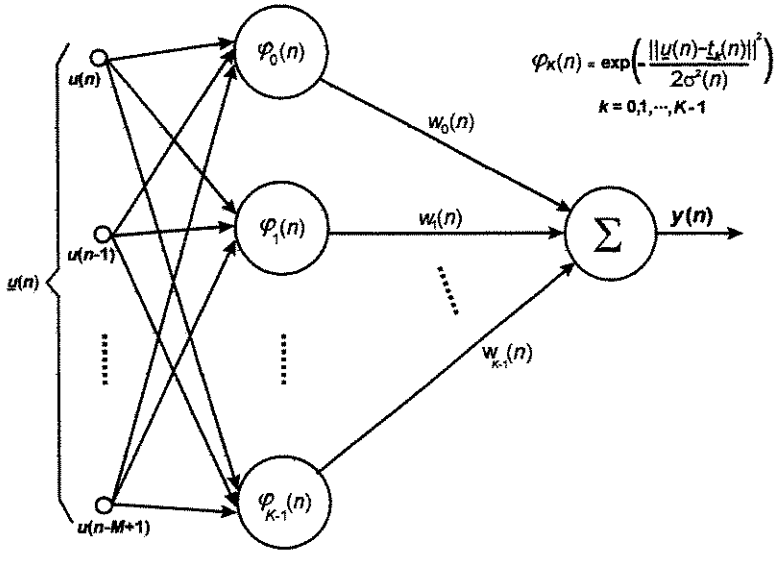

Figura 4.3: Rede neural RBF utilizada para predição não-linear de séries temporais.

Assim como a rede mostrada na Figura 4.1, esta rede RBF possui M nós de entrada e K centros Gaussianos. O vetor $\underline{t}_k \in \Re^M$ é o k–ésimo vetor centro da rede RBF ou k–ésimo vetor de estado do processo associado à série temporal S, w_k é o k–ésimo peso sináptico e σ^2 é a variância comum a todos os centros Gaussianos, com $k = 0, 1, \cdots, K-1$. No contexto de predição de séries temporais, K é o número de vetores de estado do processo e M é a ordem de predição [50][51].

A saída da rede neural RBF quando o n–ésimo vetor de entrada $\underline{u}(n) \in \Re^M$ é apresentado à sua entrada é

$$y(n) = \sum_{k=0}^{K-1} w_k(n)\varphi_k(n) = \underline{\varphi}^T(n)\underline{w}(n) \tag{4.15}$$

onde

$$\varphi_k(n) = \exp\left[-\frac{1}{2\sigma^2(n)}\left\|\underline{u}(n) - \underline{t}_k(n)\right\|^2\right] \tag{4.16}$$

é a saída do k–ésimo centro Gaussiano com o vetor $\underline{u}(n)$ aplicado à entrada da rede RBF.

Para utilizar a rede RBF no contexto de predição de séries temporais torna-se necessária a definição de algumas estruturas de dados.

Seja S uma série temporal definida por $S = \{u(0), u(1), \cdots, u(N_t - 1)\}$, onde N_t é o número de amostras de S. A um instante n qualquer, o objetivo é predizer a amostra $u(n+1)$ de S, sendo conhecidas as $N = K + M$ amostras prévias $u(n)$, $u(n-1), \cdots, u(n-M-K+1)$ que compõem a janela de predição $p(n) = \{u(n-M-K+1), \cdots, u(n-1), u(n)\}$ definida sobre S.

No instante n, seja o processo associado ao desenrolar temporal de S representado pelo conjunto $\mathbf{U}$ de $K+1$ vetores $\underline{u}(n-\delta) \in \Re^M$, $\delta = 0, 1, \cdots, K$, definido sobre a janela $p(n)$, de forma que dois vetores consecutivos de $\mathbf{U}$ estejam deslocados entre si da distância temporal entre duas amostras subseqüentes de S, conforme mostra a Figura 4.4.

Seja, ainda, o vetor de entrada da rede RBF no instante n dado por

$$\underline{u}(n)=[u(n)\ \ u(n-1)\ \cdots\ \ u(n-M+1)]^T \tag{4.17}$$

e seja, no instante n, o k–ésimo vetor de estado $\underline{t}_k(n)$ do processo associado ao desenrolar temporal de S, dado por

$$\underline{t}_k(n)=\underline{u}(n-k-1),\ \ k=0,1,\cdots,K-1 \tag{4.18}$$

A Figura 4.4 apresenta, a título ilustrativo, a janela $p(n)$ definida sobre S, a construção dos vetores $\underline{t}_k$ de estado do processo e os vetores de entrada $\underline{u}(n-j)$, $j=0,1,\cdots,K$, para o caso em que $K=4$ e $M=3$.

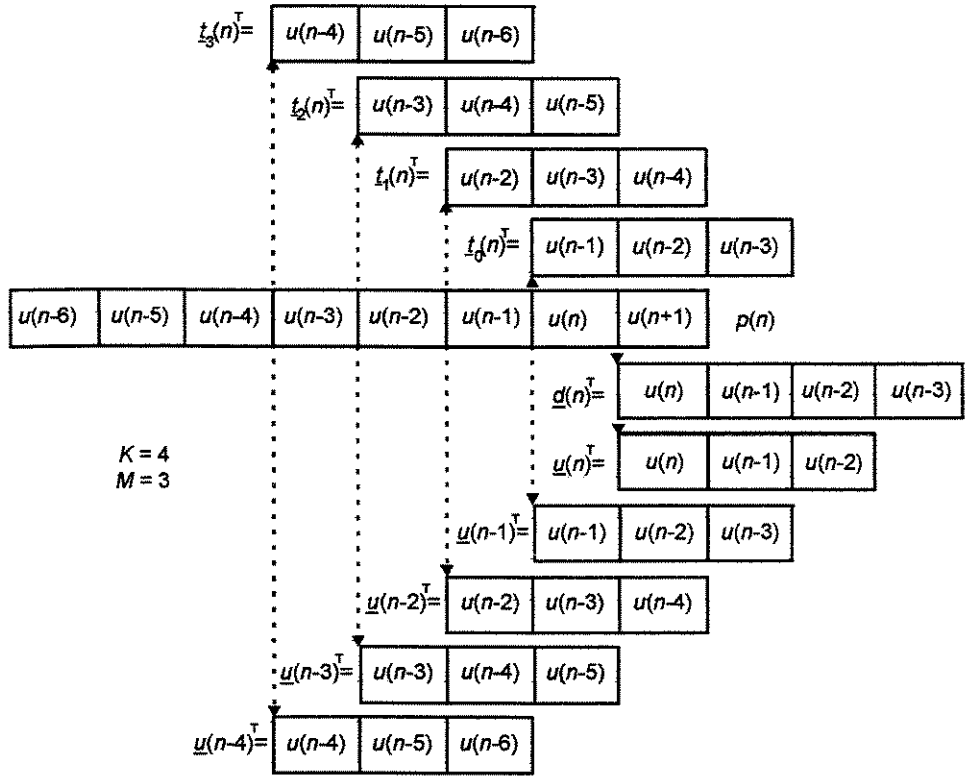

Figura 4.4: Elementos de interesse na série temporal S no instante n e construção dos vetores $\underline{t}_k$ de estado do processo associado à S e dos vetores de entrada $\underline{u}(n-j)$, $j=0,1,\cdots,K$, $k=0,1,\cdots,K-1$, para o caso $K=4$ e $M=3$. Observe que $p(n)$ é formada por $N=K+M=7$ amostras conhecidas, prévias a $u(n+1)$.

Para que se defina a variância $\sigma^2(n)$ comum a todos os centros Gaussianos no instante n, assume-se que $\sigma^2(n)$ seja proporcional ao quadrado da máxima distância Euclidiana entre todos os vetores de estado do processo [50]:

$$\sigma^2(n) = \xi \max\left\{\left\|\underline{t}_i(n) - \underline{t}_j(n)\right\|^2\right\}, \quad i, j = 0,1,\cdots,K-1 \tag{4.19}$$

onde ξ é a constante de proporção chamada Fator de Variância, a qual absorve a constante 2 da Equação (4.16).

Assim, a saída do k–ésimo centro Gaussiano, quando o vetor $\underline{u}(n)$ é aplicado à entrada da rede RBF pode ser definida como

$$\varphi_k(n) = \exp\left\{-\frac{\left\|\underline{u}(n) - \underline{t}_k(n)\right\|^2}{\xi \max\left\{\left\|\underline{t}_i(n) - \underline{t}_j(n)\right\|^2\right\}}\right\}, \quad i, j = 0,1,\cdots,K-1 \tag{4.20}$$

O conjunto $\varphi_k(n)$, $k = 0,1,\cdots,K-1$, de saídas dos K centros Gaussianos, conjunto que resulta da aplicação do vetor de entrada $\underline{u}(n)$, pode ser colocado na forma vetorial através de

$$\underline{\varphi}(n) = \left[\varphi_0(n) \;\; \varphi_1(n) \;\; \cdots \;\; \varphi_{K-1}(n)\right]^T \tag{4.21}$$

Por exemplo, $\underline{\varphi}(n-1) \in \Re^K$ é o vetor que resulta da aplicação do vetor $\underline{u}(n-1) \in \Re^M$ à entrada da rede RBF. Os elementos do vetor $\underline{\varphi}(n-1) \in \Re^K$ são as saídas de cada centro Gaussiano ao vetor $\underline{u}(n-1)$. O k–ésimo centro Gaussiano é definido por seu respectivo vetor $\underline{t}_k \in \Re^M$ de estado do processo, $k = 0,1,\cdots,K-1$. Note que, a qualquer instante arbitrário, a transformação não-linear definida pela Equação (4.20) mapeia o vetor $\underline{u}(n-\delta) \in \Re^M$ aplicado à entrada da rede RBF, δ é um atraso arbitrário qualquer, no vetor $\underline{\varphi}(n-\delta) \in \Re^K$. Portanto, a seqüência de vetores de entrada $\underline{u}(n-1)$, $\underline{u}(n-2)$, $\cdots$, $\underline{u}(n-K)$ define a seqüência de vetores $\underline{\varphi}(n-1)$, $\underline{\varphi}(n-2)$, $\cdots$, $\underline{\varphi}(n-K)$ obtidos pelo mapeamento

não-linear $\Re^M \rightarrow \Re^K$ definido pela Equação (4.20). Assim, apesar de definidos em uma dimensão diferente da dimensão original dos vetores $\underline{t}_k \in \Re^M$ de estados da série S, e apesar de obtidos através de uma transformação não-linear entre as dimensões $\Re^M$ e $\Re^K$, o conjunto de vetores $\underline{\varphi}(n-1), \underline{\varphi}(n-2), \cdots, \underline{\varphi}(n-K)$ definidos em $\Re^K$ ainda armazena informação sobre o desenrolar temporal da série. Portanto, os vetores em $\Re^K$ contêm informação implícita sobre os estados de S, definidos agora em uma outra dimensão, com S sendo "vista" através de um processo não-linear. Assim, o conjunto de vetores em $\Re^K$ pode ser agrupado em uma matriz de transição de estados da série temporal S, agora interpretada como um processo não-linear com K estados $\Re^K$ dimensionais.

A matriz de transição de estados da série temporal S, tomada como um processo não-linear com K estados $\Re^K$ dimensionais no instante n, é mostrada na Equação (4.22).

$$\Phi(n) = \begin{bmatrix} \underline{\varphi}(n-1)^T \\ \underline{\varphi}(n-2)^T \\ \vdots \\ \underline{\varphi}(n-K)^T \end{bmatrix} = \begin{bmatrix} \varphi_0(n-1) & \varphi_1(n-1) & \cdots & \varphi_{K-1}(n-1) \\ \varphi_0(n-2) & \varphi_1(n-2) & \cdots & \varphi_{K-1}(n-2) \\ \vdots & \vdots & & \vdots \\ \varphi_0(n-K) & \varphi_1(n-K) & \cdots & \varphi_{K-1}(n-K) \end{bmatrix} \tag{4.22}$$

Note que as linhas de $\Phi(n)$ correspondem aos vetores de transição de estado $\underline{\varphi}(n-1)^T, \underline{\varphi}(n-2)^T, \cdots, \underline{\varphi}(n-K)^T$ do processo não-linear, os quais resultam respectivamente da aplicação dos vetores $\underline{u}(n-1), \underline{u}(n-2), \cdots, \underline{u}(n-K)$ à entrada da rede RBF. Note também, da transformação (4.20), que o k-ésimo componente do vetor $\underline{\varphi}(n-\delta) \in \Re^K$, δ arbitrário, tende para o valor máximo 1.0 à medida em que o vetor $\underline{u}(n-\delta) \in \Re^M$ aplicado à entrada da rede RBF tende para o vetor de estado $\underline{t}_k$ da k-ésima função de base radial. Portanto, (4.20) é uma transformação não-linear de $\Re^M \rightarrow \Re^K$ que mede o quanto o desenrolar temporal momentâneo da série S se relaciona com os K estados básicos do processo à ela associado.

Para um vetor de pesos sinápticos arbitrário $\underline{w} = \underline{w}_a$, o conjunto de saídas ou o vetor de saídas $\underline{y}(n)$ para $\boldsymbol{\Phi}(n)$ dado é obtido por

$$\underline{y}(n) = \boldsymbol{\Phi}(n).\underline{w}_a \tag{4.23}$$

com

$$\underline{y}(n) = [y(n-1)\ y(n-2) \cdots\ y(n-K)]^T \tag{4.24}$$

onde $y(n-1), y(n-2), \cdots, y(n-K)$ são as saídas da rede RBF com respeito aos vetores de entrada $\underline{u}(n-1)$, $\underline{u}(n-2), \cdots$, $\underline{u}(n-K)$ associados ao desenrolar temporal momentâneo da série S, sendo dados o vetor de pesos sinápticos arbitrário $\underline{w} = \underline{w}_a$ e a matriz de estados $\boldsymbol{\Phi}(n)$.

Vamos supor que $\underline{y}(n) = \underline{d}(n)$, onde $\underline{d}(n)$ é o vetor de saídas desejadas definido por $\underline{d}(n) = [u(n)\ u(n-1) \cdots\ u(n-K+1)]^T$ conforme construção mostrada na Figura 4.4, de tal forma que

$$\underline{y}(n) = \underline{d}(n) = \begin{bmatrix} u(n) \\ u(n-1) \\ \vdots \\ u(n-K+1) \end{bmatrix} = \Phi(n).\underline{w}(n) = \tag{4.25}$$

$$= \begin{bmatrix} \varphi_0(n-1) & \varphi_1(n-1) & \cdots & \varphi_{K-1}(n-1) \\ \varphi_0(n-2) & \varphi_1(n-2) & \cdots & \varphi_{K-1}(n-2) \\ \vdots & \vdots & & \vdots \\ \varphi_0(n-K) & \varphi_1(n-K) & \cdots & \varphi_{K-1}(n-K) \end{bmatrix} \begin{bmatrix} w_0(n) \\ w_1(n) \\ \vdots \\ w_{K-1}(n) \end{bmatrix}$$

Observe-se que cada elemento em $\underline{d}(n)$ é o elemento que se coloca uma posição à frente na série S, com respeito ao vetor de entrada que gerou o correspondente vetor de transição de estado não-linear em $\boldsymbol{\Phi}$. Por exemplo, a primeira linha de $\boldsymbol{\Phi}$ na Equação

(4.25) é o vetor de transição de estado não-linear $\underline{\varphi}(n-1)^T$ que resulta da aplicação do vetor $\underline{u}(n-1)$ à entrada da rede neural. A saída desejada correspondente a este vetor de entrada é o elemento $u(n)$ de $\underline{d}(n)$, que se encontra uma posição à frente na série S, com respeito ao vetor $\underline{u}(n-1)$, conforme pode ser observado na Figura 4.4.

Portanto, através da transformação linear definida pelo vetor de pesos sinápticos $\underline{w}(n) \in \Re^K$ e através da transformação não-linear definida pela matriz $\mathbf{\Phi}(n)$, a Equação (4.25) implicitamente relaciona cada vetor $\underline{u}(n-k-1)$ formado da janela $p(n)$, $k = 0,1,\cdots,K-1$, com o elemento $u(n-k)$ de S, sendo $u(n-k)$ o elemento que está localizado na série S uma posição à frente do vetor $\underline{u}(n-k-1)$. Assim, uma vez determinado, $\underline{w}(n)$ conterá informação de como se efetua a transição partindo dos estados prévios do processo de S até o próximo elemento de S imediatamente adiante ao respectivo vetor $\underline{u}(n-k-1)$ representativo do desenrolar temporal momentâneo de S. É importante ressaltar que a informação de transição em $\underline{w}(n)$ é resultante de uma transformação não-linear $\Re^M \rightarrow \Re^K$, e, em conseqüência disto, é uma informação que envolve as estatísticas de ordem superior do processo de S. Em função disto, o método de predição não-linear aqui apresentado é potencialmente mais capaz de captar as "sutilezas estatísticas" do processo estocástico subjacente em S do que o método de predição linear baseado em estatísticas de segunda ordem visto no Capítulo 3. A obtenção de $\underline{w}(n)$ é definida pela Equação (4.26).

$$\underline{w}(n) = \mathbf{\Phi}^{-1}(n)\, \underline{d}(n) \tag{4.26}$$

Nesta tese, a inversa da matriz $\mathbf{\Phi}$ é obtida pela pseudo-inversão matricial de Moore-Penrose [50], através de decomposição em valores singulares – SVD – conforme descrito no Apêndice A. Embora a SVD minimize o problema da eventual singularidade de $\mathbf{\Phi}$ e, embora toda a série temporal seja normalizada para o intervalo [–1,1], antes de qualquer procedimento, por precaução, adiciona-se o valor 1×10^{-9} à diagonal principal de

Φ, como um parâmetro de regularização [50], visando auxiliar o tratamento das singularidades da matriz Φ.

Vamos agora efetuar o processo de predição propriamente dito. Uma vez obtido $\underline{w}(n)$ de (4.26), aplica-se o vetor $\underline{u}(n)=[u(n),\ u(n-1),\cdots,\ u(n-M+1)]^T$ à entrada da rede neural. O vetor peso sináptico $\underline{w}(n)$ obtido de (4.26) armazena informação de como ocorre a transição "estados prévios→próximo elemento" dado o vetor de entrada que descreve o desenrolar temporal momentâneo da série S. Como, por definição, uma variável de estado não sofre alteração para uma variação pequena no sistema por ela descrito [8], assume-se que os vetores de estado $\underline{t}_k$ do processo de S não sofram uma mudança significativa uma posição à frente em S. Assim, a saída da rede neural $y(n)$ ao vetor de entrada $\underline{u}(n)=[u(n)\ \ u(n-1)\cdots\ u(n-M+1)]^T$ será uma estimativa $\hat{u}(n+1)$ ou predição da amostra $u(n+1)$, dada pela Equação (4.27) com base em (4.15):

$$\hat{u}(n+1)=y(n)=\sum_{k=0}^{K-1}w_k(n)\exp\left\{\frac{-\left\|\underline{u}(n)-\underline{t}_k(n)\right\|^2}{\xi\max\left\{\left\|\underline{t}_i(n)-\underline{t}_j(n)\right\|^2\right\}}\right\}= \tag{4.27}$$

$$=\sum_{k=0}^{K-1}w_k(n)\varphi_k(n)=\underline{w}^T(n)\underline{\varphi}(n)$$

Em outras palavras, se está implicitamente deslizando a matriz Φ uma posição à frente ao longo de S, assumindo que os estados armazenados em Φ permanecem inalterados e usando-se a informação de transição armazenada em $\underline{w}$ para estimar o próximo elemento em S, a partir dos estados definidos em Φ. Obviamente, K deve ser grande o suficiente para que Φ possa armazenar todos os estados significativos. Da mesma forma, a dimensão M dos vetores de estado do processo de S deve ser suficientemente grande para representar os estados significativos do processo.

Esta técnica de predição não-linear através de redes neurais artificiais do tipo RBF com atribuição de centros definida pela Equação (4.18) doravante será referida como predição com base na Atribuição Padrão dos Centros (APC).

4.2.1 Resultados Experimentais

Nesta seção são apresentados os resultados obtidos para a predição não-linear através da heurística APC da série *Chaotic_LASER*, com $N_t = 1000$ amostras, descrita na Seção 2.2.

Foi adotada ordem de predição $M = 11$ e $K = 8$ vetores centros, necessitando portanto, de $N = K + M = 19$ amostras conhecidas, anteriores à amostra $u(n+1)$ que se deseja predizer. Para o Fator de Variância utilizou-se $\xi = 1.0$. O valor obtido para o Erro Médio Quadrático Normalizado final, conforme Equação (3.21), é $\text{NMSE}(N_t - 1) = 0.096$.

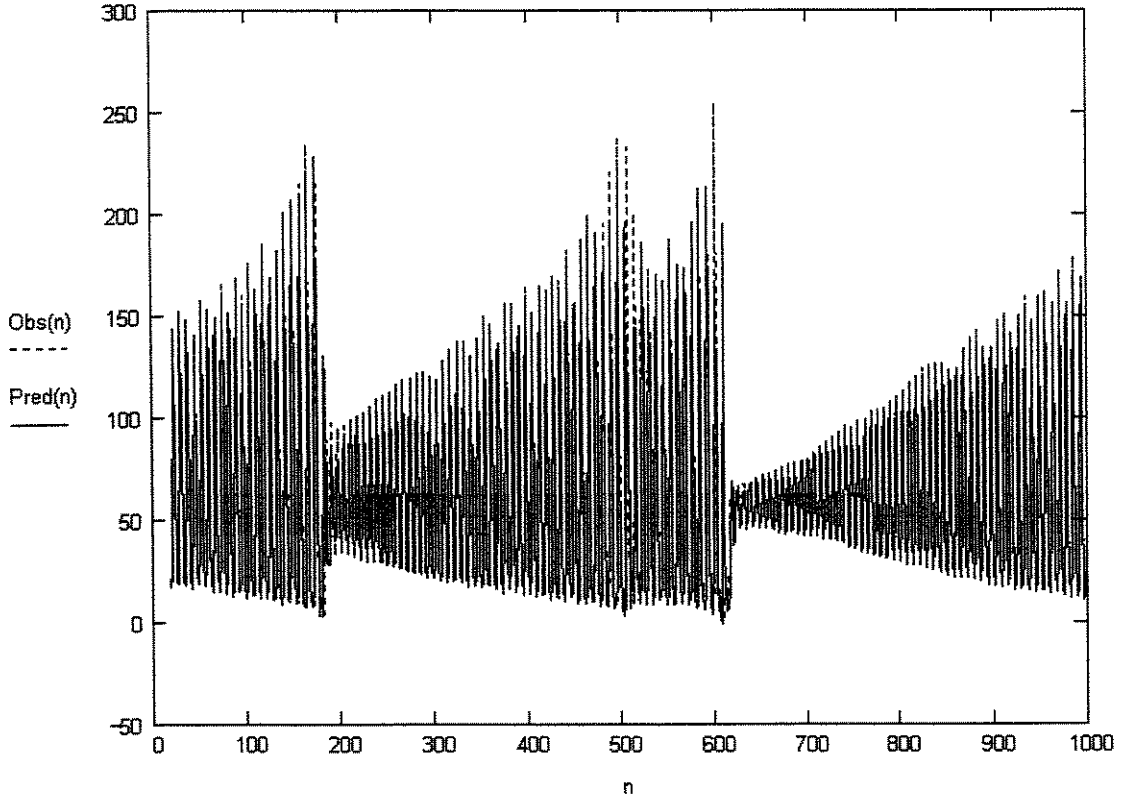

Figura 4.5: Série *Chaotic_LASER* – Predição não-linear através da heurística APC com $M = 11$, $K = 8$ e $N = 19$. $\text{NMSE}(N_t - 1) = 0.096$.

Para efeito de comparação do desempenho de predição da técnica APC, utilizaremos a predição linear descrita no Capítulo 3. Por coerência com a ordem de

predição utilizada na predição por APC, a estimativa predita $\hat{u}(n+1)$ é expressa como uma combinação linear de 11 amostras prévias, isto é, com ordem de predição $M=11$. Os valores para os coeficientes que ponderam tal combinação linear, obtidos através de (3.19), são $W_0 = -0.689614$, $W_1 = 0.545837$, $W_2 = -0.206001$, $W_3 = 0.144179$, $W_4 = -0.109438$, $W_5 = 0.054155$, $W_6 = -0.402716$, $W_7 = -0.334968$, $W_8 = 0.221793$, $W_9 = -0.248085$ e $W_{10} = 0.0365866$. Utilizando estes coeficientes no filtro FIR da Figura 3.1, o NMSE final resultante, conforme Equação (3.21), é $\text{NMSE}(N_t - 1) = 0.213$. A Figura 3.2 apresenta o resultado da predição linear para a série *Chaotic_LASER*.

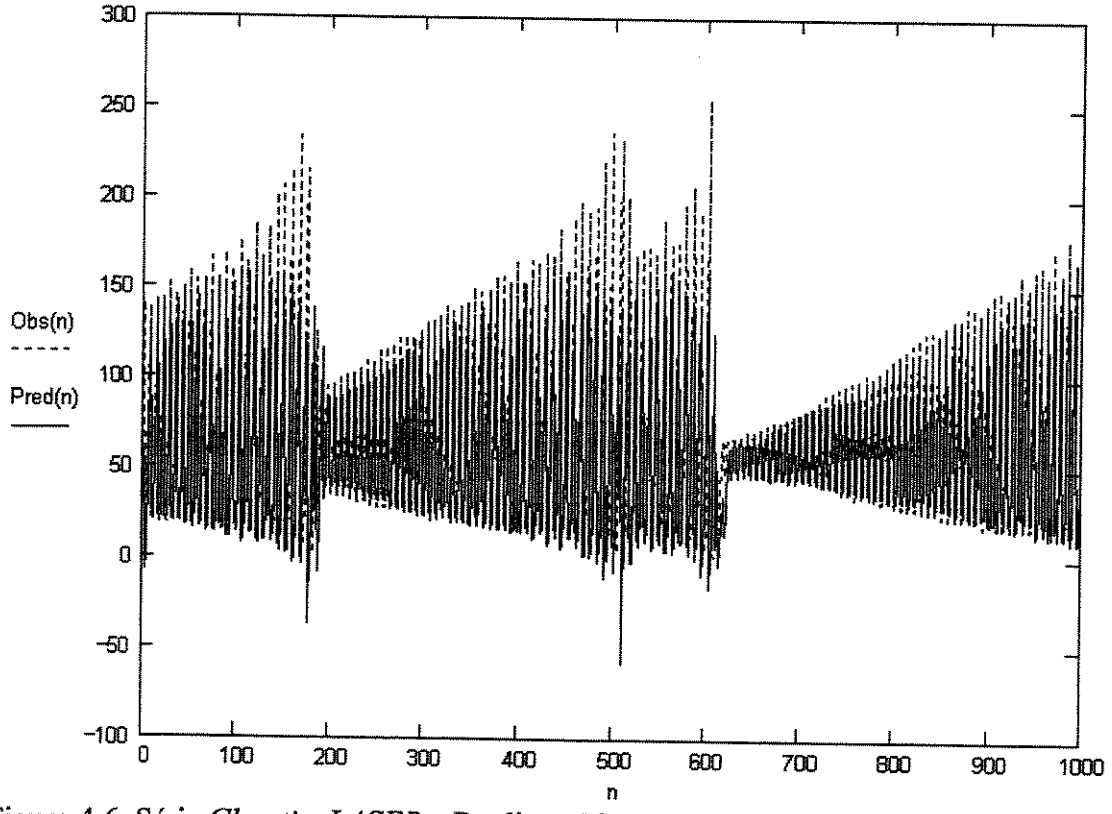

Figura 4.6: Série *Chaotic_LASER* – Predição Linear – $M=11$. $\text{NMSE}(N_t - 1) = 0.213$

Experimentou-se a predição linear da série *Chaotic_LASER* com o dobro da ordem de predição utilizada na predição por APC, isto é, $M = 22$.

Os valores dos coeficientes obtidos através de (3.19) são $W_0 = -0.670762$, $W_1 = 0.545837$, $W_2 = -0.253092$, $W_3 = 0.198692$, $W_4 = -0.0704754$, $W_5 = 0.129263$, $W_6 = -0.355902$, $W_7 = -0.510382$, $W_8 = 0.277562$, $W_9 = -0.30334$, $W_{10} = 0.217955$, $W_{11} = -0.161362$, $W_{12} = 0.092362$, $W_{13} = -0.189575$, $W_{14} = 0.0408657$, $W_{15} = 0.162011$, $W_{16} = -0.109375$, $W_{17} = 0.0276988$, $W_{18} = -0.0969306$, $W_{19} = 0.0260158$, $W_{20} = -0.0871898$ e $W_{21} = 0.101258$.

O resultado da predição linear para este caso é mostrado na Figura 4.7 e o NMSE final resultante, conforme Equação (3.21), é $\text{NMSE}(N_t - 1) = 0.185$.

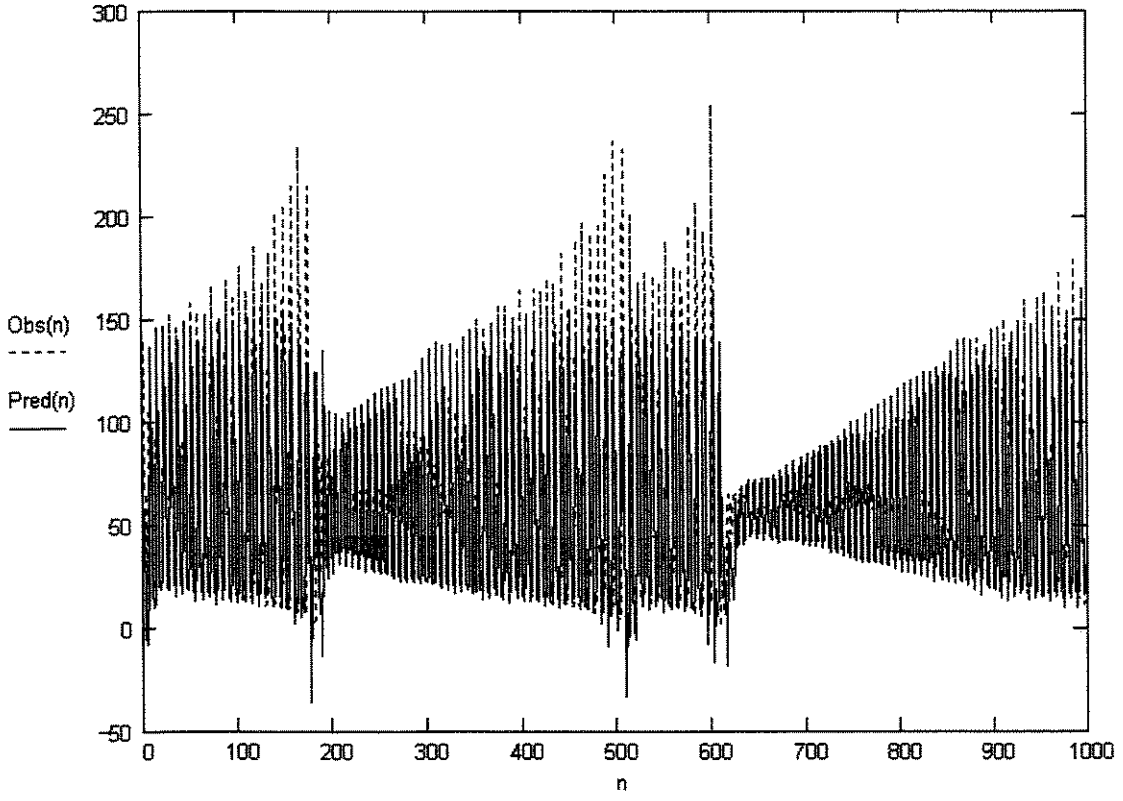

Figura 4.7: Série *Chaotic_LASER* – Predição Linear – $M = 22$. $\text{NMSE}(N_t - 1) = 0.185$.

Observe que, mesmo com o dobro da ordem de predição utilizada na heurística APC, a predição linear resulta em um NMSE final quase duas vezes maior. A diferença de performance entre a predição linear e a heurística não-linear APC fica ainda mais evidente

se lembrarmos que a heurística APC necessita, neste caso, para efetuar a predição, de apenas $N = 19$ amostras prévias conhecidas, contidas na janela de predição $p(n)$. Já a predição linear precisa conhecer todas as $N_t = 1000$ amostras da série *Chaotic_LASER* para montar a matriz de correlação $\mathbf{R}$, expressa em (3.13). Neste sentido, para uma ordem de predição $M = 11$, pode-se dizer que a heurística APC obtém um NMSE final 0.096 com uma janela de predição com apenas 19 amostras prévias conhecidas, enquanto que a predição linear obtém um NMSE final de 0.213 necessitando, para isso, uma janela de predição equivalente às 1000 amostras prévias conhecidas – todas as amostras da série.

Capítulo 5

Predição Não-Linear de Séries Temporais por Decomposição em Componentes Principais

No capítulo anterior foi apresentada a técnica clássica APC de predição não-linear através de redes neurais artificiais do tipo RBF, na qual a série temporal S é vista como um processo não-linear que possui uma matriz de transição de estados $\mathbf{\Phi}$ associada. Esta matriz de transição de estados é representada pela rede neural RBF, a qual é interpretada como um filtro não-linear com matriz de transição definida por $\mathbf{\Phi}$.

Na técnica APC, a matriz de estado $\mathbf{\Phi}$ armazena informação de transição com relação aos estados básicos do processo associado ao desenrolar temporal de S, tendo como referência o conjunto de vetores $\underline{t}_k$ centro das funções de base radial, $k = 0,1,\cdots,K-1$, vetores que definem os K estados do processo. Os centros das funções de base radial são vetores formados a partir do conjunto $\mathbf{U}$ de vetores em $\Re^M$, M é a ordem de predição, tomados seqüencialmente da janela de predição $p(n)$ na série temporal em progressão. A técnica de predição APC, portanto, não aplica nenhuma transformação ao conjunto $\mathbf{U}$ de vetores formados de $p(n)$ a serem atribuídos aos respectivos centros, no sentido de explorar as características estocásticas de $\mathbf{U}$.

No entanto, é desejável que os vetores de estado $\underline{t}_k$ detenham o máximo de informação significativa sobre o comportamento histórico do processo associado à S, de forma a representar o desenrolar temporal de S com a maior fidelidade possível.

Uma possível abordagem neste sentido é explorar as regiões de maior freqüência de ocorrência dos vetores de $\mathbf{U}$ e sua localização no espaço $\Re^M$. Para este fim, obtém-se K vetores $\underline{t}_k$ tal que definam as coordenadas centrais de K agrupamentos de vetores, cada agrupamento formado por vetores com coordenadas semelhantes no conjunto $\mathbf{U}$. Em

outras palavras, se considerarmos que a ponta de cada vetor de $\mathbf{U}$ define as coordenadas de um ponto em $\Re^M$, os K vetores $\underline{t}_k$ definem as coordenadas centrais de K nuvens de pontos em $\Re^M$, sendo a k-ésima nuvem determinada pelo k-ésimo agrupamento de vetores com coordenadas semelhantes no conjunto $\mathbf{U}$. A idéia aqui é obter as coordenadas centrais das nuvens de pontos em $\Re^M$, cada nuvem definida pelo agrupamento de vetores em $\mathbf{U}$ formado por vetores que ocorrem mais freqüentemente ao longo do desenrolar da série S e simultaneamente cujas coordenadas sejam aproximadamente semelhantes às do centro do respectivo agrupamento.

Um possível critério para a determinação dos centros das nuvens de pontos em $\Re^M$ associadas à distribuição dos vetores de $\mathbf{U}$ é o algoritmo *k-means*, já descrito no capítulo anterior. O algoritmo *k-means* computa K coordenadas, cada uma delas definindo o centro de uma região no espaço $\Re^M$ na qual os vetores de $\mathbf{U}$ apresentam alta freqüência de ocorrência e alta semelhança entre suas coordenadas. Uma tal região é denominada de *cluster* [11]. O número K de *clusters* é determinado por experimentação, por estar estreitamente associado à natureza dos dados. Ainda que esta técnica obtenha as coordenadas centrais dos *clusters* de vetores de $\mathbf{U}$ que ocorrem mais freqüentemente na janela $p(n)$, ela não utiliza informação de correlação a respeito do processo estocástico subjacente associado à S. Isto ocorre porque o algoritmo *k-means* busca os centros de forma localizada na janela $p(n)$, sem levar em consideração os modos de variação próprios do processo estocástico subjacente [34][37], isto é, sem levar em consideração as causas "escondidas" no processo subjacente que resultaram na ocorrência de um particular conjunto de vetores obtidos de $p(n)$.

A nova técnica de determinação dos centros $\underline{t}_k$ por decomposição do espaço de dados em componentes principais – ou, conforme denominamos, Decomposição em Sub-Espaços (DSE) – proposta nesta tese possibilita que os estados representados pelos centros $\underline{t}_k$ detenham o máximo de informação significativa sobre o comportamento histórico do processo associado à S, de forma que os estados assim obtidos refiram-se ao processo como

um todo e não mais de forma localizada na janela $p(n)$, como acontece com o algoritmo *k-means*, ou como acontece com a técnica APC.

Neste capítulo é descrita a técnica de predição não-linear com base na decomposição em sub-espaços. Inicialmente é apresentada a técnica para obtenção das coordenadas dos centros de conjuntos de vetores por DSE, através da KLT. Após a explicação teórica é apresentada uma comparação entre a determinação dos centros via algoritmo *k-means* e via DSE. Na seqüência do capítulo é desenvolvida a predição não-linear propriamente dita e é apresentado o critério adaptativo para atualização do fator de variância. Finalizando o capítulo, é apresentada uma possível solução para o problema de encontrar a arquitetura da rede RBF mais apta a acompanhar não-estacionariedades presentes em muitas séries temporais. As diversas etapas do algoritmo são acompanhadas de exemplos ilustrativos das técnicas descritas.

5.1 Atribuição dos Centros de Redes RBF para Predição Não-Linear através de Decomposição em Componentes Principais

Em trabalhos prévios [34][35][37] estudou-se a Transformada Karhunen-Loève (KLT) para análise dos componentes principais, ou sub-espaços, de um conjunto de dados. O Apêndice B desta tese descreve a KLT no contexto de Decomposição em Sub-Espaços.

A decomposição em sub-espaços provisionada pela KLT consiste em obter o conjunto de M auto-vetores e auto-valores da matriz de covariância $\mathbf{C}$ do conjunto $\mathbf{U}$ de média zero, composto por L vetores de dados $\underline{u}_i \in \Re^M, i = 0,1,\cdots,L-1$. A restrição de que o conjunto $\mathbf{U}$ apresente média vetorial $\underline{0} \in \Re^M$ é transparente a nível de procedimento, já que o vetor média pode ser restaurado ao final. O conjunto de M auto-vetores $\underline{e}_m \in \Re^M$ obtido pela KLT, $m = 0,1,\cdots,M-1$, define uma base de vetores ortonormais em $\Re^M$ e,

portanto, define um conjunto de M eixos ortogonais Cartesianos. A coordenada de origem deste sistema Cartesiano é a coordenada de origem dos vetores da base ortonormal definida pelo conjunto de M auto-vetores $\underline{e}_m$. Como a média do conjunto $\mathbf{U}$ é assumida zero, a coordenada de origem do sistema Cartesiano é $\underline{0} \in \Re^M$. Ao projetar o conjunto original de vetores $\underline{u}_i \in \Re^M$ sobre o m-ésimo eixo Cartesiano, a variância do conjunto projetado – ou variância do sub-espaço – é dada pelo m-ésimo auto-valor λ_m. Ainda, a variância do conjunto $\mathbf{U}$ é igual à soma das variâncias das projeções de $\mathbf{U}$ [37]. Isto é, a média do quadrado da norma Euclidiana dos vetores de $\mathbf{U}$ é igual à soma dos M auto-valores λ_m, onde λ_m equivale à média do quadrado da norma Euclidiana da projeção dos vetores $\underline{u}_i \in \Re^M$ sobre o m-ésimo eixo.

A título de interpretação da KLT, se quiséssemos obter a KLT através de um processo manual e experimental, tomaríamos um vetor arbitrário de módulo unitário $\underline{e} \in \Re^M$ com origem em $\underline{0} \in \Re^M$, o qual definiria uma direção arbitrária em que o conjunto $\mathbf{U}$ de vetores $\underline{u}_i \in \Re^M$ de dados seria projetado. Projetaríamos, então, a totalidade do conjunto $\mathbf{U}$ sobre a direção dada por $\underline{e}$ e mediríamos a variância λ da projeção. Após tentarmos todas as direções possíveis no espaço $\Re^M$, haverá uma direção dada por $\underline{e}$ na qual é obtida a maior variância λ_0. O vetor $\underline{e}$ que define tal direção é igual ao auto-vetor $\underline{e}_0$ associado ao maior auto-valor, obtidos pela KLT, com valor do maior auto-valor dado por λ_0. O processo é repetido novamente para a obtenção do segundo maior auto-valor λ_1, com a restrição de que a busca da direção de maior variância em $\Re^M$ seja feita em direções ortogonais à do auto-vetor $\underline{e}_0$ associado ao maior auto-valor λ_0, recém determinados. A busca da direção de maior variância em $\Re^M$ para obtenção do terceiro maior auto-valor λ_2 é feita com a restrição de que as direções testadas sejam ortogonais às direções dos dois auto-vetores $\underline{e}_0$ e $\underline{e}_1$ associados aos maiores auto-valores λ_0 e λ_1 previamente encontrados. E assim prosseguiríamos neste processo recursivo até que os M auto-valores e auto-vetores fossem determinados.

Assim, como os sub-espaços obtidos pela KLT estão alinhados com as direções ortogonais de maior variância possível no espaço original $\Re^M$, a KLT é considerada uma transformação ótima no sentido do erro médio quadrático MSE [37][38][51] para efeito de reconstrução do espaço original $\Re^M$ a partir de suas M projeções ou componentes principais. Ou seja, o conjunto de M sub-espaços representa de maneira ótima, no sentido do MSE, o conjunto $\mathbf{U}$ de L vetores $\underline{u}_i \in \Re^M$, $i = 0,1,\cdots,L-1$. Como vimos, os sub-espaços obtidos pela KLT encontram-se alinhados com os eixos Cartesianos que definem as direções de maior variância possíveis no espaço original $\Re^M$. Em conseqüência, isto resulta em uma maior concentração de pontos definidos pelos vetores $\underline{u}_i \in \Re^M$ do espaço original nas vizinhanças dos eixos Cartesianos que definem cada sub-espaço.

Como o m-ésimo eixo Cartesiano define a m-ésima região em $\Re^M$ de maior variância λ_m, aqueles vetores cuja média do quadrado de suas normas Euclidianas é próxima ao valor λ_m (λ_m é a média do quadrado das normas Euclidianas das projeções destes vetores sobre o m-ésimo eixo) caracterizarão um sub-conjunto de vetores do espaço $\Re^M$ aproximadamente alinhados com o m-ésimo eixo.

Parte destes vetores estará aproximadamente congruente (*i.e.*, alinhados no mesmo sentido) com o m-ésimo semi-eixo positivo, e parte dos vetores estará aproximadamente congruente com o m-ésimo semi-eixo negativo. Mas, independentemente do sentido positivo ou negativo, a média do quadrado da norma Euclidiana destes vetores será, em maior ou menor grau, próxima ao valor λ_m, grau que depende do quanto os vetores alinham-se com o m-ésimo eixo Cartesiano. Adicionalmente, a média dos vetores que são congruentes com o semi-eixo negativo tende a ser igual à média dos vetores que são congruentes com o semi-eixo positivo, já que $\mathbf{U}$ apresenta média vetorial $\underline{0} \in \Re^M$. Assim, deve-se esperar um certo equilíbrio entre os vetores alinhados com cada um dos dois semi-eixos.

Portanto, o m-ésimo sub-espaço identifica uma nuvem de pontos em $\Re^M$ nas vizinhanças do m-ésimo eixo Cartesiano, com coordenada de cada ponto definida pelo respectivo vetor $\underline{u}_i \in \Re^M$ do conjunto $\mathbf{U}$ nas vizinhanças do m-ésimo eixo. Como a média do quadrado da norma Euclidiana dos vetores associados a estes pontos tende em maior ou menor grau para λ_m, a norma – ou distância – Euclidiana média destes pontos à origem aproxima-se do valor $\sqrt{\lambda_m}$.

Ao projetar a totalidade dos vetores $\underline{u}_i \in \Re^M$ do conjunto $\mathbf{U}$ sobre o m-ésimo eixo Cartesiano definido por $\underline{e}_m$, a variância do conjunto projetado, ou variância do sub-espaço, é dada pelo m-ésimo auto-valor λ_m. Conforme visto, haverá uma nuvem de pontos nas vizinhanças do m-ésimo eixo Cartesiano, com distância Euclidiana média destes pontos à origem aproximada por $\sqrt{\lambda_m}$. As coordenadas do centro geométrico dos pontos da m-ésima nuvem é a origem $\underline{0} \in \Re^M$, mas, como existe equilíbrio entre os pontos projetados nos dois semi-eixos Cartesianos (por $\mathbf{U}$ apresentar média vetorial $\underline{0} \in \Re^M$), isto implica que a m-ésima nuvem resulta em dois *clusters*, *i.e.*, duas regiões de maior concentração dos pontos que se desviam da origem – uma com coordenada central dada por $\underline{\psi}_m = \underline{e}_m \cdot \sqrt{\lambda_m}$ e a outra com coordenada central dada por $\underline{\psi}_m = -\underline{e}_m \cdot \sqrt{\lambda_m}$.

Por exemplo, seja o conjunto $\mathbf{U}$ de $L = 6$ vetores $\underline{u}_i \in \Re^2$, $i = 0,1,\cdots,L-1$, de média zero definido por $\mathbf{U} = \left\{ \begin{bmatrix} 0.425 \\ -1.063 \end{bmatrix}, \begin{bmatrix} 0.595 \\ -0.943 \end{bmatrix}, \begin{bmatrix} 0.327 \\ -0.843 \end{bmatrix}, \begin{bmatrix} -0.497 \\ 0.820 \end{bmatrix}, \begin{bmatrix} -0.491 \\ 1.133 \end{bmatrix}, \begin{bmatrix} -0.359 \\ 0.897 \end{bmatrix} \right\}$. Ao aplicar a KLT sobre $\mathbf{U}$, os auto-valores resultam em $\lambda_0 = 1.118$ e $\lambda_1 = 7.035 \times 10^{-3}$, com auto-vetores associados dados por $\underline{e}_0 = \begin{bmatrix} -0.427 \\ 0.904 \end{bmatrix}$ e $\underline{e}_1 = \begin{bmatrix} 0.904 \\ 0.427 \end{bmatrix}$. A Figura 5.1 mostra o conjunto $\mathbf{U}$, os $M = 2$ auto-vetores escalonados $\underline{\psi}_m = \underline{e}_m \cdot \sqrt{\lambda_m}$, $m = 0,1,\cdots,M-1$ e os eixos Cartesianos por eles definidos.

Note que a variância do conjunto **U** é $\frac{1}{L}\sum_{i=0}^{L-1}\|\underline{u}_i\|^2 = 1.125 = \lambda_0 + \lambda_1$, onde $\|\underline{u}\| = \sqrt{\sum_{m=0}^{M-1}(u_m)^2} = \sqrt{\underline{u}^T \cdot \underline{u}}$ é a norma Euclidiana do vetor $\underline{u} \in \Re^M$ [8].

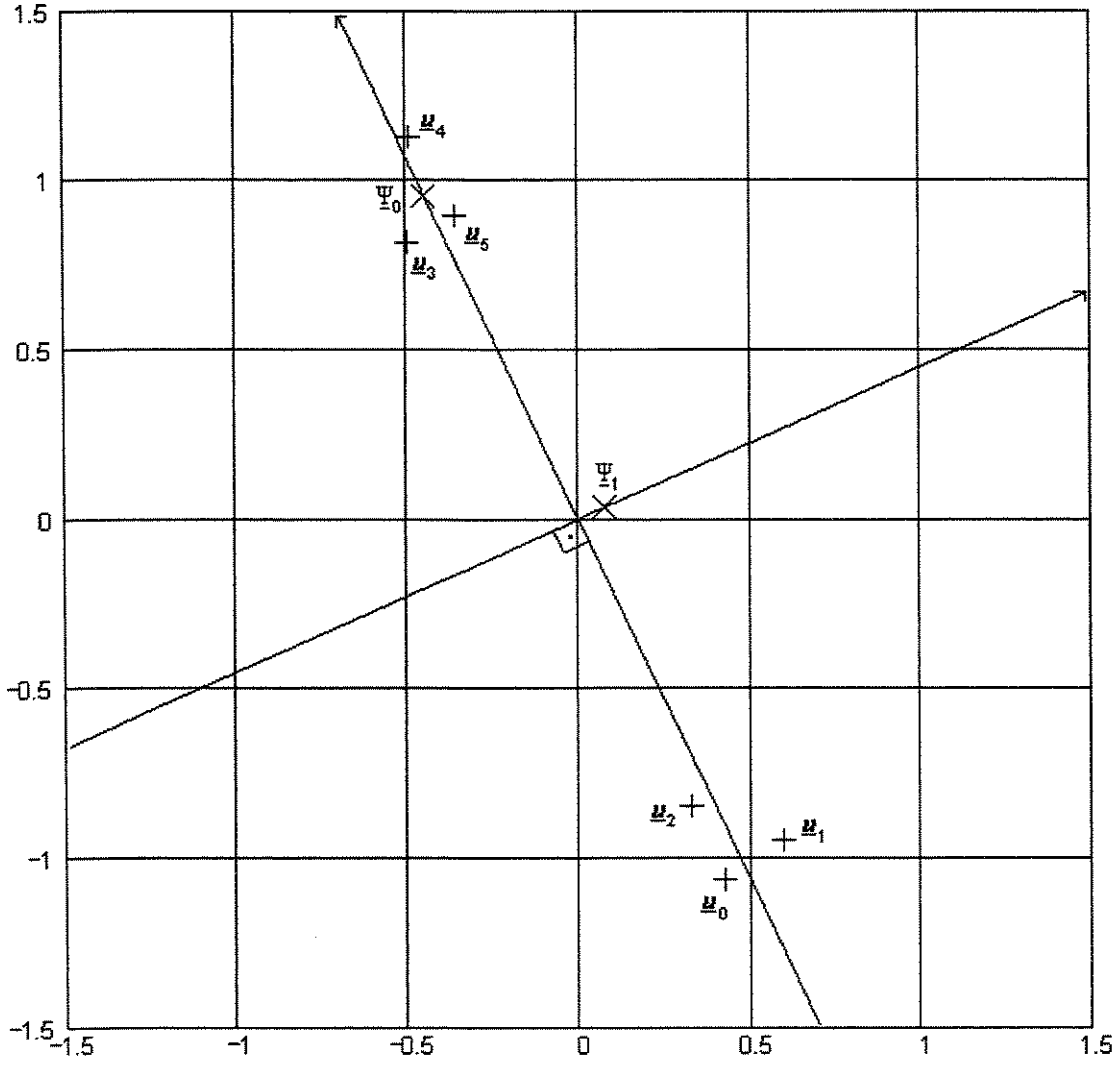

Figura 5.1: Conjunto **U**, auto-vetores escalonados $\underline{\psi}_0$ e $\underline{\psi}_1$ e os eixos Cartesianos por eles definidos.

A Tabela 5.1 mostra a valor da projeção de cada vetor $\underline{u}_i \in \Re^2$ do conjunto **U** sobre os eixos Cartesianos definidos por $\underline{e}_0$ e $\underline{e}_1$.

i	$\underline{u}_i^T \cdot \underline{e}_0$	$\underline{u}_i^T \cdot \underline{e}_1$
0	-1.142	-0.07
1	-1.107	0.135
2	-0.902	-0.065
3	0.953	-0.099
4	1.234	0.04
5	0.964	0.059

Tabela 5.1: Projeções de **U** sobre os eixos Cartesianos definidos por $\underline{e}_0$ e $\underline{e}_1$.

Observe que, como os auto-vetores $\underline{e}_0$ e $\underline{e}_1$ têm norma unitária e são adimensionais [8], o valor absoluto da projeção de cada $\underline{u}_i \in \Re^2$ sobre cada eixo define a norma Euclidiana da i-ésima projeção. Note ainda que as variâncias das projeções de **U** resultam em $\frac{1}{L}\sum_{i=0}^{L-1}\left(\underline{u}_i^T \cdot \underline{e}_0\right)^2 = 1.118 = \lambda_0$ e que $\frac{1}{L}\sum_{i=0}^{L-1}\left(\underline{u}_i^T \cdot \underline{e}_1\right)^2 = 7.035\times10^{-3} = \lambda_1$. Portanto, a variância do conjunto projetado, ou variância do sub-espaço, é dada pelo m-ésimo auto-valor λ_m. Ainda, a distância Euclidiana média dos vetores da m-ésima projeção à origem, $\frac{1}{L}\sum_{i=0}^{L-1}\left|\underline{u}_i^T \cdot \underline{e}_m\right|$, pode ser aproximada por $\sqrt{\frac{1}{L}\sum_{i=0}^{L-1}\left(\underline{u}_i^T \cdot \underline{e}_m\right)^2} = \sqrt{\lambda_m}$. No caso, $\frac{1}{L}\sum_{i=0}^{L-1}\left|\underline{u}_i^T \cdot \underline{e}_0\right| = 1.051$ para $\sqrt{\lambda_0} = 1.057$ e $\frac{1}{L}\sum_{i=0}^{L-1}\left|\underline{u}_i^T \cdot \underline{e}_1\right| = 0.078$ para $\sqrt{\lambda_1} = 0.083$.

Na técnica de predição não-linear via DSE apresentada nesta tese, o conjunto de vetores centro $\underline{t}_k$ das funções de base radial da rede RBF é obtido dos auto-vetores escalonados $\underline{\psi}_m = \underline{e}_m \cdot \sqrt{\lambda_m}$, isto é $\underline{t}_k = \underline{\psi}_k$. Os centros $\underline{t}'_k = \underline{\psi}'_k$, simétricos aos centros $\underline{t}_k = \underline{\psi}_k$, que seriam obtidos das coordenadas dadas por $\underline{\psi}'_m = -\underline{e}_m \cdot \sqrt{\lambda_m}$ não são considerados. Tal procedimento – confirmado por resultados experimentais – é justificado quando se insere a DSE no contexto temporal de predição. Primeiramente, note que os

centros $\underline{\psi}_k$ assim determinados resultam dos auto-vetores e auto-valores da matriz de covariância **C** do conjunto **U** de vetores. No contexto de predição do elemento $u(n+1)$ da série temporal S, os vetores $\underline{u}_i \in \Re^M$ de **U** são obtidos da janela de predição $p(n)$. Portanto, a matriz de covariância **C** acumula informação de correlação temporal entre os vetores $\underline{u}_i \in \Re^M$. Desta maneira, cada sub-espaço de **U** representa uma nuvem de pontos em $\Re^M$ com alta correlação temporal entre si. Se considerarmos também como centros aqueles definidos por $\underline{t}'_k = \underline{\psi}'_k$, além dos definidos por $\underline{t}_k = \underline{\psi}_k$, estaremos definindo duas funções de base radial para caracterizar a mesma região de alta correlação em $\Re^M$. Portanto estaríamos usando duas bases, que no contexto de extrapolação por redes RBF deveriam representar duas classes independentes, para representar pontos em $\Re^M$ de "mesma espécie" sob o ponto de vista de correlação temporal. Experimentalmente, muito embora o erro de aproximação em $p(n)$ algumas vezes tenha sido reduzido, a grande maioria das vezes em que se usou os centros $\underline{t}'_k = \underline{\psi}'_k$ além dos centros $\underline{t}_k = \underline{\psi}_k$, o erro de predição permaneceu praticamente inalterado ou aumentou. Até porque, como veremos a seguir, quando consideramos as coordenadas $\underline{\psi}_k$ no espaço $\Re^M$, e não somente no m-ésimo sub-espaço unidimensional, elas acabam por definir centros de nuvens de pontos sob um posto de vista global em $\Re^M$.

Sejam os M auto-valores do conjunto de vetores **U** definidos por $\{\lambda_0, \lambda_1, \cdots, \lambda_p, \cdots, \lambda_{M-1}\}$ com respectivos auto-vetores definidos por $\{\underline{e}_0, \underline{e}_1, \cdots, \underline{e}_p, \cdots, \underline{e}_{M-1}\}$, $\underline{e}_m \in \Re^M$, $m = 0, 1, \cdots, M-1$, tal que $\lambda_0 > \lambda_1 > \cdots > \lambda_{p-1} > \lambda_p > \cdots > \lambda_{M-2} > \lambda_{M-1}$. Sejam duas nuvens quaisquer a_p e a_{p-1} de pontos em $\Re^M$ que se concentram nas vizinhanças das direções definidas respectivamente por $\underline{e}_p$ e $\underline{e}_{p-1}$. Ambas as nuvens possuem centro geométrico na origem $\underline{0} \in \Re^M$. A largura da região ocupada em $\Re^M$ pela nuvem a_{p-1} é $2\sqrt{\lambda_{p-1}}$ que é maior que a largura $2\sqrt{\lambda_p}$ da região ocupada em $\Re^M$ pela nuvem a_p. A coordenada $\underline{\psi}_{p-1} = \underline{e}_{p-1} \cdot \sqrt{\lambda_{p-1}}$ define o centro de um dos dois *clusters* de maior

concentração de pontos da nuvem a_{p-1}. Cada *cluster* da nuvem a_{p-1} tem seus pontos concentrados respectivamente nas vizinhanças dos semi-eixos positivo e negativo pertencentes ao eixo definido por $\underline{e}_{p-1}$. O vetor $\underline{\psi}_p = \underline{e}_p \cdot \sqrt{\lambda_p}$ é ortogonal à direção ao longo da qual a nuvem a_{p-1} se espalha e a distância Euclidiana $\sqrt{\lambda_p}$, medida da coordenada $\underline{\psi}_p$ até a origem $\underline{0} \in \Re^M$, é menor do que a largura $2\sqrt{\lambda_{p-1}}$ da região ocupada em $\Re^M$ pela nuvem a_{p-1}. Portanto, as coordenadas do ponto dado por $\underline{\psi}_p$ não só distanciam-se igualmente dos centros dos dois *clusters* da nuvem a_{p-1}, como encontram-se próximas ao centro geométrico da nuvem a_{p-1}.

Assim, as coordenadas $\underline{\psi}_p$ podem ser consideradas como um centro em $\Re^M$ para os pontos da nuvem a_{p-1}, a menos de um pequeno desvio à origem $\underline{0} \in \Re^M$ dado por $\sqrt{\lambda_p}$ – pequeno com relação à largura $2\sqrt{\lambda_{p-1}}$ da nuvem a_{p-1}. Tanto mais essa interpretação será válida quanto maior for λ_{p-1} com relação a λ_p. Note que, sob esta definição, a coordenada $\underline{\psi}_0 = \underline{e}_0 \cdot \sqrt{\lambda_0}$ poderia ser interpretada como um centro em $\Re^M$ para os pontos de uma nuvem a_{-1} em $\Re^{M+1}$ projetada sobre $\Re^M$, a qual apresentaria concentração de pontos nas vizinhanças de uma direção ortogonal a todos os vetores $\{\underline{e}_0, \underline{e}_1, \cdots, \underline{e}_p, \cdots, \underline{e}_{M-1}\}$, com largura da região ocupada por a_{-1} em $\Re^{M+1}$ maior que $2\sqrt{\lambda_0}$. Se esta interpretação soa um tanto "esotérica", podemos simplesmente considerar a coordenada $\underline{\psi}_0$ apenas como o centro de um dos dois *clusters* de maior concentração de pontos da nuvem a_0.

É interessante observar que, experimentalmente, ao se remover a função de base radial com vetor centro definido por $\underline{\psi}_0$, deixando somente as demais, cujos vetores centro possuem normas menores, o erro de predição de uma rede RBF para a predição da série temporal S aumenta significativamente na grande maioria dos casos. Sob o ponto de vista da interpretação baseada em uma nuvem a_{-1} em $\Re^{M+1}$, fica experimentalmente implícito que os pontos desta nuvem contribuem na caracterização do processo associado à S, já que

a remoção do centro $\underline{\psi}_0$ que representa a coordenada central da nuvem a_{-1} em $\Re^{M+1}$ aumenta o erro de predição da série *S*. Sob o ponto de vista de que a coordenada $\underline{\psi}_0$ representa apenas o centro de um dos dois *clusters* da nuvem a_0 diríamos que, experimentalmente, os pontos de um destes dois *clusters* contribuem na caracterização do processo associado à *S*, já que a remoção do centro $\underline{\psi}_0$ que representa a coordenada central de um dos *clusters* da nuvem a_0 aumenta o erro de predição da série *S*.

É claro que o centro geométrico de todas as *M* nuvens no espaço $\Re^M$, obtidas por DSE do conjunto **U**, localiza-se em $\underline{0} \in \Re^M$, coordenada que representa o vetor média do conjunto **U**. Mas, justamente por isto, esta coordenada é inútil para localização dos centros das funções de base radial da rede RBF com vistas à predição da série temporal *S*. Primeiro porque esta coordenada é única e estaríamos limitando a rede RBF a ter um só centro. E, segundo, porque a predição seria feita com base na média – um estimador muito pobre do processo associado à série *S*. Isto não acontece quando os centros das *M* funções de base radial são definidos pelos auto-vetores escalonados $\underline{\psi}_m = \underline{e}_m \cdot \sqrt{\lambda_m}$. Os *M* sub-espaços definidos pelos auto-vetores $\underline{e}_m$ e auto-valores λ_m descrevem os *M* modos próprios de variação temporal do processo associado à *S* em torno de sua média, e portanto contêm muito mais informação sobre o desenrolar temporal do processo do que sua média contém. Como todas as *M* nuvens definidas pelos *M* sub-espaços possuem centro geométrico em $\underline{0} \in \Re^M$, a distinção entre as nuvens, para efeito de definição de funções de base radial que devam representar bases distintas para modos de variação temporal independentes, deverá ser feita através de um conjunto de parâmetros que as identifiquem univocamente, não importando o valor médio $\underline{0} \in \Re^M$ comum a todas elas. Assim, em termos das coordenadas dos centros das funções de base radial, o par de parâmetros $(\lambda_m, \underline{e}_m)$ que identifica o *m*-ésimo modo de variação temporal do processo em torno de sua média traduz-se na coordenada $\underline{\psi}_m = \underline{e}_m \cdot \sqrt{\lambda_m}$.

Esta característica única da nova técnica de predição não-linear apresentada nesta tese possibilita que os estados da série S sejam representados como um todo e não mais de forma localizada no tempo, como através do algoritmo *k-means*, ou como através da técnica APC.

5.1.1 DSE para Obtenção dos Centros de um Conjunto de Vetores

Nesta seção a técnica DSE é definida com base na KLT. Na Seção 5.2, esta técnica será aplicada para a atribuição dos centros às funções de base radial de redes neurais artificiais RBF.

Vimos que a atribuição de centros por DSE está associada ao conceito de norma Euclidiana do sub-espaço. Como a média do quadrado da norma Euclidiana dos vetores de um conjunto A é associada ao conceito de variância ou energia média do conjunto é instrutivo examinarmos como a distância Euclidiana se relaciona com o conceito de desvio-padrão de um conjunto de escalares.

Seja um conjunto χ de dados com L elementos escalares, definido pelos pontos $\{x_i\}$, $i=0,1,\cdots,L-1$, representado por uma variável aleatória X, com média μ_x, desvio padrão σ_x estimado por

$$\sigma_x = \sqrt{\frac{1}{L}\sum_{i=0}^{L-1}(x_i-\mu_x)^2}\ , \tag{5.1}$$

e variância σ_x^2. Os limites da probabilidade P de encontrar X fora de um intervalo $\mu_x \pm k\sigma_x$, $k>0$, podem ser determinados através da desigualdade de Tchebycheff [31],

$$P\left(\left|(X-\mu_x)\right| \geq k\sigma_x\right) \leq \frac{1}{k^2}. \tag{5.2}$$

Note que (5.2) define um limite superior para a probabilidade de um ponto x_i do conjunto χ ter um valor que desvie da média μ_x por mais do que k vezes o desvio padrão σ_x da variável aleatória X. Assim, (5.2) justifica o uso de σ_x como uma medida do raio de espalhamento dos elementos de χ ao redor de μ_x, isto é, o raio que delimita a região de χ com centro em μ_x onde é possível encontrar "a grande maioria" dos elementos de χ que se desviam da média. A variância σ_x^2 é associada à média do quadrado da norma Euclidiana dos elementos x_i do conjunto χ com relação ao valor médio μ_x do conjunto. Assim, o desvio padrão σ_x medirá a distância Euclidiana média dos pontos x_i com relação ao valor médio μ_x. Portanto, de (5.1) e (5.2), pode-se esperar que nas vizinhanças de σ_x (ou de $-\sigma_x$, já que $\sigma_x = \pm\sqrt{\sigma_x^{\,2}}$) ocorra a maior concentração dos pontos do conjunto χ que se caracterizem por se desviar da média μ_x.

Considere-se agora um conjunto A de vetores de dados, de média zero, composto de L vetores $\underline{x}_i \in \Re^M$, $i = 0,1,\cdots,L-1$, tal que $\mathrm{A} = \{\underline{x}_0, \underline{x}_1, \cdots, \underline{x}_{L-1}\}$. Através da aplicação da KLT ao conjunto A obtém-se os componentes principais ou sub-espaços de A, os quais são as projeções do conjunto dos vetores de A sobre a direção definida por cada um de seus M auto-vetores. A matriz de covariância $\mathbf{C}$ de A é obtida através de

$$\mathbf{C} = \frac{1}{L}\sum_{i=0}^{L-1} \underline{x}_i \cdot \underline{x}_i^{\,T} \tag{5.3}$$

De acordo com a Transformação Karhunen-Loève (Apêndice B), os auto-valores e auto-vetores da matriz de covariância $\mathbf{C}$ são obtidos através da solução da Equação (5.4) para λ_m e $\underline{e}_m$,

$$\mathbf{C}\,\underline{e}_m = \lambda_m\, \underline{e}_m \tag{5.4}$$

onde $\underline{e}_m$ são os auto-vetores da matriz $\mathbf{C}$ e λ_m são os auto-valores associados a estes auto-vetores, com $m = 0,1,\cdots,M-1$.

O conjunto de M auto-vetores $\underline{e}_m \in \Re^M$ gerado pela KLT, obtidos da solução de (5.4), forma uma base ortonormal para os M sub-espaços de A, os quais, conforme discutido anteriormente, definem nuvens de maior concentração de pontos em $\Re^M$. O m-ésimo auto-valor λ_m associado ao auto-vetor $\underline{e}_m$ aproxima, conforme visto, a média do quadrado da norma Euclidiana dos pontos da m-ésima nuvem à origem do sistema de M eixos Cartesianos definido pela base ortonormal.

Assim, conforme discutido na Seção 5.1, estando os auto-valores λ_m ordenados tal que seu valor decresça com o incremento do índice m, a DSE aproxima a coordenada central em $\Re^M$ da nuvem $m-1$ do conjunto A de vetores por

$$\underline{\Psi}_m = \sqrt{\lambda_m}\,\underline{e}_m \tag{5.5}$$

onde $\sqrt{\lambda_m}$ representa o desvio-padrão σ_m dos pontos obtidos pela projeção dos vetores de A sobre o m-ésimo eixo Cartesiano.

5.1.2 Resultados Comparativos

Nesta seção são apresentados os resultados experimentais referentes à aplicação da Decomposição em Sub-Espaços baseada na Transformada Karhunen-Loève para determinação dos centros das M nuvens de dados no espaço $\Re^M$, obtidas por DSE do conjunto U de vetores em $\Re^M$.

Para efeito de comparação são utilizados os resultados obtidos na determinação dos centros através do algoritmo *k-means*, previamente discutido nesta tese.

As Figuras 5.2, 5.3 e 5.4 apresentam resultados da determinação dos centros através dos algoritmos *k-means* e DSE.

A Figura 5.2 mostra um conjunto **U** de vetores em $\Re^2$, de média zero, com dois *clusters* de vetores A e B. São também mostradas as coordenadas E_0 e E_1 definidas pelos vetores $\underline{\Psi}_0 = \sqrt{\lambda_0}\,\underline{e}_o$ e $\underline{\Psi}_1 = \sqrt{\lambda_1}\,\underline{e}_1$, onde $\underline{e}_0$ e $\underline{e}_1$ são os auto-vetores da matriz de covariância de **U** associados aos auto-valores λ_0 e λ_1. As coordenadas K_0 e K_1 são obtidas através do aplicação do algoritmo *k-means* sobre o conjunto **U**. Note que o algoritmo *k-means* considera os *clusters* A e B como nuvens individuais e identifica seus centros respectivamente como K_0 and K_1. A determinação dos centros através da DSE considera a composição de *clusters* A+B como uma nuvem com centro em E_1 e considera o *cluster* A como a outra nuvem com centro em E_0. Sob a interpretação de que a coordenada E_0 aproxima a coordenada central em $\Re^2$ de uma nuvem em $\Re^3$, podemos visualizar esta nuvem em $\Re^3$ como a composição de duas regiões tubulares de pontos em $\Re^3$. O eixo de ambos os tubos é perpendicular ao plano da Figura 5.2. O centro da seção transversal de um dos tubos é o centro do *cluster* A sendo o raio do tubo relacionado com a distância da coordenada E_1 à origem. O centro da seção transversal do outro tubo é o centro do *cluster* B sendo o raio do tubo também relacionado com a distância da coordenada E_1 à origem. Não sabemos quanto os pontos se espalham ao longo dos tubos, acima e abaixo do plano da Figura 5.2, apenas sabemos que eles se espalham por uma distância que é maior do que a distância da coordenada E_0 à origem.

Na Figura 5.3 o número de *clusters* de vetores é aumentado para três (A, B e C), enquanto o número de centros permitido para o algoritmo *k-means* é mantido em dois. Nesta nova situação, o algoritmo *k-means* considera a composição de *clusters* A+B como uma nuvem com centro em K_0 e o *cluster* C como a outra nuvem com centro em K_1. A determinação dos centros via DSE considera a composição de *clusters* A+B como uma nuvem com centro em E_0 e a composição de *clusters* A+B+C como a outra nuvem com centro em E_1.

A Figura 5.4 apresenta quatro *clusters* de vetores (A, B, C e D) com dois centros atribuídos pelo algoritmo *k-means*. A DSE considera a composição de *clusters* A+B como uma nuvem com centro em E_0 e a composição de *clusters* A+B+C+D como uma nuvem com centro em E_1. O *k-means* considera a composição de *clusters* A+B como uma nuvem com centro em K_0 e a composição de *clusters* C+D como uma nuvem com centro em K_1.

Note que, em todos os casos, a atribuição dos centros das nuvens de dados pela DSE necessita, na média, um menor raio de alcance para poder acessar todos os pontos em **U** do que a atribuição dos centros através do algoritmo *k-means*. Este fato pode ser uma vantagem na atribuição dos centros de redes RBF já que os centros atribuídos pela DSE tendem a ser *kernels* globais em **U** [11]. Portanto, o uso do algoritmo *k-means* para atribuição dos centros, e em maior grau, a atribuição dos centros via técnica APC, tende a gerar *kernels* locais em **U** [50].

Em contraste ao uso de *kernels* globais, o uso de *kernels* locais como centro de redes neurais RBF freqüentemente leva o sistema de equações que determina os pesos sinápticos da rede a uma condição mais sobre-determinada, aumentando o erro de aproximação da rede RBF. A não-localidade geométrica dos centros obtidos pela DSE é uma vantagem adicional e independente da já discutida não-localidade temporal resultante da habilidade da DSE em captar a correlação de longo prazo do processo associado à série S através da *eigen*–decomposição da matriz de covariância de **U**, sendo **U** o conjunto de vetores formado de uma janela de predição definida sobre S.

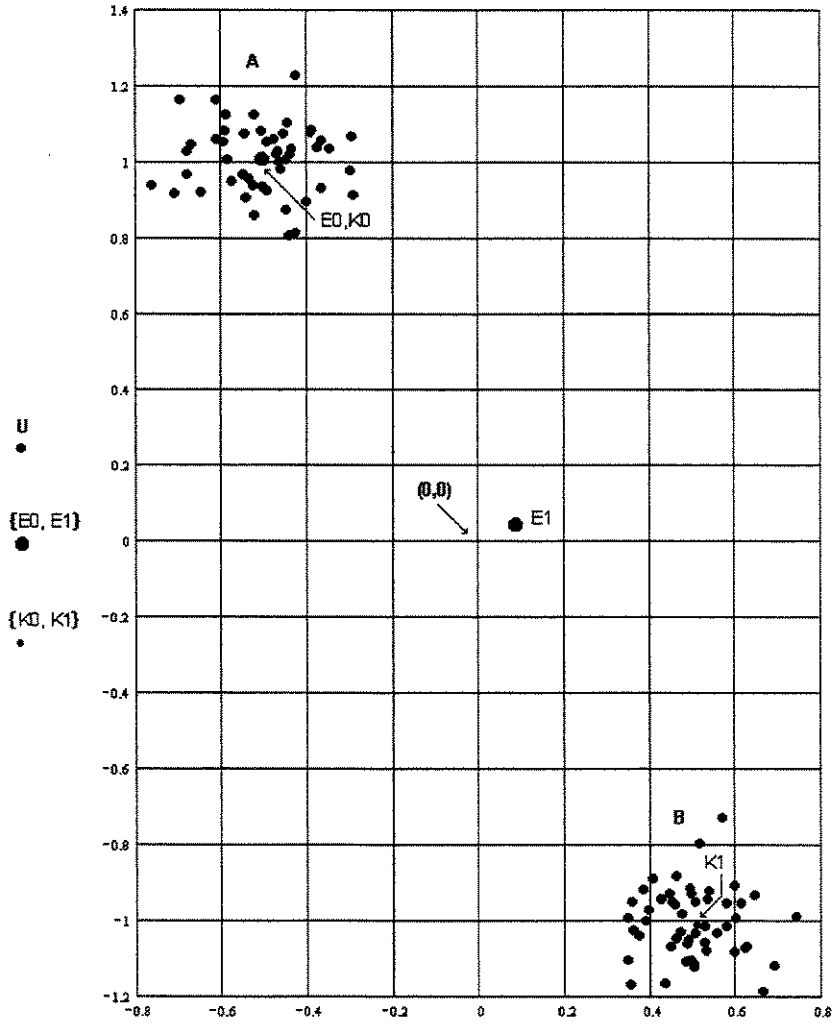

Figura 5.2: Conjunto de vetores em $\Re^2$, com dois *clusters* de vetores.

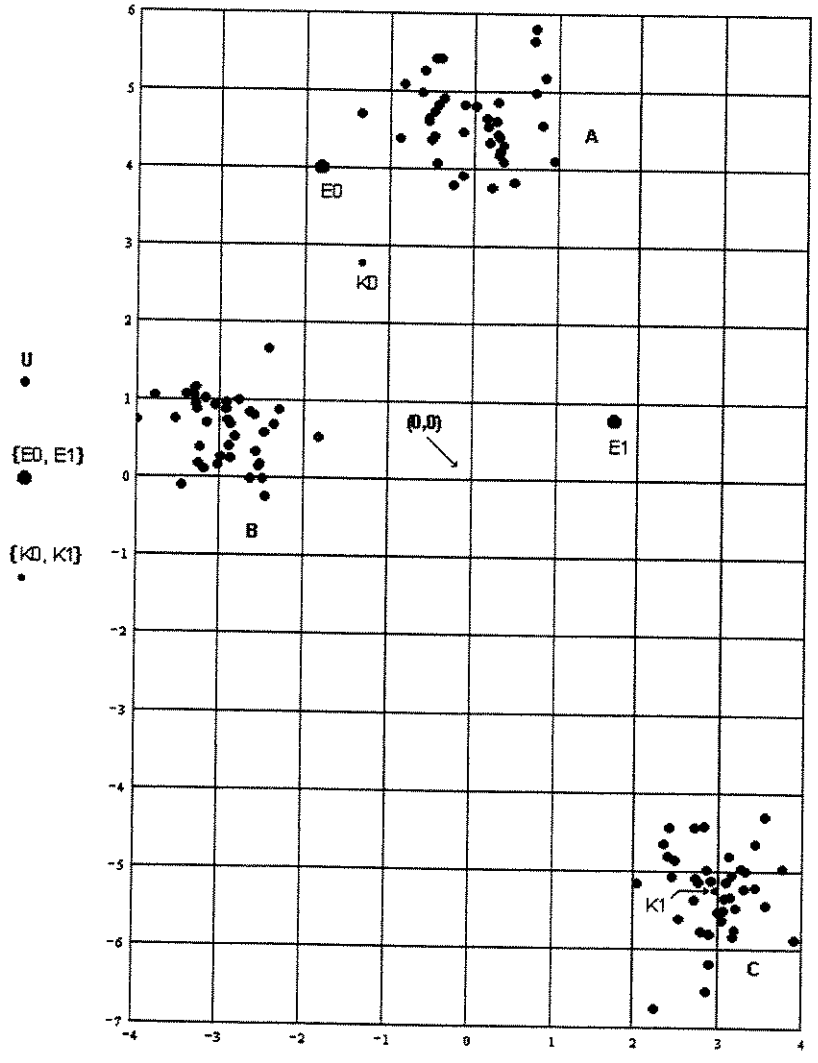

Figura 5.3: Conjunto de vetores em $\Re^2$, com três *clusters* de vetores.

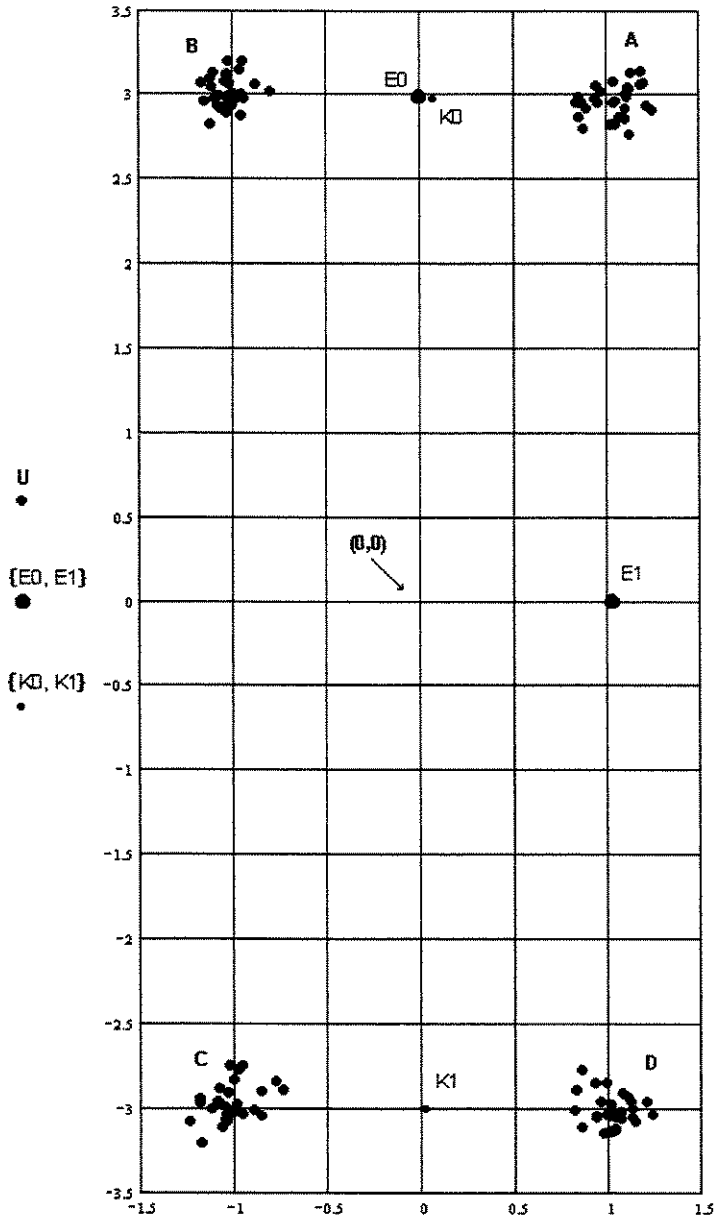

Figura 5.4: Conjunto de vetores em $\Re^2$, com quatro *clusters* de vetores.

5.1.3 O Efeito da Não-Localidade Geométrica dos Centros Obtidos por DSE via KLT no Erro de Aproximação da Pseudo-Inversão de Moore-Penrose

Conforme observado nos exemplos apresentados na seção anterior, a técnica DSE situa os centros dos conjuntos de dados em coordenadas mais centrais em relação a todo o conjunto de dados do que os centros obtidos pelo algoritmo *k-means* ou pela técnica APC. Isto pode reduzir o erro de aproximação da pseudo-inversão matricial usada na determinação do vetor de pesos sinápticos $\underline{w}(n)$ obtido no instante n com base na janela de predição $p(n)$. Assim, o objetivo desta seção é comparar experimentalmente o erro de aproximação de duas redes RBF, R_{Ψ} e R_{K}, com os centros das funções de base radial da rede R_{Ψ} obtidos por DSE e com os centros das funções de base radial da rede R_{K} obtidos pelo algoritmo *k-means*.

A Figura 5.5 representa uma função definida por 100 pontos representando a transição linear de um patamar de 25 pontos no nível –1 para um patamar de 25 pontos no nível +1, sendo adicionado ruído Gaussiano com desvio padrão 0.02 e média zero. Vamos considerar esta função como modelo para uma janela de predição genérica $p(n)$, de tamanho $N=100$, definida no instante n sobre uma série temporal S. Esta suposição não implica em perda de generalidade do modelo, já que uma série temporal nada mais é do que uma sucessão de transições e sempre pode-se fazer o tamanho N da janela tal que esta seja representativa de uma transição assim modelada. Por simplicidade, assume-se que o conjunto $\mathbf{U}$ de vetores formados de $p(n)$ são vetores em $\Re^M$, $M=2$, cujos pontos por eles definidos em $\Re^2$ são mostrados na Figura 5.6.

Os $K=M$ vetores centros $\underline{\Psi}_0 \in \Re^2$ e $\underline{\Psi}_1 \in \Re^2$, obtidos por DSE de $\mathbf{U}$, são respectivamente identificados pelas coordenadas E_0 e E_1 na Figura 5.6. Da mesma forma, os K vetores centros $\underline{\mathbf{K}}_0 \in \Re^2$ e $\underline{\mathbf{K}}_1 \in \Re^2$, obtidos pela aplicação do algoritmo *k-means* ao conjunto $\mathbf{U}$, são respectivamente identificados pelas coordenadas K_0 e K_1 na Figura 5.6.

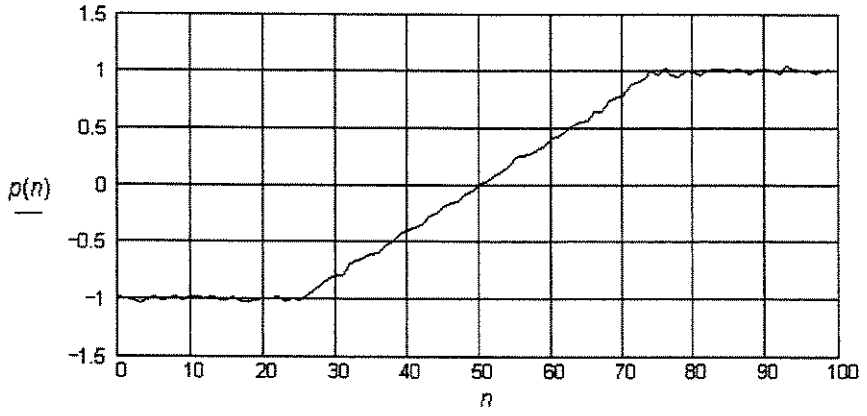

Figura 5.5: Modelo de transição para uma série temporal genérica.

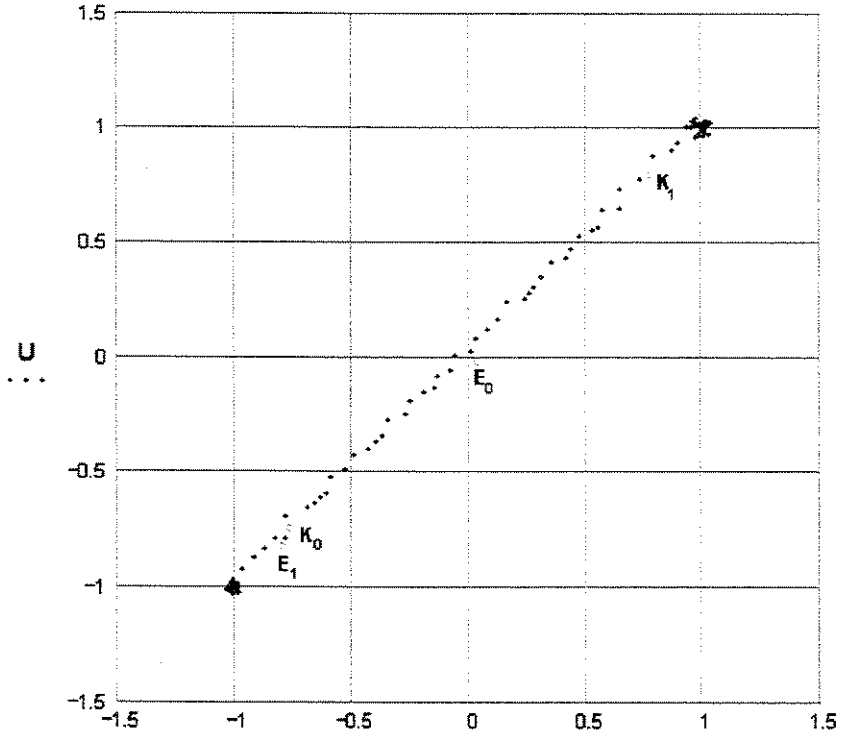

Figura 5.6: Conjunto de vetores em $\Re^2$ obtidos da janela de predição e localização dos centros.

O vetor de saídas desejadas comum a ambas as redes R_{Ψ} e R_{K} é dado por $\underline{d}(n)=[u(n)\ \ u(n-1)\cdots\ u(n-N_d+1)]^T$, sendo $u(n-i),\ \ i=0,1,\cdots,N_d-1$, o elemento que está localizado na série S uma posição à frente do vetor $\underline{u}(n-i-1)\in\Re^2$ do conjunto $\mathbf{U}$. $N_d=N-2$ é o número de vetores $\underline{u}(n-i-1)\in\Re^2$ possíveis de serem definidos sobre a janela $p(n)$, de forma que dois vetores consecutivos de $\mathbf{U}$ estejam deslocados entre si da distância temporal entre dois elementos de S.

A matriz $\mathbf{\Phi}(n)$, no instante n em que é definida a janela $p(n)$ sobre S, é obtida de

$$\mathbf{\Phi}_{i,j}(n)=\exp\left(\frac{-1}{2\sigma^2(n)}\left\|\underline{u}(n-i-1)-\underline{\mathbf{t}}_j(n)\right\|^2\right),\quad i=0,1,\cdots,N_d-1, \tag{5.6}$$

$$j=0,1,\cdots,K-1$$

onde $\|\cdot\|$ é o operador que representa a norma Euclidiana do vetor argumento, e

$$\sigma^2(n)=\xi\max\left\{\left\|\underline{t}_l(n)-\underline{t}_m(n)\right\|^2\right\},\quad l,m=0,1,\cdots,K-1 \tag{5.7}$$

sendo ξ denominado Fator de Variância, o qual é obtido experimentalmente de forma a minimizar o erro de aproximação obtido.

Para $\underline{\mathbf{t}}_j=\underline{\Psi}_j$ e $\xi=\xi_{\Psi}$ em (5.6) e (5.7) obtemos a matriz $\mathbf{\Phi}_{\Psi}$, representativa da rede R_{Ψ} e, para $\underline{\mathbf{t}}_j=\underline{\mathbf{K}}_j$ e $\xi=\xi_{K}$ em (5.6) e (5.7) obtemos a matriz $\mathbf{\Phi}_{K}$, representativa da rede R_{K}.

Os vetores de pesos sinápticos $\underline{w}_{\Psi}$ e $\underline{w}_{K}$, respectivos às redes RBF R_{Ψ} e R_{K}, são obtidos através de

$$\underline{w}_{\Psi}=\mathbf{\Phi}_{\Psi}^{+}\cdot\underline{d} \tag{5.8}$$

$$\underline{w}_{\mathbf{K}} = \mathbf{\Phi}_{\mathbf{K}}^{+} \cdot \underline{d} \tag{5.9}$$

onde $\{\cdot\}^{+}$ é o operador que representa a pseudo-inversa [50] de Moore-Penrose da matriz argumento.

O erro médio quadrático respectivo às aproximações obtidas com as redes RBFs R_{Ψ} e $\mathrm{R}_{\mathbf{K}}$ é definido como

$$MSE_{\Psi} = \frac{1}{N_d} \left\| \underline{d} - \mathbf{\Phi}_{\Psi} \cdot \underline{w}_{\Psi} \right\|^2 \tag{5.10}$$

$$MSE_{\mathbf{K}} = \frac{1}{N_d} \left\| \underline{d} - \mathbf{\Phi}_{\mathbf{K}} \cdot \underline{w}_{\mathbf{K}} \right\|^2 \tag{5.11}$$

Variando-se o parâmetro ξ em (5.7) determinou-se que, para o conjunto de vetores $\mathbf{U}$ da Figura 5.6, obtido da janela de predição $p(n)$ da Figura 5.5, os erros de aproximação MSE_{Ψ} e $MSE_{\mathbf{K}}$ são minimizados respectivamente para $\xi_{\Psi} = 12.0$ e $\xi_{\mathbf{K}} = 7.0$. Os valores obtidos para os erros de aproximação assim minimizados são $MSE_{\Psi} = 8.6 \times 10^{-4}$ e $MSE_{\mathbf{K}} = 1.4 \times 10^{-3}$, resultando em uma razão entre os erros de aproximação $\frac{MSE_{\mathbf{K}}}{MSE_{\Psi}} = 1.63$.

Portanto, a rede RBF com atribuição de centros via algoritmo *k-means* resultou em um erro de aproximação da janela $p(n)$ 63% maior do que o erro obtido com a rede RBF com atribuição de centros via DSE. É importante relembrar aqui que um erro de aproximação baixo não significa que o erro de predição também o seja. Embora o erro de predição dependa do erro de aproximação, o primeiro depende muito mais da capacidade da rede RBF extrapolar temporalmente a janela de predição $p(n)$ do que depende do segundo.

Por exemplo, toda vez que se utilizou o algoritmo Gradiente Estocástico para determinar os centros de uma RBF preditora de uma série S (aliás, sob um alto custo

computacional resultante), o erro de aproximação obtido para cada janela $p(n)$ foi baixíssimo (muito menor do que o usualmente obtido para a DSE e *k-means*), mas o erro de predição resultante foi catastrófico. Isto pode ser interpretado da seguinte maneira: mesmo quando "adere" perfeitamente aos pontos em $\Re$ que antecipam o processo em $\Re^M$ associado à S para uma particular janela $p(n)$, a superfície não-linear que define o mapeamento $\Re^M \rightarrow \Re$ gerado pela rede RBF pode não representar fielmente os demais pontos em $\Re$ que antecipam o processo em $\Re^M$ fora de $p(n)$.

5.2 Predição Não-Linear através de Redes Neurais Artificiais RBF com Determinação dos Centros por DSE via KLT

Esta seção define e apresenta detalhadamente o método de predição a um passo de uma série temporal *S*, através de redes RBF com centros atribuídos por DSE.

Dada uma série temporal $S = \{u(0), u(1), \cdots\}$, a qualquer instante n o objetivo é predizer a próxima amostra $u(n+1)$ a uma janela de predição $p(n) = \{u(n-N+1), \cdots, u(n-1), u(n)\}$ de N amostras prévias conhecidas em *S*.

Com este objetivo em vista, são assumidas as seguintes condições:

1- Que exista um processo estocástico vetorial $\mathbf{U}$, definido em $\Re^M$, associado ao desenrolar temporal de *S*.

2- Que o mapeamento não-linear $\Im : \mathbf{U} \in \Re^M \rightarrow y \in \Re$ gerado pela rede RBF seja capaz de antecipar temporalmente, mediante o treino da rede RBF, o processo $\mathbf{U}$ através do valor obtido para y. No instante n, a antecipação temporal de $\mathbf{U}$ consiste em y aproximar a componente desconhecida do vetor de $\mathbf{U}$ que ocorre no instante $n+1$.

3- Que, embora localizada no tempo, a janela $p(n)$ definida sobre S tenha abrangência suficiente para que os modos de variação básicos de $\mathbf{U}$ possam ser representados via

DSE. Portanto, o conjunto de vetores em $\Re^M$ formados de $p(n)$ será referido como o próprio conjunto **U** de vetores representativo do processo estocástico vetorial **U** associado ao desenrolar temporal de *S*.

Para representação de **U**, a janela de predição $p(n)$ é estruturada em um conjunto contendo todas as $L = N - M + 1$ seqüências distintas e consecutivas de M amostras adjacentes, dispostas em ordem crescente de índice de ocorrência, de tal forma que duas seqüências consecutivas representem duas sub-janelas de M amostras deslocadas entre si de uma amostra. O conjunto de vetores **U** é então definido pelo conjunto de L vetores $\underline{u}_i \in \Re^M$, $i = 0,1,\cdots,L-1$, formado quando considera-se cada uma das L seqüências assim obtidas de $p(n)$ como um vetor $\Re^M$ dimensional.

Por exemplo, vamos supor que, no instante *n*, queiramos construir o conjunto **U** de L vetores $\underline{u}_i \in \Re^M$, $i = 0,1,\cdots,L-1$, a partir de uma janela $p(n)$ de tamanho $N = L + M - 1$, objetivando prever o elemento $u(n+1)$ imediatamente adiante de $p(n)$ em *S*. Para tanto, formamos de $p(n)$ os L vetores $\underline{u}_i \in \Re^M$, cada um contendo M amostras consecutivas em $p(n)$ tal que $\underline{u}_i(n) = \underline{u}(n-i)$, isto é, $\mathbf{U} = \{\underline{u}(n), \underline{u}(n-1), \underline{u}(n-2), \cdots, \underline{u}(n-L+1)\}$ onde $\underline{u}(n-i) = [u(n-i)\ u(n-i-1) \cdots u(n-i-M+1)]^T$. A Figura 5.7 mostra uma janela $p(n)$ de tamanho $N = 8$ e o resultante conjunto **U** de L vetores $\underline{u}(n-i) \in \Re^3$ com $L = N - M + 1 = 6$, $i = 0,1,\cdots,L-1$, tal que $\mathbf{U} = \{\underline{u}(n), \underline{u}(n-1), \underline{u}(n-2), \cdots, \underline{u}(n-5)\}$.

No instante *n*, os centros das K funções de base radial da rede RBF, $K = M$, são determinados pelas coordenadas $\underline{\psi}_k = \underline{e}_k \cdot \sqrt{\lambda_k}$, $k = 0,1,\cdots,K-1$, sendo λ_k e $\underline{e}_k$ respectivamente os auto-valores e auto-vetores da matriz de covariância **C** do conjunto **U** formado de $p(n)$.

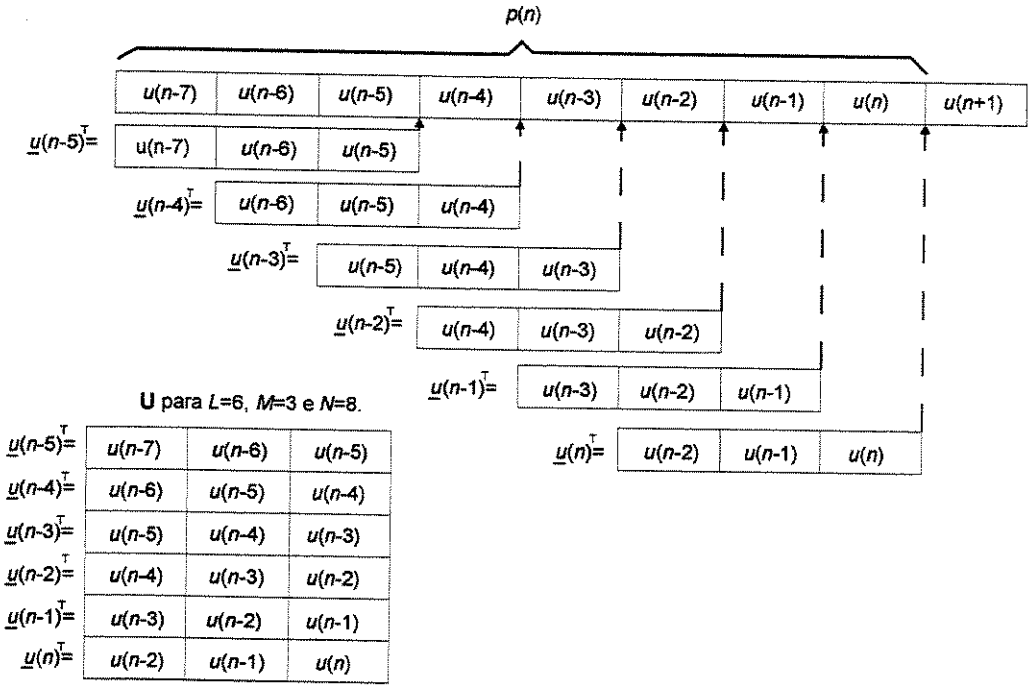

Figura 5.7: $p(n)$ e $\mathbf{U}$ para $L=6$ e $M=3$. Note que, em função da construção gráfica adotada para facilitar a compreensão do método, os vetores $\underline{u}(n-i)$ devem ser lidos da direita para a esquerda nesta figura. Observe que $p(n)$ é formada por $N=L+M-1=8$ amostras.

Para tal, primeiramente é calculado o vetor média do conjunto $\mathbf{U}$, conforme Equação (5.12).

$$\overline{\underline{u}(n)}=\frac{1}{L}\sum_{i=0}^{L-1}\underline{u}(n-i) \tag{5.12}$$

A seguir, o conjunto $\mathbf{U}$ é centralizado em torno de sua média, formando o conjunto $\mathbf{X}$ de vetores $\underline{x}_i \in \Re^M$ dado por,

$$\underline{x}_i(n)=\underline{x}(n-i)=\underline{u}(n-i)-\overline{\underline{u}(n)},\quad i=0,1,\cdots,L-1 \tag{5.13}$$

A matriz covariância $\mathbf{C}(n)$ é, então, determinada por

$$\mathbf{C}(n)=\frac{1}{L}\sum_{i=0}^{L-1}\underline{x}(n-i)\underline{x}(n-i)^T \tag{5.14}$$

sendo $\mathbf{C}(n)$ uma matriz de ordem $[M\times M]$, conforme a Equação (5.15).

$$\mathbf{C}(n)=\begin{bmatrix} C_{0,0}(n) & C_{0,1}(n) & C_{0,2}(n) & \cdots & C_{0,(M-1)}(n) \\ C_{1,0}(n) & C_{1,1}(n) & C_{1,2}(n) & \cdots & C_{1,(M\_1)}(n) \\ C_{2,0}(n) & C_{2,1}(n) & C_{2,2}(n) & \cdots & C_{2,(M-1)}(n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{(M-1),0}(n) & C_{(M-1),1}(n) & C_{(M-1),2}(n) & \cdots & C_{(M-1),(M-1)}(n) \end{bmatrix} \tag{5.15}$$

Os M auto-valores e auto-vetores de $\mathbf{C}(n)$ são obtidos da solução da Equação 5.16 para $\lambda_m(n)$ e $\underline{e}_m(n)$.

$$\mathbf{C}(n)\underline{e}_m(n)=\lambda_m(n)\underline{e}_m(n), \quad m=0,1,...,M-1. \tag{5.16}$$

A solução da Equação (5.16) para λ_m e $\underline{e}_m$ no instante n apresenta elevado custo computacional, se executada sem algum cuidado. Para resolver este problema aplica-se a transformação de Householder [59] . Inicialmente, o algoritmo de Householder transforma a matriz $\mathbf{C}(n)$, que é simétrica, à forma tridiagonal, através de $[(M\times M)-1]$ transformações. À matriz tridiagonal é aplicada uma transformação **QL**, resultando em uma matriz ortogonal **Q** e uma matriz triangular inferior **L**. A matriz **Q** obtida é pré-multiplicada pela matriz **L** e, novamente, a transformação **QL** é aplicada, resultando em novas matrizes **Q** e **L**. O algoritmo assim procede, sucessivamente, até que um número suficiente de transformações seja aplicado. Os auto-valores da matriz $\mathbf{C}(n)$ estarão, no final deste processo, na diagonal principal da matriz **L**. Com esta técnica, em média, são necessárias menos de duas iterações por auto-valor.

Da solução de (5.16) resultam os M auto-vetores $\underline{e}_0(n), \underline{e}_1(n), ..., \underline{e}_{M-1}(n)$, com auto-valores associados $\lambda_0(n), \lambda_1(n), ..., \lambda_{M-1}(n)$, sendo $\lambda_{p-1}(n)>\lambda_p(n)$, $p=1,2,\cdots,M-1$.

Os vetores $\underline{\Psi}_k(n)=\underline{e}_k(n)\cdot\sqrt{\lambda_k(n)}$, $k=0,1,\cdots,K-1$, com $K\leq M$, são, então, atribuídos aos K centros das funções de base radial da rede RBF. Assim, os centros determinados por DSE serão,

$$\begin{aligned}
&\underline{\Psi}_0(n)=\sqrt{\lambda_0(n)}\,\underline{e}_0(n),\\
&\underline{\Psi}_1(n)=\sqrt{\lambda_1(n)}\,\underline{e}_1(n),\\
&\cdots\\
&\underline{\Psi}_{(K-1)}(n)=\sqrt{\lambda_{(K-1)}(n)}\,\underline{e}_{(K-1)}(n)
\end{aligned} \tag{5.17}$$

Em geral, o número K de centros é igual ao número M de sub-espaços de $\mathbf{U}$. No entanto, para séries temporais com superposição de ruído aproximadamente branco, uma significativa melhora do erro de predição é obtida quando considera-se somente os maiores auto-valores λ_k na formação dos centros $\underline{\Psi}_k=\underline{e}_k\cdot\sqrt{\lambda_k}$, fazendo-se $K<M$. Isto ocorre porque o ruído de baixa correlação é representado nos sub-espaços de menores auto-valores, e sem uma função de base radial que a ele represente, resulta que a rede RBF acaba por desconsiderar o ruído no processo de predição.

A Figura 5.8 apresenta a arquitetura da rede neural RBF utilizada na técnica de predição por DSE. No instante n, a rede neural tem uma camada de entrada formada por M nós, a qual recebe o vetor $\underline{x}\in\Re^M$ pertencente ao conjunto $\mathbf{X}$ resultante da centralização do conjunto $\mathbf{U}$ em torno de sua média, e uma camada escondida formada por K neurônios não-lineares. Portanto, a rede assim estruturada representa K vetores de estado do processo $\mathbf{U}$ com uma ordem de predição M.

O vetor $\underline{\Psi}_k \in \Re^M$ é o k-ésimo vetor centro da rede RBF ou vetor de estado do processo, w_k é o k-ésimo peso sináptico e σ^2 é a variância comum a todos os centros Gaussianos, com $k=0,1,\cdots,K-1$.

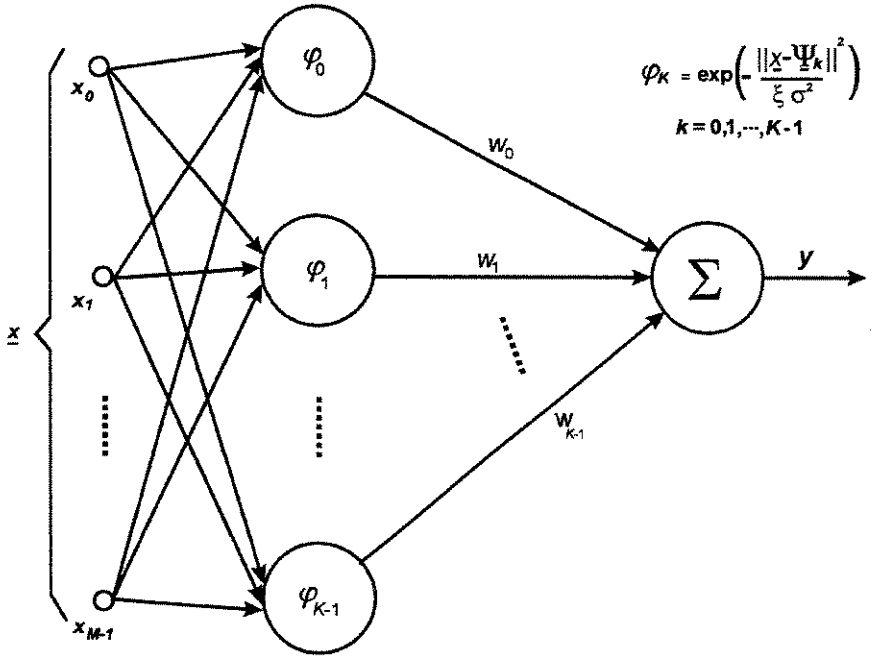

Figura 5.8: Arquitetura da rede neural RBF para predição por DSE.

A k-ésima unidade escondida computa a distância entre o vetor centro $\underline{\Psi}_k \in \Re^M$ e o vetor de entrada $\underline{x} = [x_0 \; x_1 \; \ldots \; x_{M-1}]^T$, distância que é utilizada como argumento do mapeamento não-linear expresso por (5.18), resultando na saída φ_k da referida unidade.

$$\varphi_k(n) = \exp\left[-\frac{1}{\xi(n)\sigma^2(n)}\|\underline{x}(n) - \underline{\Psi}_k(n)\|^2\right] \tag{5.18}$$

onde $\sigma^2(n)$ é a variância e ξ é o fator de variância (observe, na Equação (5.18), que o fator $\xi(n)$ absorveu a constante 2, presente no denominador da função Gaussiana.). Este último é inicializado com um valor constante ξ_{init} e, como será visto na próxima seção deste capítulo (Seção 5.2.1), é adaptativamente atualizado a cada instante n ao longo do processo de predição, através do algoritmo LMS [50].

O produto interno entre o vetor formado pelas saídas dos neurônios da camada escondida $\underline{\varphi} = [\varphi_0 \ \varphi_1 \ \cdots \ \varphi_{K-1}]^T$ e o vetor de pesos da camada de saída $\underline{w} = [w_0 \ w_1 \ \cdots \ w_{K-1}]^T$ resulta na saída y da rede RBF, expressa na Equação (5.19).

$$y = \underline{\varphi}^T \cdot \underline{w} \tag{5.19}$$

O treino da rede neural RBF consiste especificamente na determinação dos pesos sinápticos no instante n, os quais são obtidos a partir da aplicação de cada vetor $\underline{x}(n-j) = [x(n-j) \ x(n-j-1) \cdots x(n-j-M+1)]^T$ do conjunto **X** de vetores de treino, $j = 1,2,\cdots,L-1$, à camada de entrada da rede, exceto o vetor mais recente $\underline{x}(n) = [x(n) \ x(n-1) \cdots x(n-M+1)]^T$.

A cada vetor $\underline{x}(n-j)$ aplicado, obtém-se K valores $\varphi_k(n-j)$, $k = 0,1,\cdots,K-1$, na saída da camada escondida, valores dados por (5.18). Estes K valores, organizados em forma vetorial, definem o vetor de saída da camada escondida $\underline{\varphi}(n-j) = [\varphi_0(n-j) \ \varphi_1(n-j) \ \ldots \ \varphi_{K-1}(n-j)]^T$. Assim, a matriz de transição $\boldsymbol{\Phi}(n)$ da rede RBF pode ser definida como

$$\boldsymbol{\Phi}(n) = \begin{bmatrix} \underline{\varphi}(n-L+1)^T \\ \vdots \\ \underline{\varphi}(n-2)^T \\ \underline{\varphi}(n-1)^T \end{bmatrix} \tag{5.20}$$

e o vetor de pesos sinápticos da camada de saída $\underline{w}(n)$ é dado por

$$\underline{w}(n) = \boldsymbol{\Phi}^+(n)\underline{d}(n) \tag{5.21}$$

onde $\underline{d}(n) = [x(n-L+2) \cdots x(n-1) \ x(n)]^T$ é o vetor saída-desejada e $\{\cdot\}^+$ é o operador que representa a pseudo-inversa de Moore-Penrose da matriz argumento [50]. Note de (5.20) que, como $\boldsymbol{\Phi}(n) \cdot \underline{w}(n) = \underline{d}(n)$, cada elemento do vetor

$\underline{d}(n) = [x(n-L+2) \cdots x(n-1)\ x(n)]^T$ é associado à amostra que está uma posição à frente na janela de predição $p(n)$ com respeito ao vetor $\underline{x}(n-j)$, $j = 1,2,\cdots,L-1$, vetor este que deu origem à linha $\underline{\varphi}(n-j)^T$ em $\boldsymbol{\Phi}(n)$. Neste trabalho, conforme observado no Capítulo 4, a operação de pseudo-inversão matricial é obtida por Decomposição em Valores Singulares, a qual é discutida no Apêndice A.

Note que a matriz $\boldsymbol{\Phi}(n)$ definida por (5.20) tem interpretação idêntica à da matriz definida por (4.22), apenas a ordem das linhas $\underline{\varphi}(n-j)^T$ encontra-se invertida. Portanto, a rede neural RBF com vetor de pesos sinápticos $\underline{w}(n)$ dado por (5.21) age como a matriz de transição de um filtro não-linear. Assim, ao aplicar o mais recente vetor do conjunto de treino $\underline{x}(n) = [x(n)\ x(n-1) \cdots\ x(n-M+1)]^T$ à camada de entrada desta rede RBF, a sua saída $y(n)$ será a estimativa $\hat{x}(n+1)$ da próxima amostra $x(n+1)$, imediatamente adiante de $\underline{x}(n)$.

Ao restaurar o vetor média, o conjunto **X** é transformado no conjunto original **U** representativo do processo associado à série *S*. Portanto, a estimativa $\hat{u}(n+1)$ da próxima amostra $u(n+1)$ a uma janela de predição $p(n) = \{u(n-N+1), \cdots, u(n-1), u(n)\}$ de N amostras prévias conhecidas em S é dada por

$$\hat{u}(n+1) = y(n) + \overline{u_0(n)} \tag{5.22}$$

onde $\overline{u_0(n)}$ é o componente mais recente do vetor média $\overline{\underline{u}(n)}$ e $y(n)$ é a saída da rede neural RBF para o vetor de entrada $\underline{x}(n) = [x(n)\ x(n-1) \cdots\ x(n-M+1)]^T$, definida por

$$y(n) = \underline{\varphi}(n)^T \cdot \underline{w}(n) \tag{5.23}$$

com $\underline{w}(n)$ dado pela Equação (5.21).

A totalidade da série temporal S é predita através de repetidos deslizamentos de $p(n)$ uma amostra adiante sobre S, isto é, repetidamente incrementando n de um e obtendo a estimativa da próxima amostra, $\hat{u}(n+1)$.

Note que, na predição a um passo de uma série temporal S através de redes neurais RBF com centros atribuídos por DSE , o tipo de aprendizado utilizado na determinação dos centros é Não-Supervisionado enquanto que o aprendizado utilizado na determinação dos pesos sinápticos é Supervisionado [51]. A variância dos centros é proporcional ao quadrado da máxima norma Euclidiana entre eles, com constante de proporcionalidade dada pelo Fator de Variância ξ. Quando ξ é constante podemos dizer que a variância dos centros é ajustada de modo Não-Supervisionado, mas quando ξ é ajustado pelo algoritmo LMS, conforme será visto na próxima seção, a variância dos centros será ajustada de modo Supervisionado.

5.2.1 O Fator de Variância Adaptativo

Conforme já discutido em seções anteriores, um baixo erro de aproximação não significa necessariamente um baixo erro de predição. Comprovamos experimentalmente inúmeras vezes este comportamento da rede RBF como preditora a um passo de uma série S, em especial quando se trata da variância atribuída aos centros. Na esperança de reduzir o erro de predição de redes RBF com K centros determinados tanto por APC, *k-means* ou DSE, experimentou-se atribuir variâncias distintas para cada centro, e ajustá-las através de inúmeras iterações do algoritmo Gradiente Estocástico, dentro de cada janela $p(n)$, objetivando reduzir o erro de aproximação a cada n, conforme já visto na Seção 4.1. Embora o erro de aproximação obtido para cada janela de predição $p(n)$ resultasse baixo, o erro de predição era muito maior do que quando se utilizava uma única variância σ^2 comum a todos os centros, à qual era atribuído um valor proporcional ao quadrado da máxima distância Euclidiana entre os vetores centros no instante n. Denominou-se Fator de

Variância ξ à constante de proporcionalidade que define, no instante n, a variância σ^2 a partir do quadrado da máxima distância Euclidiana entre os centros.

No entanto, se um baixo erro de aproximação não significa necessariamente um baixo erro de predição, um alto erro de aproximação quase sempre implica em um alto erro de predição. Buscando uma melhora de desempenho neste sentido, observou-se que, ao ajustar experimentalmente ξ de forma a minimizar o erro de aproximação, obtinha-se uma redução concomitante do erro de predição. Porém, a redução do erro de predição só era observada quando ξ era experimentalmente otimizado para uma dada série temporal S, mas mantido constante ao longo do processo de predição de S, à medida que n varia. A tentativa de utilizar o Gradiente Estocástico para atualizar ξ a cada n através de inúmeras iterações do algoritmo Gradiente Estocástico dentro de cada janela $p(n)$, visando reduzir o erro de aproximação a cada n, resultou em um erro de predição alto, embora menor do que para o caso em que variâncias específicas são atribuídas para cada centro.

Na Seção 5.1.3 o Fator de Variância ξ foi experimentalmente ajustado de forma a minimizar o erro de aproximação obtido. A heurística de ajuste adotada naquela seção foi simplesmente aplicar pequenas variações a partir de um valor inicial para ξ, ao longo de um intervalo amplo de valores, e adotar aquele valor de ξ que resulta no menor erro de aproximação.

No contexto de predição a um passo da série temporal S com base em uma janela de predição $p(n)$, poderíamos aplicar esta heurística, mas adotando aquele valor de ξ que resulta no menor erro de predição medido pelo NMSE(n), dado por (3.21). Isto porque, conforme experimentalmente observado, quando a variância σ^2 dos centros é definida pelo fator de proporcionalidade ξ, a redução do erro de aproximação por otimização de ξ implica na concomitante redução do erro de predição. Portanto, como ambos os erros correlacionam-se através de ξ, o uso do NMSE(n) como referência para a otimização experimental de ξ é não só possível como também aconselhável, já que o que se busca

efetivamente é a redução do erro de predição e não a eventual redução do erro de aproximação por janela $p(n)$.

No entanto, esta heurística experimental para otimização de ξ é proibitivamente demorada para séries temporais longas. Assim, é desejável uma abordagem em que o ajuste de ξ seja executado com base no gradiente de descida de alguma função de custo que dê a medida do erro de predição, sujeito à restrição de que o valor de ξ deve ser mantido o máximo possível constante ao longo do processo de predição S, à medida que n varia.

Uma possível aproximação para esta abordagem é iniciar a predição de S com um valor inicial ξ_{init} para o Fator de Variância ξ. A cada n é computado o NMSE(n), dado por (3.21). Se, a um dado n, NMSE(n) > NMSE max o algoritmo LMS [50] entra em ação para otimizar ξ até que NMSE(n) ≤ NMSE min, sendo NMSE max e NMSE min limiares arbitrários e experimentais, ou até que a razão de redução de NMSE(n) torne-se insignificante. Como o denominador de (3.21) cresce com n, o NMSE(n) tende a ser maior para as primeiras janelas de predição $p(n)$. Portanto, o limiar NMSE max tende a ser ultrapassado somente durante o início do processo de predição S. Se ξ_{init} é tal que NMSE(n) > NMSE max, o algoritmo LMS ajustará ξ para que durante as primeiras janelas de predição $p(n)$ o NMSE(n) seja minimizado. Após este intervalo inicial de ajuste, ξ tende a permanecer inalterado ao longo do resto do processo de predição de S. Assim fica obedecida a restrição de que ξ seja mantido o máximo possível constante à medida que n varia. Como o NMSE(n) é cumulativo, o valor de ξ que minimiza o NMSE(n) para os instantes n iniciais tenderá a ser o mesmo para os instantes n finais da série S.

Especificamente, dado $p(n)$ e os resultantes centros $\underline{\psi}_k(n)$, o vetor de pesos sinápticos é obtido por $\underline{w}(n) = \Phi^+(n)\underline{d}(n)$, sendo $\Phi(n)$ calculada com base em $\xi(n)$. Determinada a estimativa da próxima amostra $\hat{u}(n+1) = y(n) + \overline{u_0(n)} = \underline{\varphi}(n)^T \cdot \underline{w}(n) + \overline{u_0(n)}$, testa-se se NMSE($n$) > NMSE max. Se o resultado do teste é verdadeiro, o algoritmo LMS

otimiza $\xi(n)$ objetivando minimizar o NMSE(n). Observe, porém, que durante o processo de otimização de $\xi(n)$ os centros $\underline{\psi}_k(n)$ são mantidos constantes, pois não dependem de $\xi(n)$. O vetor de pesos sinápticos $\underline{w}(n)$, ao contrário, é considerado uma variável ao longo do processo, por depender de ξ. O valor $\xi_{OTIMO}(n)$ resultante da otimização de $\xi(n)$ no instante n é utilizado no instante $n+1$ para definir o Fator de Variância $\xi(n+1)$ tal que $\xi(n+1)=\xi_{OTIMO}(n)$, presumindo-se que, uma vez que $\xi_{OTIMO}(n)$ minimiza NMSE(n) para $p(n)$, também minimizará NMSE($n+1$) para $p(n+1)$.

O desenvolvimento que segue apresenta a derivação do procedimento recursivo adotado para otimizar $\xi(n)$, no instante n, objetivando minimizar NMSE(n) > NMSE max. Considere-se a rede neural RBF descrita anteriormente na Seção 5.2, com saída definida por

$$y(n)=\sum_{k=0}^{K-1} w_k(n)\varphi_k(n)=w_k(n)\exp\left\{-\frac{1}{\xi(n)\sigma^2(n)}\left\|\underline{x}(n)-\underline{\Psi}_k(n)\right\|^2\right\} \tag{5.24}$$

$$\sigma^2(n)=\max_{a,b}\left\{\left\|\underline{\Psi}_a(n)-\underline{\Psi}_b(n)\right\|^2\right\},\quad a,b=0,1,..,K-1 \tag{5.25}$$

Substituindo $\sigma^2(n)$, expresso na Equação (5.25), na Equação (5.24), obtemos

$$y(n)=\sum_{k=0}^{K-1} w_k(n)\exp\left\{\frac{-\left\|\underline{x}(n)-\underline{\Psi}_k(n)\right\|^2}{\xi(n)\max_{a,b}\left[\left\|\underline{\Psi}_a(n)-\underline{\Psi}_b(n)\right\|^2\right]}\right\},\quad a,b=0,1,...,K-1 \tag{5.26}$$

A Equação (5.26) pode ser convenientemente escrita como

$$y(n)=\sum_{k=0}^{K-1} w_k(n)\exp\left\{\frac{-C_k(n)}{\xi(n)}\right\} \tag{5.27}$$

onde

$$C_k(n) = \frac{\left\| \underline{x}(n) - \underline{\Psi}_k(n) \right\|^2}{\max_{a,b} \left[\left\| \underline{\Psi}_a(n) - \underline{\Psi}_b(n) \right\|^2 \right]} \tag{5.28}$$

A minimização do NMSE no instante n será levada a efeito através do algoritmo Gradiente Estocástico, mantendo o indexador n "congelado", isto é, n é mantido constante ao longo do processo de minimização. A cada passo do processo, o Gradiente Estocástico ajusta ξ de forma que o NMSE, após o ajuste, seja menor que o anterior. Portanto, é definida uma trajetória percorrida sobre a curva NMSE(ξ) na busca do mínimo NMSE. Esta trajetória é tal que, a cada instante s, o movimento é feito na direção oposta do gradiente de NMSE(ξ) com respeito à ξ. Enquanto o NMSE é minimizado, n é mantido constante, e o índice s que indexa os pontos da trajetória é incrementado até que $\text{NMSE}(s) \leq \text{NMSE min}$.

A otimização de $\xi(s)$ no instante n será dada, portanto, por

$$\xi(s+1) = \xi(s) - \eta \nabla(s) \tag{5.29}$$

onde s é o índice do passo de recursão, η controla o tamanho da correção incremental do Gradiente Estocástico e $\nabla(s)$ é o gradiente instantâneo dado por

$$\nabla(s) = \frac{\partial\{\text{NMSE}(s)\}}{\partial \xi(s)} \tag{5.30}$$

sendo NMSE(s) o NMSE a ser minimizado no instante n, dado por

$$NMSE(s) = \frac{\sum_{i=0}^{n} (u(i+1) - \hat{u}(i+1))^2}{\sum_{i=0}^{n} (u(i+1) - u(i))^2} = \frac{\sum_{i=0}^{n} \left(u(i+1) - \left[y(s)\right]_i + \overline{u_0(n)}\right]^2}{\sum_{i=0}^{n} (u(i+1) - u(i))^2} \tag{5.31}$$

e como durante a otimização de $\xi(n)$ o vetor de pesos sinápticos $\underline{w}(n)$ é variável e os centros $\underline{\psi}_k(n)$ são constantes, o termo que expressa esta situação será dado por $y(s)|_i$, que, com base em (5.27), é definido por

$$y(s)|_i = \begin{cases} \sum_{k=0}^{K-1} w_k(s)\exp\left\{\dfrac{-C_k(i)}{\xi(s)}\right\}, & i = n \\ \sum_{k=0}^{K-1} w_k(i)\exp\left\{\dfrac{-C_k(i)}{\xi(i)}\right\}, & i < n \end{cases} \tag{5.32}$$

Assim, a função de custo específica que se deseja minimizar no instante n é dada por

$$NMSE(s) = \frac{\sum_{i=0}^{n}\left[u(i+1) - y(s)|_i - \overline{u_0(n)}\right]^2}{\sum_{i=0}^{n}\left(u(i+1) - u(i)\right)^2} \tag{5.33}$$

Aplicando o operador gradiente expresso em (5.30) à expressão do NMSE(s) dada por (5.33), temos

$$\nabla(s) = \frac{\partial}{\partial\xi(s)}\left\{\frac{\sum_{i=0}^{n}\left[u(i+1) - y(s)|_i - \overline{u_0(n)}\right]^2}{\sum_{i=0}^{n}\left(u(i+1) - u(i)\right)^2}\right\} \tag{5.34}$$

Mas, como $\xi(n)$ é otimizado no instante n, o operador $\partial\{\cdot\}$ anula todos os termos do numerador de (5.34) exceto o que ocorre em n, isto é

$$\nabla(s) = \frac{\dfrac{\partial}{\partial\xi(s)}\left[u(n+1) - \sum_{k=0}^{K-1} w_k(s)\exp\left(\dfrac{-C_k(n)}{\xi(s)}\right) - \overline{u_0(n)}\right]^2}{\sum_{i=0}^{n}\left(u(i+1) - u(i)\right)^2} \tag{5.35}$$

mas

$$\frac{\partial}{\partial\xi(s)}\left[u(n+1)-\sum_{k=0}^{K-1}w_k(s)\exp\left(\frac{-C_k(n)}{\xi(s)}\right)-\overline{u_0(n)}\right]^2=\frac{\partial e^2(s)}{\partial\xi(s)}=2e(s)\frac{\partial e(s)}{\partial\xi(s)} \tag{5.36}$$

com

$$e(s)=u(n+1)-\hat{u}(n+1)=u(n+1)-\sum_{k=0}^{K-1}w_k(s)\exp\left(\frac{-C_k(n)}{\xi(s)}\right)-\overline{u_0(n)} \tag{5.37}$$

e, então

$$\frac{\partial e(s)}{\partial\xi(s)}=\frac{\partial}{\partial\xi(s)}\begin{Bmatrix}u(n+1)-w_0(s)e^{\frac{-C_0(n)}{\xi(s)}}-w_1(s)e^{\frac{-C_1(n)}{\xi(s)}}-\ldots\\ \ldots-w_{K-1}(s)e^{\frac{-C_{K-1}(n)}{\xi(s)}}-\overline{u_0(n)}\end{Bmatrix}= \tag{5.38}$$

$$=\begin{Bmatrix}-\left[\frac{w_0(s)C_0(n)}{\xi^2(s)}+\frac{\partial w_0(s)}{\partial\xi(s)}\right]e^{\frac{-C_0(n)}{\xi(s)}}-\left[\frac{w_1(s)C_1(n)}{\xi^2(s)}+\frac{\partial w_1(s)}{\partial\xi(s)}\right]e^{\frac{-C_1(n)}{\xi(s)}}-\ldots\\ \ldots-\left[\frac{w_{K-1}(s)C_{K-1}(n)}{\xi^2(s)}+\frac{\partial w_{K-1}(s)}{\partial\xi(s)}\right]e^{\frac{-C_{K-1}(n)}{\xi(s)}}\end{Bmatrix}=$$

$$=-\sum_{k=0}^{K-1}\left[\frac{w_k(s)C_k(n)}{\xi^2(s)}+\frac{\partial w_k(s)}{\partial\xi(s)}\right]e^{\frac{-C_k(n)}{\xi(s)}}=$$

$$=-\sum_{k=0}^{K-1}\left[\frac{w_k(s)C_k(n)}{\xi^2(s)}+\frac{w_k(s)-w_k(s-1)}{\xi(s)-\xi(s-1)}\right]e^{\frac{-C_k(n)}{\xi(s)}}$$

Note que, na Equação (5.38), o termo $\frac{\partial w_k(s)}{\partial\xi(s)}$ foi aproximado por $\frac{w_k(s)-w_k(s-1)}{\xi(s)-\xi(s-1)}$. Esta aproximação é razoável porque η na Equação (5.29) é sempre um pequeno valor.

Substituindo (5.38), (5.37) e (5.36) em (5.35) temos

$$\nabla(s)=\frac{-2e(s)}{\sum_{i=0}^{n}(u(i+1)-u(i))^2}\sum_{k=0}^{K-1}\left(\frac{w_k(s)C_k(n)}{\xi^2(s)}+\frac{w_k(s)-w_k(s-1)}{\xi(s)-\xi(s-1)}\right)\exp\left(\frac{-C_k(n)}{\xi(s)}\right) \tag{5.39}$$

e substituindo (5.28) em (5.39),

$$\nabla(s)=\frac{-2e(s)}{\sum_{i=0}^{n}(u(i+1)-u(i))^2}\times \tag{5.40}$$

$$\times\sum_{k=0}^{K-1}\left(\frac{w_k(s)}{\xi^2(s)}\frac{\left\|\underline{u}(n)-\underline{\Psi}_k(n)\right\|^2}{\max\limits_{a,b}\left[\left\|\underline{\Psi}_a(n)-\underline{\Psi}_b(n)\right\|^2\right]}+\frac{w_k(s)-w_k(s-1)}{\xi(s)-\xi(s-1)}\right)\exp\left(\frac{-\left\|\underline{u}(n)-\underline{\Psi}_k(n)\right\|^2}{\xi(s)\max\limits_{a,b}\left[\left\|\underline{\Psi}_a(n)-\underline{\Psi}_b(n)\right\|^2\right]}\right)$$

mas, de (5.24) e (5.26)

$$\varphi_k(s)=\exp\left(\frac{-\left\|\underline{u}(n)-\underline{\Psi}_k(n)\right\|^2}{\xi(s)\max\limits_{a,b}\left[\left\|\underline{\Psi}_a(n)-\underline{\Psi}_b(n)\right\|^2\right]}\right) \tag{5.41}$$

portanto (5.40) pode ser expressa por

$$\nabla(s)=\frac{-2e(s)}{\sum_{i=0}^{n}(u(i+1)-u(i))^2}\times \tag{5.42}$$

$$\times\sum_{k=0}^{K-1}\left(\frac{w_k(s)}{\xi^2(s)}\frac{\left\|\underline{u}(n)-\underline{\Psi}_k(n)\right\|^2}{\max\limits_{a,b}\left[\left\|\underline{\Psi}_a(n)-\underline{\Psi}_b(n)\right\|^2\right]}+\frac{w_k(s)-w_k(s-1)}{\xi(s)-\xi(s-1)}\right)\varphi_k(s)$$

ou

$$\nabla(s)=\frac{-2e(s)}{\sum_{i=0}^{n}(u(i+1)-u(i))^2}\times \tag{5.43}$$

$$\times\left[\frac{\sum_{k=0}^{K-1}\varphi_k(s)w_k(s)\left\|\underline{u}(n)-\underline{\Psi}_k(n)\right\|^2}{\xi^2(s)\max_{a,b}\left[\left\|\underline{\Psi}_a(n)-\underline{\Psi}_b(n)\right\|^2\right]}+\sum_{i=0}^{K-1}\varphi_k(s)\frac{w_k(s)-w_k(s-1)}{\xi(s)-\xi(s-1)}\right]$$

Substituindo a Equação (5.43) na Equação (5.29):

$$\xi(s+1)=\xi(s)+\frac{2\eta\, e(s)}{\sum_{i=0}^{n}(u(i+1)-u(i))^2}\times \tag{5.44}$$

$$\times\left[\frac{\sum_{k=0}^{K-1}\varphi_k(s)w_k(s)\left\|\underline{u}(n)-\underline{\Psi}_k(n)\right\|^2}{\xi^2(s)\max_{a,b}\left[\left\|\underline{\Psi}_a(n)-\underline{\Psi}_b(n)\right\|^2\right]}+\sum_{i=0}^{K-1}\varphi_k(s)\frac{w_k(s)-w_k(s-1)}{\xi(s)-\xi(s-1)}\right]$$

Conforme se infere de (5.33), a dependência do NMSE como função de ξ, pode ser expressa no instante n por

$$\mathrm{NMSE}(\xi)=\left(c_1-\sum_{k=0}^{K-1}w_k(\xi)\exp\left\{\frac{-C_k(n)}{\xi}\right\}\right)^2+c_0 \tag{5.45}$$

onde c_0 é uma constante que representa o efeito de todos os termos anteriores ao instante n e c_1 representa o termo $u(n+1)-\overline{u_0(n)}$. Por simplicidade, considerou-se o denominador $\sum_{i=0}^{n}(u(i+1)-u(i))^2$ como sendo unitário.

De (5.45), infere-se que a função de custo $\mathrm{NMSE}(\xi)$ será forçosamente não-quadrática e, em conseqüência, apresentará mínimos locais cujo número cresce com o aumento do número K de centros.

A existência de mínimos locais na função de custo $\mathrm{NMSE}(\xi)$ resulta na probabilidade não nula de a trajetória percorrida pelo gradiente ficar presa em algum mínimo local, situação em que o mínimo global, que efetivamente minimiza $\mathrm{NMSE}(\xi)$, jamais será alcançado. Isto ocorre porque a trajetória do gradiente move-se sempre na direção de maior descida sobre a curva $\mathrm{NMSE}(\xi)$, e, ao entrar em algum "vale local" de $\mathrm{NMSE}(\xi)$, esta ficará presa no ponto de declividade nula no "fundo do vale". Este problema é resolvido através da adição de um fator de inércia ou *momentum* à Equação (5.29), isto é

$$\xi(s+1)=\xi(s)-\eta\nabla(s)+\alpha\cdot\Delta\xi(s-1) \tag{5.46}$$

onde o fator positivo α controla a quantidade de *momentum* aplicada à trajetória e $\Delta\xi(s-1)=\xi(s)-\xi(s-1)$. A interpretação de (5.46) é a seguinte: Assume-se que no instante $s-1$ a trajetória do Gradiente Estocástico esteja descendo a curva $\mathrm{NMSE}(\xi)$ em algum ponto. No instante s a trajetória encontra um mínimo local, o que faz nulo o termo $\eta\nabla(s)$. Sem o termo $\alpha\cdot\Delta\xi(s-1)$, o próximo valor de ξ, isto é, $\xi(s+1)$, seria igual ao valor atual $\xi(s)$ e o algoritmo ficaria eternamente nesta condição. No entanto, note que o termo $\alpha\cdot\Delta\xi(s-1)$ é não-nulo, pois é assumido que o algoritmo não esteja preso a um mínimo local no instante $s-1$. Note também que o sinal algébrico de $\alpha\cdot\Delta\xi(s-1)$ coincide com o sentido de variação de ξ que minimiza $\mathrm{NMSE}(\xi)$. Portanto, este valor não nulo de $\alpha\cdot\Delta\xi(s-1)$ tende a dar continuidade à trajetória de minimização mesmo quando ela coincide com um mínimo local.

Portanto, de (5.43) e (5.46) temos

$$\xi(s+1)=\xi(s)+\alpha\,\Delta\xi(s-1)+\frac{2\eta\,e(s)}{\sum_{i=0}^{n}(u(i+1)-u(i))^2}\times \tag{5.47}$$

$$\times\left[\frac{\sum_{k=0}^{K-1}\varphi_k(s)w_k(s)\left\|\underline{u}(n)-\underline{\Psi}_k(n)\right\|^2}{\xi^2(s)\max_{a,b}\left[\left\|\underline{\Psi}_a(n)-\underline{\Psi}_b(n)\right\|^2\right]}+\sum_{i=0}^{K-1}\varphi_k(s)\frac{w_k(s)-w_k(s-1)}{\xi(s)-\xi(s-1)}\right]$$

5.2.2 Os Pré-Processamentos $\nabla\mathbf{U}$ e $\nabla^2\mathbf{U}$

Nesta tese, conforme já discutido, o processo de predição do elemento $u(n+1)$ imediatamente adiante de uma janela $p(n)$ de N elementos, definida em um instante n sobre uma série temporal S, fundamenta-se no mapeamento não-linear $\Im:\mathbf{U}\in\Re^M\rightarrow y\in\Re$ gerado por uma rede neural RBF. O mapeamento $\Im$ antecipa temporalmente o processo vetorial $\mathbf{U}$ associado à S quando a rede RBF é treinada para que o valor obtido em sua saída y aproxime os elementos seguintes em $p(n)$ a cada vetor $\underline{u}_i\in\Re^M$, $i=0,1,\cdots,L-1$, pertencentes ao conjunto $\mathbf{U}$ descritor do processo $\mathbf{U}$, sendo M a ordem de predição.

O conjunto $\mathbf{U}$ é formado a partir de uma sub-janela J_i de M elementos sobre S localizada sobre a i-ésima posição de $p(n)$, tal que J_i seja totalmente contida por $p(n)$. Ao associar cada sub-janela J_i ao vetor $\underline{u}_i\in\Re^M$, $i=0,1,\cdots,L-1$, $L=N-M+1$, fica definido o conjunto $\mathbf{U}$ de L vetores. Portanto, no instante n, cada vetor $\underline{u}_i\in\Re^M$ é definido por $\underline{u}_i(n)=\underline{u}(n-i)$, sendo $\underline{u}(n-i)=[u(n-i)\ u(n-i-1)\ \cdots\ u(n-i-M+1)]^T$ de tal forma que $\mathbf{U}=\{\underline{u}(n),\underline{u}(n-1),\underline{u}(n-2),\cdots,\underline{u}(n-L+1)\}$.

O conjunto **U** é centralizado em torno de sua média, formando o conjunto **X** de vetores $\underline{x}_i \in \Re^M$ dado por $\underline{x}_i(n) = \underline{x}(n-i) = \underline{u}(n-i) - \overline{\underline{u}(n)}$, onde $\overline{\underline{u}(n)}$ é o vetor média de **U** no instante *n*. A rede RBF é então treinada, isto é, são determinados os vetores centros, a variância dos centros e o vetor de pesos sinápticos, de tal forma que o valor obtido em sua saída y aproxime os elementos $x(n-j+1)$ seguintes a cada vetor $\underline{x}(n-j) = [x(n-j)\ x(n-j-1) \cdots x(n-j-M+1)]^T$, $j = 1,2,\cdots,L-1$.

A estimativa do elemento $u(n+1)$ imediatamente adiante da janela $p(n)$ é aproximada por $\hat{u}(n+1) = y(n) + \overline{u_0(n)}$ onde $\overline{u_0(n)}$ é o componente mais recente do vetor média $\overline{\underline{u}(n)}$ e *y(n)* é a saída da rede neural RBF para o vetor de entrada $\underline{x}(n) = [x(n)\ x(n-1) \cdots\ x(n-M+1)]^T$.

Os parágrafos anteriores resumem o processo básico de predição apresentado nesta tese. No entanto, seria razoável supor que uma determinada série *S* talvez tenha um melhor grau de associação à velocidade da variação do processo vetorial **U** do que simplesmente ao processo **U**. Assim como seria razoável supor que uma outra determinada série *S* talvez tenha um melhor grau de associação à aceleração da variação do processo vetorial **U** do que simplesmente ao processo **U**. Para que estas duas possíveis situações sejam consideradas no processo de predição, introduz-se a seguir os conceitos de pré-processamento $\nabla\mathbf{U}$ e pré-processamento $\nabla^2\mathbf{U}$.

Dado o conjunto **U** com L vetores $\underline{u}_i \in \Re^M$, $\underline{u}_i(n) = \underline{u}(n-i)$, $i = 0,1,\cdots,L-1$, formado no instante *n* a partir da janela $p(n)$, conforme discutido nos parágrafos anteriores, seja o conjunto $\nabla\mathbf{U}$, com $L-1$ vetores $\underline{u}'_j \in \Re^M$, $\underline{u}'_j(n) = \underline{u}'(n-j)$, $j = 0,1,\cdots,L-2$, formado de **U** através de

$$\underline{u}'(n-j) = \underline{u}(n-j) - \underline{u}(n-j-1),\ j = 0,1,\cdots,L-2 \tag{5.48}$$

O conjunto $\nabla\mathbf{U}$ é centralizado em torno de sua média, formando o conjunto $\mathbf{X}$ de vetores $\underline{x}_i \in \Re^M$, a rede RBF é treinada conforme descrito nos parágrafos anteriores e a estimativa do elemento $u'(n+1)$ é então aproximada por

$$\hat{u}'(n+1) = y(n) + \overline{u'_0(n)} \tag{5.49}$$

onde $\overline{u'_0(n)}$ é o componente mais recente do vetor média do conjunto $\nabla\mathbf{U}$. Mas, da lei de formação do conjunto $\nabla\mathbf{U}$ dada por (5.30), fica implícito que $\hat{u}'(n+1) = \hat{u}(n+1) - u(n)$. Aplicando este resultado em (5.49) pode-se obter a estimativa $\hat{u}(n+1)$ do elemento $u(n+1)$ imediatamente adiante da janela $p(n)$, dada por

$$\hat{u}(n+1) = y(n) + u(n) + \overline{u'_0(n)} \tag{5.50}$$

O pré-processamento $\nabla\mathbf{U}$, portanto, obtém a estimativa $\hat{u}(n+1)$ do elemento $u(n+1)$ com base no conjunto de vetores de velocidade do processo $\mathbf{U}$.

Seja, agora, o conjunto $\nabla\mathbf{U}$, com $L-1$ vetores $\underline{u}'_j \in \Re^M$, $\underline{u}'_j(n) = \underline{u}'(n-j)$, $j = 0,1,\cdots,L-2$, formado do conjunto $\mathbf{U}$ conforme descrito anteriormente, e seja o conjunto $\nabla^2\mathbf{U}$, com $L-2$ vetores $\underline{u}''_k \in \Re^M$, $\underline{u}''_k(n) = \underline{u}''(n-k)$, $k = 0,1,\cdots,L-3$, formado de $\nabla\mathbf{U}$ através de

$$\underline{u}''(n-k) = \underline{u}'(n-k) - \underline{u}'(n-k-1), \quad k = 0,1,\cdots,L-3 \tag{5.51}$$

O conjunto $\nabla^2\mathbf{U}$ é centralizado em torno de sua média, formando o conjunto $\mathbf{X}$ de vetores $\underline{x}_i \in \Re^M$, a rede RBF é treinada conforme descrito nos parágrafos anteriores e a estimativa do elemento $u''(n+1)$ é então aproximada por

$$\hat{u}''(n+1) = y(n) + \overline{u''_0(n)} \tag{5.52}$$

onde $\overline{u_0''(n)}$ é o componente mais recente do vetor média do conjunto $\nabla^2\mathbf{U}$. Mas, da lei de formação do conjunto $\nabla^2\mathbf{U}$ dada por (5.51), fica implícito que $\hat{u}''(n+1)=\hat{u}'(n+1)-u'(n)$ e, aplicando este resultado em (5.52), obtém-se

$$\hat{u}'(n+1)=y(n)+u'(n)+\overline{u_0''(n)} \tag{5.53}$$

Mas, lembrando que $\hat{u}'(n+1)=\hat{u}(n+1)-u(n)$ e que $u'(n)=u(n)-u(n-1)$, temos de (5.53) que

$$\hat{u}(n+1)-u(n)=y(n)+u(n)-u(n-1)+\overline{u_0''(n)} \tag{5.54}$$

E assim, de (5.54), a estimativa $\hat{u}(n+1)$ do elemento $u(n+1)$ imediatamente adiante da janela $p(n)$ é dada por

$$\hat{u}(n+1)=y(n)+2u(n)-u(n-1)+\overline{u_0''(n)} \tag{5.55}$$

O pré-processamento $\nabla^2\mathbf{U}$, portanto, obtém a estimativa do elemento $u(n+1)$ com base no conjunto de vetores de aceleração do processo $\mathbf{U}$.

5.2.3 Resultados Experimentais

Nesta seção são apresentados resultados experimentais comparativos da predição não-linear através de redes neurais RBF quanto à heurística de atribuição dos centros das funções de base radial. São comparados os resultados da predição para os casos em que os centros são determinados por DSE, *k-means* e APC, com Fator de Variância ξ ajustado conforme Seção 5.2.1.

Utilizou-se no conjunto de experimentos as séries temporais *Sunspots*, *Dollar2Franc* e *Starlight*, descritas no Capítulo 2.

Dada a série S com N_t amostras, o objetivo é obter o menor NMSE final $\text{NMSE}_f = \text{NMSE}(N_t - 1)$ com o menor N possível, sendo N o número de amostras conhecidas na janela de predição $p(n)$. Para tanto, os parâmetros das heurísticas DSE e *k-means* – a ordem de predição M, o número K de centros Gaussianos e o número de vetores L em $p(n)$, bem como os da heurística APC – a ordem de predição M e o número K de centros Gaussianos, são experimentados até que se obtenha o menor NMSE_f com o menor N.

Para todos os casos, ξ é ajustado pela Equação (5.46) com $\alpha = 0.1$. O limiar de ativação do ajuste de ξ é $\text{NMSE}\max = 1.0$ e o limiar de convergência é $\text{NMSE}\min = 0.9$, conforme descrito na Seção 5.2.1. O valor inicial ξ_{init} de ξ e o passo de adaptação η em (5.46) serão especificados para cada caso.

Nos exemplos que seguem é avaliado o tempo de execução relativo ρ da heurística h de atribuição dos centros – onde h pode referir-se às heurísticas DSE, *k-means* ou APC – adotada para a predição de uma particular série S. Para uma dada série S define-se $\rho(h)$ como a razão entre o tempo de execução da heurística h e o tempo de execução da heurística mais rápida na tarefa computacional de predição de toda a série S.

Nos resultados mostrados nas Figuras 5.9, 5.10 e 5.11 a curva Obs(n) representa a observação da série S e a curva Pred(n) representa a predição de S através da heurística DSE. Respectivamente, nas Tabelas 5.2, 5.3 e 5.4 encontra-se a especificação dos parâmetros e os resultados comparativos obtidos.

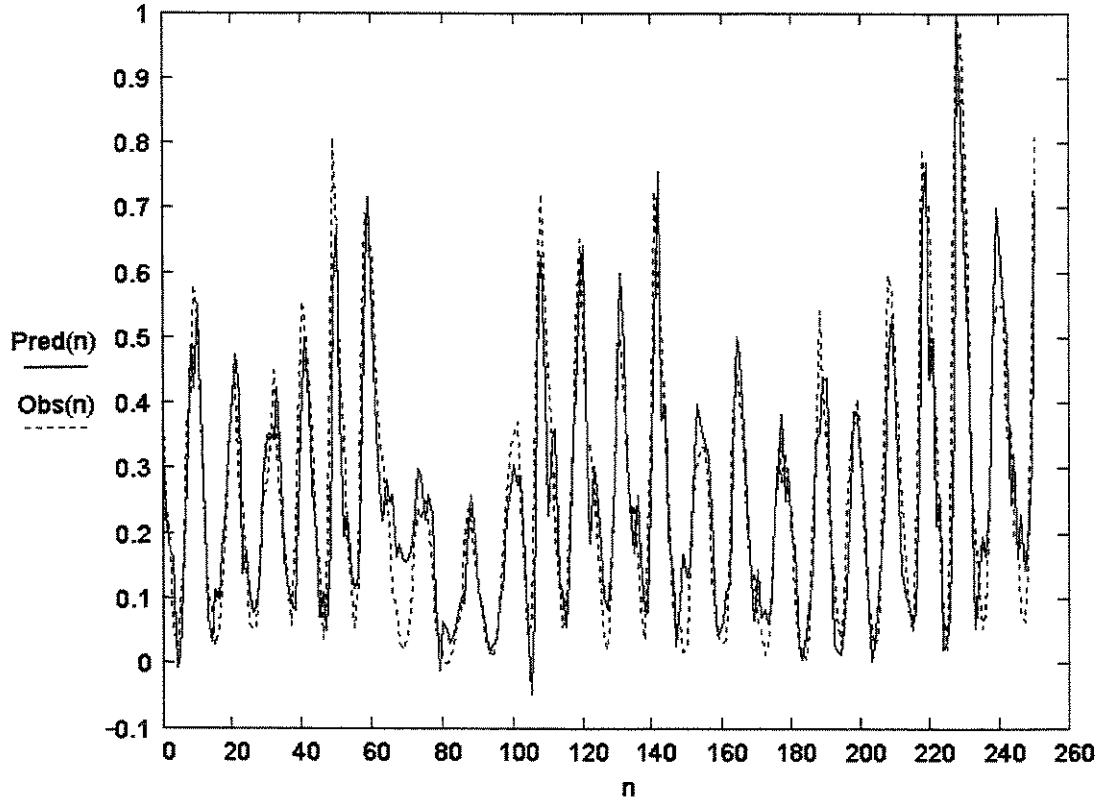

Figura 5.9: Série *Sunspots,* predição por DSE, $N = 29$, $\text{NMSE}_f = 0.484$.

Atribuição de centros via heurística DSE								
M	K	L	ξ_{init}	η	Pré Processamento	$N = M + L - 1$	ρ	NMSE_f
4	4	26	8.0	0.1	–	29	1.0	0.484
Atribuição de centros via heurística *k-means*								
M	K	L	ξ_{init}	η	Pré Processamento	$N = M + L - 1$	ρ	NMSE_f
4	5	26	40.0	0.1	–	29	79.5	0.608
Atribuição de centros via heurística APC								
M	K	ξ_{init}		η		$N = M + K$	ρ	NMSE_f
18	14	1.0		0.01		32	6.9	0.658

Tabela 5.2: Especificação de parâmetros e resultados de predição para a série *Sunspots*.

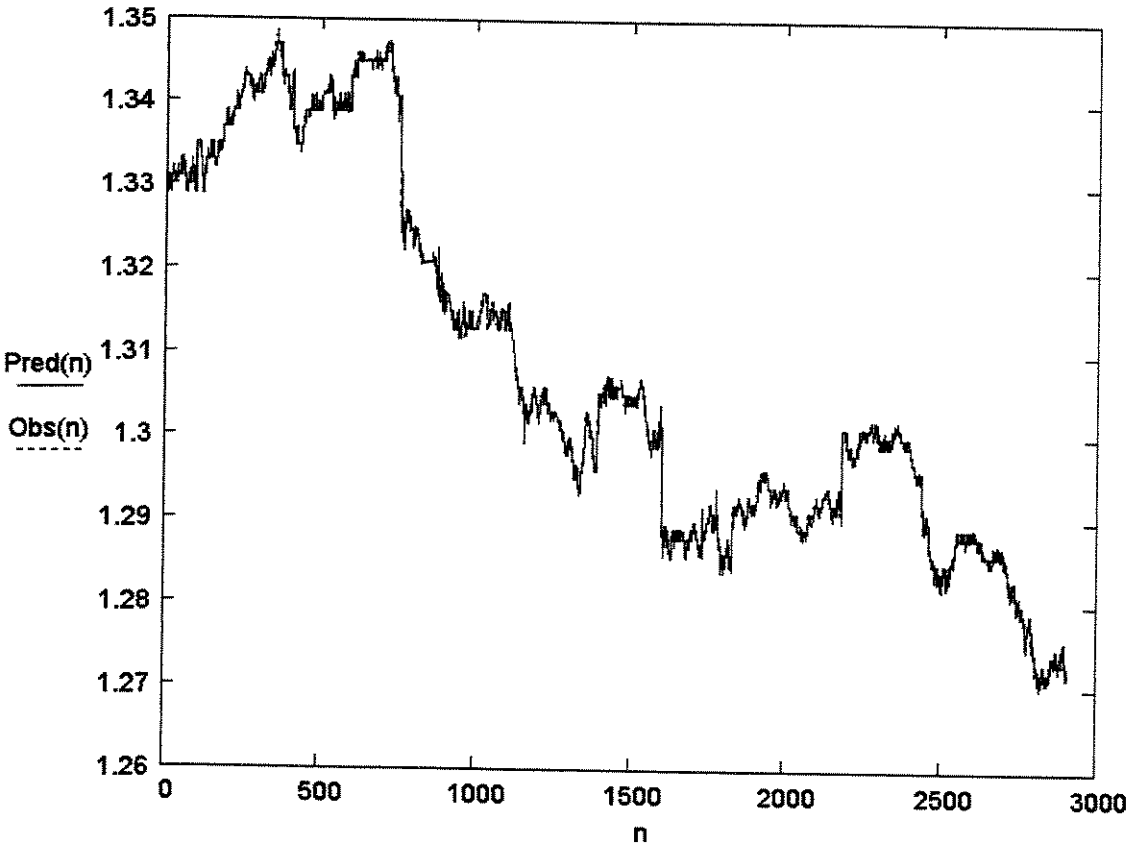

Figura 5.10: Série *Dollar2Franc*, predição por DSE, $N = 91$, $\text{NMSE}_f = 0.984$.

Atribuição de centros via heurística DSE								
M	K	L	ξ_{init}	η	Pré Processamento	$N = M + L - 1$	ρ	NMSE_f
3	3	89	15.0	0.1	∇U	91	2.3	0.984
Atribuição de centros via heurística *k-means*								
M	K	L	ξ_{init}	η	Pré Processamento	$N = M + L - 1$	ρ	NMSE_f
3	4	89	10.0	0.1	∇U	91	318.0	1.02
Atribuição de centros via heurística APC								
M	K	ξ_{init}		η		$N = M + K$	ρ	NMSE_f
88	4	50.0		0.1		92	1.0	1.06

Tabela 5.3: Especificação de parâmetros e resultados de predição para a série *Dollar2Franc*.

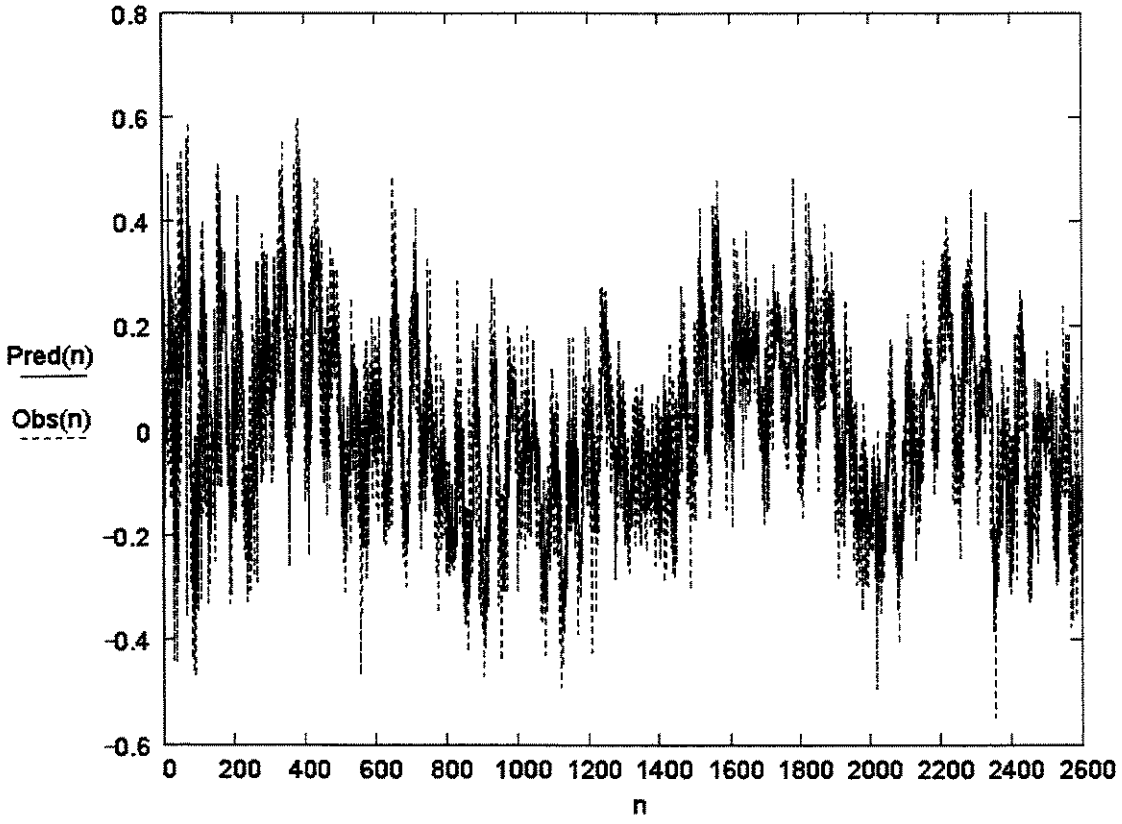

Figura 5.11: Série *Starlight*, predição por DSE, $N = 9$, $\text{NMSE}_f = 0.644$.

Atribuição de centros via heurística DSE								
M	K	L	ξ_{init}	η	Pré Processamento	$N = M + L - 1$	ρ	NMSE_f
5	1	5	1.0	0.1	–	9	1.0	0.644
Atribuição de centros via heurística *k-means*								
M	K	L	ξ_{init}	η	Pré Processamento	$N = M + L - 1$	ρ	NMSE_f
5	2	5	5.0	0.1	–	9	113.6	0.758
Atribuição de centros via heurística APC								
M	K	ξ_{init}		η		$N = M + K$	ρ	NMSE_f
36	4	2.0		0.01		40	2.8	0.664

Tabela 5.4: Especificação de parâmetros e resultados de predição para a série *Starlight*.

Note que, em todas as situações, o NMSE_f obtido com predição por DSE é menor do que o obtido pelas demais heurísticas. Note também que o tempo de execução relativo ρ da predição com atribuição de centros por *k-means* é, no mínimo, uma ordem de grandeza maior do que o tempo de execução das demais heurísticas.

Ainda, é interessante observar que a série *Starlight* apresenta considerável nível de ruído de medição, conforme descrito no Capítulo 2. Em função disto, melhores resultados foram obtidos com $K < M$ – no caso, utilizou-se um único centro $\underline{\psi}_0$ associado ao maior auto-valor λ_0. Isto ocorre porque o ruído de baixa correlação é representado nos sub-espaços de menores auto-valores, e sem uma função de base radial que a ele represente, a rede RBF atua como um *eigen*-filtro [50], desconsiderando o ruído no processo de predição.

5.2.4 Interpretação da Predição por Decomposição em Componentes Principais

A função de auto-correlação de uma série temporal $U(n)=\{u(0),u(1),\cdots,u(N_t-1)\}$ com N_t amostras pode ser aproximada pela Equação (5.56), a qual mede o grau de interdependência no domínio tempo entre as amostras da série $U(n)$ à medida que $U(n)$ é "deslizada" sobre si mesma de um número δ de amostras, $\delta = 0,1\cdots,N_t-1$ [50].

$$R_{uu}(\delta)=\frac{1}{N_t}\sum_{n=0}^{N_t-1}u(n)u(n-\delta) \tag{5.56}$$

A função de auto-correlação $R_{uu}(\delta)$ de uma série temporal $U(n)$ é uma função decrescente com δ, e, em geral, a partir de um valor $\delta=\delta_c$ sofre um significativo decréscimo. O valor δ_c indica o intervalo de instantes amostrais em que os elementos da

série $U(n)$ mantêm influência mútua. Assim, pode-se dizer que $U(n)$ tem uma correlação de δ_c amostras [31].

A predição não-linear a um passo de $U(n)$ por DSE assume que exista um processo estocástico vetorial $\mathbf{U}$ em $\Re^M$, associado ao desenrolar temporal de $U(n)$, representável a partir de uma janela de predição $p(n)$ de N amostras definida sobre $U(n)$.

Conforme anteriormente discutido na Seção 5.2, para representação de $\mathbf{U}$ é definida uma sub-janela J_i de M elementos sobre $U(n)$ localizada sobre a i-ésima posição de $p(n)$ tal que J_i seja totalmente contida por $p(n)$. Ao "deslizar" J_i sobre $p(n)$ fazendo $i = 0,1,\cdots,L-1$, $L = N-M+1$, o conjunto de vetores $\mathbf{U}$ fica definido pelo conjunto de L vetores $\underline{u}_i \in \Re^M$ formado a partir da associação $\underline{u}_i \leftarrow J_i$. É obtida, então, a matriz de covariância $\mathbf{C}$ de $\mathbf{U}$, de cuja *eigen*–decomposição resultam os M sub-espaços de $\mathbf{U}$.

Note a semelhança do processo de obtenção da matriz $\mathbf{C}$ do conjunto $\mathbf{U}$, no qual uma sub-janela J_i é "deslizada" ao longo de L amostras sobre uma janela de predição $p(n)$ definida sobre $U(n)$, com o processo de determinação de $R_{uu}(\delta)$, na qual $U(n)$ é "deslizada" ao longo de N_t amostras sobre si mesma. No caso de $R_{uu}(\delta)$ é expressa a correlação entre os pontos de $U(n)$, enquanto que no caso de $\mathbf{C}$ é expressa a correlação entre regiões J_i de tamanho M definidas em $U(n)$. É esta particularidade que potencialmente habilita $\mathbf{C}$ a captar a correlação entre elementos de $U(n)$ fora da região delimitada por $p(n)$. Por exemplo, dada uma janela $p(n)$, a região J_L, cujo elemento mais recente é $u(n+1)$, não está totalmente contida em $p(n)$. Mas J_L contém elementos comuns à região J_{L-1} imediatamente anterior e totalmente contida em $p(n)$, cujo elemento mais recente é $u(n)$. Como $\mathbf{C}$ expressa a correlação entre regiões J_i, $\mathbf{C}$ expressa informação *a priori* a respeito da região J_L decorrente de J_{L-1}.

Ao efetuar a *eigen*–decomposição da matriz de covariância $\mathbf{C}$ de $\mathbf{U}$ nos M sub-espaços de $\mathbf{U}$, o conjunto de L regiões J_i obtido de $p(n)$ é dividido em M classes

distintas, cada classe correspondendo a um sub-espaço de $\mathbf{U}$. Cada sub-espaço de $\mathbf{U}$ reúne vetores $\underline{u}_i \in \Re^M$ de tamanho semelhante entre si e que alinham-se aproximadamente na direção do sub-espaço, tendo como centro geométrico o vetor média de $\mathbf{U}$. Portanto, como a cada vetor $\underline{u}_i$ associa-se uma região J_i, cada classe reúne regiões semelhantes sob o ponto de vista de seus elementos, tendo como referência a região média (média entre todas as L regiões de M elementos cada) da janela $p(n)$. Como as regiões J_i resultam do "deslizamento" ao longo de L amostras sobre $p(n)$, o conjunto de M classes expressa M possíveis modos de maior correlação temporal entre regiões de tamanho M definidas sobre $p(n)$.

O mapeamento não-linear $\Im : \mathbf{U} \in \Re^M \rightarrow y \in \Re$ gerado pela rede RBF antecipa temporalmente o processo $\mathbf{U}$ quando a RBF é treinada para que o valor obtido em sua saída y aproxime os elementos seguintes em $p(n)$ às regiões J_i. Esta é uma característica comum a todos os métodos de predição a um passo com redes RBF a partir de uma janela de predição $p(n)$.

No entanto, a RBF com centros obtidos através de DSE armazena muito mais informação *a priori* a respeito da série $U(n)$ do que os demais métodos. Isto ocorre porque, no método DSE, o par de parâmetros $(\lambda_m, \underline{e}_m)$, $m = 0,1,\cdots,M-1$, que identifica o m-ésimo sub-espaço, ou o m-ésimo modo de variação do processo $\mathbf{U}$ em torno de sua média, traduz-se na coordenada $\underline{\psi}_m = \underline{e}_m \cdot \sqrt{\lambda_m}$ do m-ésimo centro do conjunto de funções de base radial da rede RBF. Como o conjunto de M sub-espaços expressa M possíveis modos de maior correlação temporal entre regiões de tamanho M definidas sobre $p(n)$, esta informação *a priori* a respeito de $U(n)$ fica implícita nos centros $\underline{\psi}_m$.

Na técnica de predição a um passo de $U(n)$ através de redes RBF com atribuição de centros por DSE, a ordem de predição M, o número de estados K do processo vetorial $\mathbf{U}$ associado a $U(n)$ e o tamanho N da janela de predição $p(n)$ constituem um conjunto de

parâmetros que identificam um modelo de predição. Em outras palavras, cada modelo define uma particular arquitetura da rede RBF .

Portanto, é razoável considerar que, na predição de uma série temporal $U(n)$ haverá distintas combinações dos parâmetros M, K e N que poderão conduzir a diferentes desempenhos de predição.

Como, em geral, $K = M$ e como o número de vetores L que se assume caracterizar o processo $\mathbf{U}$ é dado por $L = N - M + 1$, doravante nesta tese iremos nos referir ao par de parâmetros (M, L) como identificadores de um modelo de predição, associado à uma particular arquitetura da rede RBF.

O bom senso sugere que, para avaliação dos modelos de predição, seja considerado de melhor desempenho aquele modelo (M, L) que resulte em um menor $\mathrm{NMSE_f} = \mathrm{NMSE}(N_t - 1)$ e que simultaneamente necessite do menor número possível de amostras prévias conhecidas $N = L + M - 1$, onde $\mathrm{NMSE}(\cdot)$ é definido pela Equação (3.21) e N_t é o número de elementos na série $U(n)$.

A escolha da arquitetura de uma rede neural que pretenda representar espaços de dados é uma tarefa reconhecidamente árdua. Devido às peculiaridades da nova técnica de predição apresentada nesta tese, no entanto, este problema não é de difícil solução. Na verdade, a abordagem deste problema resultou no aprimoramento de nossa técnica de predição.

Além do fato de particulares modelos de predição (M, L) resultarem em melhor desempenho do que outros, consideramos também que, em séries temporais com características não-estacionárias, um modelo (M, L) pudesse ser mais adequado para a predição de um determinado intervalo da série do que outros.

A escolha de modelos (M, L) diferenciados por intervalos permite acompanhar de forma mais acurada as não-estacionariedades presentes ao longo da série $U(n)$. O princípio aqui é similar ao implícito no uso da matriz de transição $\mathbf{\Phi}$ para predição a um passo: A

matriz Φ armazena o máximo de informação, obtida da janela $p(n)$, sobre o processo que caracteriza $U(n)$. Então, é assumido que a informação armazenada em Φ seja invariante a um passo o que a habilita a estimar o próximo elemento $u(n+1)$ adiante de $p(n)$. A nível de modelo de predição (M,L), podemos executar um procedimento que é um super-conjunto do princípio implícito no uso da matriz de transição Φ para captar a correlação de $U(n)$. Antes de estimar o elemento $u(n+1)$ adiante de $p(n)$, podemos testar um conjunto Ω de modelos (M,L), $\Omega = \{(M_0,L_0),(M_1,L_1),\cdots\}$, e escolher aquele modelo (M_*,L_*) que resulta no menor erro para predizer o elemento conhecido $u(n)$, ainda dentro da janela $p(n)$. Ao utilizar (M_*,L_*) para prever o elemento desconhecido $u(n+1)$ adiante de $p(n)$, estaremos baseando o procedimento na esperança que (M_*,L_*) também seja o modelo mais adequado para prever $u(n+1)$ já que o foi para prever $u(n)$.

Assim, buscamos a situação ideal que seria aquela em que se pudesse selecionar o modelo (M,L) mais adequado para a predição a um passo de uma janela $p(n)$ definida sobre $U(n)$, sob o ponto de vista das estatísticas presentes nesta particular janela.

O próximo capítulo apresenta a solução encontrada para este problema, à qual demos o nome "Janela de Predição Seletiva" ou, resumidamente, JPS.

Capítulo 6

Janela de Predição Seletiva

Conforme discutido ao final do Capítulo 5, a dimensão dos vetores do conjunto de treino M, assim como o número de vetores L que o compõem são parâmetros que definem o modelo de predição (M, L) adotado para a predição por DSE de uma dada série temporal $U(n)$ com N_t amostras, modelo que define uma particular arquitetura para a rede RBF.

Modelos de predição diferentes resultam em diferentes desempenhos de predição. O desempenho da predição é avaliado pelo $\text{NMSE}_f = \text{NMSE}(N_t - 1)$ que resulta da predição com base no conhecimento de N amostras prévias na janela de predição $p(n)$.

A Tabela 6.1 apresenta alguns resultados experimentais da predição da série temporal *Tspc6f*, descrita na Seção 2.6, pela técnica DSE utilizando pré-processamento $\nabla\mathbf{U}$ e Fator de Variância $\xi_{INIT} = 15$, com o objetivo de ilustrar os desempenhos de predição resultantes da utilização de diferentes modelos (M, L). A série *Tspc6f* é originária dos valores de fechamento das ações PNB da Telesp Celular, no período de 01/07/1998 a 16/11/1999, obtidos dos boletins de mercado de ações e será utilizada ao longo desta seção para exemplificar as diversas etapas do algoritmo. A série *Tspc6f* é uma das séries temporais de maior dificuldade de predição dentre as que testamos. Embora os resultados obtidos para a predição desta particular série não sejam os mais expressivos, a série foi escolhida como um exemplo didaticamente ilustrativo por apresentar resultados parciais visíveis ao longo da aplicação de todas as etapas que envolvem a heurística de predição por DSE usando o conceito de Janela de Predição Seletiva.

Modelos (M,L)		NMSE_f	Número N de Amostras Prévias Conhecidas
Dimensão M dos Vetores	Número L de Vetores		
2	89	0.99	90
2	88	0.991	89
4	104	0.997	107
2	82	0.998	83
4	99	0.999	102
3	84	1	86
2	90	1	91
2	84	1	85
2	94	1	95
7	94	1.01	100
2	105	1.01	106
...	...	...	...

Tabela 6.1: Exemplos de desempenhos de predição para a série temporal *Tspc6f*, resultantes da utilização de diferentes modelos (M,L), ordenados em ordem crescente de NMSE_f.

Conforme constatado experimentalmente e observado na Tabela 6.1, alguns modelos efetivamente resultam em melhores resultados de predição do que outros. Observe na Tabela 6.1 que o modelo de predição que utiliza 89 vetores de dimensão 2, isto é, o modelo $(2,89)$, resulta em $\text{NMSE}_f = 0.99$, necessitando do conhecimento prévio de $N = 90$ amostras da série temporal. Já o modelo $(2,105)$, por exemplo, que utiliza 105 vetores de dimensão 2, resulta em um $\text{NMSE}_f = 1.01$, necessitando de $N = 106$ amostras prévias conhecidas. O modelo $(2,105)$ resulta em $\text{NMSE}_f > 1$, que é um resultado pior do que aquele que seria obtido através da predição pela última amostra, não sendo considerado, individualmente, um preditor útil.

Como muitas vezes tratamos do problema da predição de séries temporais não-estacionárias, é razoável supormos que – da mesma forma que alguns modelos (M,L) resultam em melhor desempenho do que outros na predição ao longo de toda a série – possam haver modelos mais adequados para a predição de um determinado intervalo na série temporal do que outros.

A partir da consideração de que mais de um modelo possa ser utilizado ao longo da predição de diferentes intervalos de uma mesma série temporal, o problema imediatamente posto é escolher alguns modelos – dentre o universo de todos os modelos (M,L) possíveis – para que estes modelos selecionados possam ser utilizados de forma conjunta e cooperativa na predição dos diversos intervalos, estatisticamente diferenciados, da série temporal. É preciso, portanto, estabelecer um critério de seleção que permita restringir a dimensão do universo de possíveis soluções – modelos (M,L) – a serem aplicadas à predição.

A tarefa pretendida compreende, portanto:

1- Selecionar um conjunto $\mathbf{\Omega}$ de soluções – conjunto de modelos (M,L) – mais adequadas à tarefa de predizer uma específica série temporal dentre o universo de todas as soluções possíveis.

2- Estabelecer um critério para a operação conjunta e cooperativa destas soluções na tarefa de predição da série em questão.

Por simplificação, chamaremos a predição da série $U(n)$ por DSE, através de um único modelo (M,L), fixo ao longo de todo o processo de predição de $U(n)$, de heurística de predição por modelo fixo, ou quando conveniente, heurística MF.

A característica do problema a ser abordado, no que diz respeito à escolha de modelos mais adequados dentre um universo de possíveis soluções, sugere que o critério de seleção dos modelos (M,L) seja inspirado na filosofia dos Algoritmos Genéticos [12][14][15].

No algoritmo proposto, o universo das possíveis soluções – ou o conjunto de todos os possíveis modelos (M,L) – é visto como uma população de indivíduos. Alguns indivíduos da população serão selecionados para serem mantidos ou eliminados da população, operação esta levada a termo através da avaliação comparativa da aptidão absoluta de cada modelo. Os indivíduos selecionados dentre a população global para serem mantidos formarão um conjunto $\mathbf{\Omega}$ de modelos (M,L), $\mathbf{\Omega} = \{(M_0,L_0),(M_1,L_1),\cdots\}$,

considerados os mais aptos para a tarefa de predição da série temporal, conforme veremos na Seção 6.2.

Para a escolha do modelo (M_*, L_*) a ser eleito para a predição da amostra $u(n+1)$ seguinte à janela $p(n)$ é avaliado o erro de predição de cada um dos modelos (M, L) que compõem o conjunto Ω na predição da amostra $u(n)$, já observada (conhecida), seguinte à janela $p(n-1)$. O modelo do conjunto $\Omega = \{(M_0, L_0), (M_1, L_1), \cdots\}$ que tiver apresentado o menor erro quadrático de predição $[u(n) - \hat{u}(n)]^2$ é considerado como o modelo (M_*, L_*) que será utilizado para a predição da amostra $u(n+1)$ seguinte à janela $p(n)$. A utilização de (M_*, L_*) para prever a amostra desconhecida $u(n+1)$ adiante de $p(n)$, fundamenta-se na premissa de que, em havendo correlação temporal entre intervalos da série, (M_*, L_*) seja o modelo mais adequado para prever $u(n+1)$ já que o foi para prever $u(n)$.

A heurística de predição da série temporal $U(n)$ na qual é selecionado o modelo $(M_*, L_*) \in \Omega$ mais adequado para a predição a um passo de uma janela $p(n)$ sobre $U(n)$, a partir da janela $p(n-1)$ e do conjunto Ω, será denominada Janela de Predição Seletiva ou, resumidamente, JPS.

Ao operador que define a predição da série temporal $U(n)$ com base na heurística JPS e com base no conjunto Ω iremos nos referir como JPS$\{\Omega, U\}$. Diferentemente da heurística MF, na qual o Fator de Variância ξ é atualizado pelo algoritmo LMS (Seção 5.2.1) a partir de um valor inicial ξ_{INIT}, a heurística JPS utiliza um parâmentro ξ a ser especificado que é constante durante todo o processo de predição.

6.1 Semelhanças e Diferenças entre a Heurística de Definição do Conjunto Ω e os Algoritmos Genéticos

Em um sentido amplo do termo, um algoritmo genético é qualquer modelo que trata uma população de soluções de um determinado problema, aplicando determinadas operações aos membros da população, de forma que a população resultante seja composta de indivíduos mais aptos do que a população inicial (ou até que se selecione apenas o indivíduo mais apto) [12][14][15].

Cada um dos indivíduos que compõem a população representa uma solução completa do problema que se deseja tratar e tem avaliada a sua aptidão na execução da tarefa. A avaliação é feita não apenas de forma absoluta (*evaluation*), mas também de forma relativa aos demais indivíduos da população (*fitness*) [12][14][15].

Em um algoritmo genético são alocadas propriedades seletivas e reprodutivas, de tal forma que os indivíduos que representem melhores soluções para o problema alvo tenham mais chances de persistir ou de se reproduzir no conjunto de soluções, do que aqueles indivíduos que representem piores soluções.

À semelhança dos algoritmos genéticos, a heurística de definição do conjunto Ω de modelos (M,L), a ser definida na Seção 6.2, opera sobre uma população de indivíduos, que são todos os modelos possíveis, buscando selecionar um conjunto Ω de modelos mais aptos, no sentido de adequação da população à tarefa.

Cada modelo (M,L), que representa uma solução completa ao problema de predição da série temporal pela heurística MF é, portanto, um indivíduo (M,L,NMSE_f) da população, descrito por três parâmetros: o número L de vetores envolvidos na predição, a dimensão M de cada um destes vetores e a medida da aptidão absoluta deste indivíduo na tarefa de predição a qual varia inversamente ao NMSE_f do indivíduo.

O algoritmo de definição de Ω não busca definir uma única solução para o problema que está sendo tratado, mas sim, a exemplo dos algoritmos genéticos, busca

encorajar as melhores soluções a subsistirem no conjunto de soluções, por meio de – específicas – operações seletivas. Como veremos na Seção 6.2, ao longo do processo de definição de Ω, membros da população de indivíduos serão continuamente selecionados para serem mantidos ou eliminados da população, por meio da avaliação comparativa de seu desempenho.

A exemplo das semelhanças, é importante, também, salientar as diferenças básicas entre a heurística de definição de Ω e os Algoritmos Genéticos. As semelhanças, na verdade, estão relacionadas à caracterização do problema e não à operação do algoritmo.

Neste trabalho, diferentemente do praticado em Algoritmos Genéticos, a medida da aptidão de um indivíduo na tarefa de predição – o inverso do NMSE_f do indivíduo – é sempre uma medida absoluta, e não uma medida relativa à aptidão dos demais indivíduos. Dito de outra forma, diferentemente do conceito de *fitness* usado em Algoritmos Genéticos, esta medida não é normalizada pela aptidão média da população.

Outra diferença entre a heurística de definição de Ω e os Algoritmos Genéticos é o fato de que a população de indivíduos (M, L, NMSE_f) não é gerada de forma aleatória e sim compreende todos os possíveis modelos a serem explorados na predição de uma específica série temporal. Em razão de a população já compreender todo o espaço de soluções, não são executadas operações como cruzamento e mutação, que levam à geração de novos indivíduos.

Por rigor científico, consideramos muito importante afirmar que a heurística de definição de Ω é inspirada – e apenas inspirada – na filosofia dos Algoritmos Genéticos, mas não se trata de um Algoritmo Genético. Devido às fortes semelhanças entre alguns conceitos, no entanto, toma-se a liberdade nesta tese de emprestar dos Algoritmos Genéticos, alguns jargões, por se mostrarem extremamente adequados à caracterização do problema aqui tratado.

6.2 A Heurística de Definição de Ω e o Algoritmo JPS

A heurística de definição de Ω busca identificar o conjunto de modelos (M,L) mais adequado para a predição a um passo de uma janela $p(n)$ definida sobre uma série temporal $U(n)$, sob o ponto de vista das estatísticas presentes nesta particular janela e sob o ponto de vista de todas as outras estatísticas presentes nas demais janelas $p(n)$ que ocorrem ao longo do processo de predição de $U(n)$.

O custo computacional da definição de Ω para predição de $U(n)$ sob heurística JPS é algo maior do que o custo computacional da predição de $U(n)$ sob heurística MF. No entanto, uma vez definido Ω, a predição sob heurística JPS apresenta maior capacidade para lidar com as não-estacionaridades de $U(n)$ do que a predição sob heurística MF. Como será mostrado na Seção 6.3 , a capacidade de generalização da heurística JPS é tal que, definido Ω para um intervalo da série a ser predita, ao ser usado o mesmo Ω para predizer outro – desconhecido – intervalo, o desempenho que resulta é semelhante.

O primeiro passo na heurística de definição de Ω é a geração da população de indivíduos $(M,L,\mathrm{NMSE_f})$ que compõem o universo de modelos (M,L) possíveis. Especificamente, os indivíduos da População I são associados a um conjunto Ω^{I} de modelos (M,L) através de $(M,L)\leftarrow(M,L,\mathrm{NMSE_f})$, tal que Ω^{I} compreenda todos os possíveis modelos considerados.

Por exemplo, a Tabela 6.2 ilustra o espectro dos modelos usados na predição da série temporal *Tspc6f* com $\xi_{INIT}=15$ e pré-processamento $\nabla\mathbf{U}$. O espectro de modelos usados corresponde à população inicial de indivíduos – ou População I – sobre a qual a heurística irá operar. A População I compreende o conjunto de todos os possíveis indivíduos, para o caso em que $L=20,21,\cdots,149$ e $M=2,3,\cdots,12$. Por serem 1430 os indivíduos que compõem a população inicial para esta série temporal, apenas alguns destes

modelos são mostrados nesta seção, na Tabela 6.2. O Apêndice C apresenta a Tabela 6.2 na íntegra.

Indivíduo (M, L, NMSE_f)		
Modelos (M, L)		NMSE_f
Dimensão M dos vetores	Número L de vetores	(Heurística MF)
2	20	1.15
2	21	1.08
2	22	1.1
2	23	1.1
2	24	1.09
2	25	1.08
2	26	1.04
2	27	1.05
2	28	1.05
2	29	1.07
2	30	1.03
2	31	1.05
2	32	1.08
2	33	1.08
2	34	1.05
2	35	1.06
2	36	1.07
.....	.....	.....
12	143	1.18
12	144	1.21
12	145	1.18
12	146	1.2
12	147	1.14
12	148	1.19
12	149	1.17

Tabela 6.2: População inicial de indivíduos para a série temporal *Tspc6f*, formada por todos os possíveis modelos (M, L), definidos por $L = 20,21,\cdots,149$ **e** $M = 2,3,\cdots,12$. Note que esta população de indivíduos implica que o número N de amostras prévias conhecidas na série assumirá os valores $N = 21, 22,\cdots, 160$, já que $N = L + M - 1$.

Estabelecida a população inicial de modelos – População I – a primeira operação seletiva é eliminar desta população, temporariamente, aqueles indivíduos cujo $\mathrm{NMSE_f}$ ultrapasse um determinado valor. É importante lembrar que, quanto maior o $\mathrm{NMSE_f}$ menor a aptidão do indivíduo em questão para efeito de predição. Este procedimento é equivalente a aplicar um corte à superfície gerada pela função $\mathrm{NMSE_f}(M,L)$ representada graficamente em $\Re^3$ sobre o domínio (M,L) em $\Re^2$, excluindo os indivíduos para os quais $\mathrm{NMSE_f}(M,L) > \mathrm{NMSE_c}$, sendo $\mathrm{NMSE_c}$ o valor de NMSE que define o corte na superfície.

A Tabela 6.3 apresenta, para a mesma série temporal, a população inicial de indivíduos em ordem ascendente de $\mathrm{NMSE_f}$ e o corte aplicado. Assim como adotado para a Tabela 6.2, por serem 1430 os indivíduos que compõem a população inicial para esta série temporal, apenas alguns destes modelos são mostrados nesta seção, na Tabela 6.3. O Apêndice C apresenta a Tabela 6.3 na íntegra. A Figura 6.1 ilustra a superfície espectral de modelos representativa da população inicial.

Os indivíduos da População II, resultantes do corte $\mathrm{NMSE_c} = 1$ na População I, são associados a um conjunto $\boldsymbol{\Omega}^{\mathrm{II}}$ de modelos (M,L) através de $(M,L) \leftarrow (M,L,\mathrm{NMSE_f})$, tal que $\boldsymbol{\Omega}^{\mathrm{II}} = \{(2,89),(2,88),(4,104),(2,82),(4,99),(3,84),(2,90),(2,84),(2,94)\}$.

A série *Tspc6f* é, então, predita através da heurística JPS com base no conjunto $\boldsymbol{\Omega}^{\mathrm{II}}$ e utilizando pré-processamento $\nabla\mathbf{U}$ com $\xi = 15$, isto é, é executada a operação definida pelo operador $\mathrm{JPS}\{\boldsymbol{\Omega}^{\mathrm{II}}, Tspc6f\}$, conforme já discutido anteriormente. O $\mathrm{NMSE_f}$ resultante de $\mathrm{JPS}\{\boldsymbol{\Omega}^{\mathrm{II}}, Tspc6f\}$, a ser referido como $\mathrm{NMSE_f^{II}}$, será utilizado na próxima etapa da heurística de definição de $\boldsymbol{\Omega}$.

Cada etapa da heurística de definição de $\boldsymbol{\Omega}$ será avaliada com respeito à curva de NMSE(n) obtida da predição de série *Tspc6f* sob heurística MF, utilizando o modelo (M,L) que corresponde ao melhor indivíduo $(M,L,\mathrm{NMSE_f}) = (2,89,0.99)$ da População I, conforme Tabela 6.3.

Indivíduo (M, L, NMSE_f)		
Dimensão M dos Vetores	Número L de Vetores	NMSE_f (Heurística MF)
2	**89**	**0.99**
2	88	0.991
4	104	0.997
2	82	0.998
4	99	0.999
3	84	1
2	90	1
2	84	1
2	94	1
3	97	1.01
7	94	1.01
2	105	1.01
2	100	1.01
2	104	1.01
3	101	1.01
3	91	1.01
3	88	1.01
.....	.....	.....
12	23	2.14
10	22	2.24
9	20	2.29
11	22	2.57
10	20	2.6
11	21	2.73
12	20	3.65

Tabela 6.3: Série *Tspc6f* – População I ordenada em ordem crescente de NMSE_f e o corte seletivo nos indivíduos desta população para $\text{NMSE}_c = 1$, formando a População II – elementos em realce na tabela. O algoritmo de ordenação utilizado é o *Quicksort* [59]. Observe-se que o indivíduo que ocupa a primeira posição na Tabela 6.3, cujas características são $(M, L, \text{NMSE}_f) = (2,89,0.99)$ – destacado em negrito – é aquele que apresenta melhor desempenho de predição, sob a heurística de predição MF.

A Figura 6.2 apresenta a comparação entre a curva de $\text{NMSE}(n)$ obtida com o modelo $(2,89)$ sob heurística MF – traço contínuo – e a curva $\text{NMSE}^{\text{II}}(n)$ resultante de $\text{JPS}\{\boldsymbol{\Omega}^{\text{II}}, Tspc6f\}$ – tracejado.

Observe que nesta etapa da heurística de definição de $\boldsymbol{\Omega}$, $\text{NMSE}_f^{\text{II}} = 0.999$, resultante de $\text{JPS}\{\boldsymbol{\Omega}^{\text{II}}, Tspc6f\}$, ainda é maior do que o NMSE_f resultante da predição sob heurística MF ($\text{NMSE}_f = 0.99$). A predição $\text{JPS}\{\boldsymbol{\Omega}^{\text{II}}, Tspc6f\}$ tem como fundamento a seleção de diferentes modelos (M, L) em $\boldsymbol{\Omega}^{\text{II}}$ para diferentes intervalos de *Tspc6f* de forma a acompanhar mais acuradamente as não-estacionariedades presentes ao longo da série. Portanto, é possível que alguns dos modelos incluídos na População II (isto é, alguns modelos de $\boldsymbol{\Omega}^{\text{II}}$) não sejam adequados às estatísticas dos diferentes intervalos da série e possam estar prejudicando o desempenho da predição $\text{JPS}\{\boldsymbol{\Omega}^{\text{II}}, Tspc6f\}$.

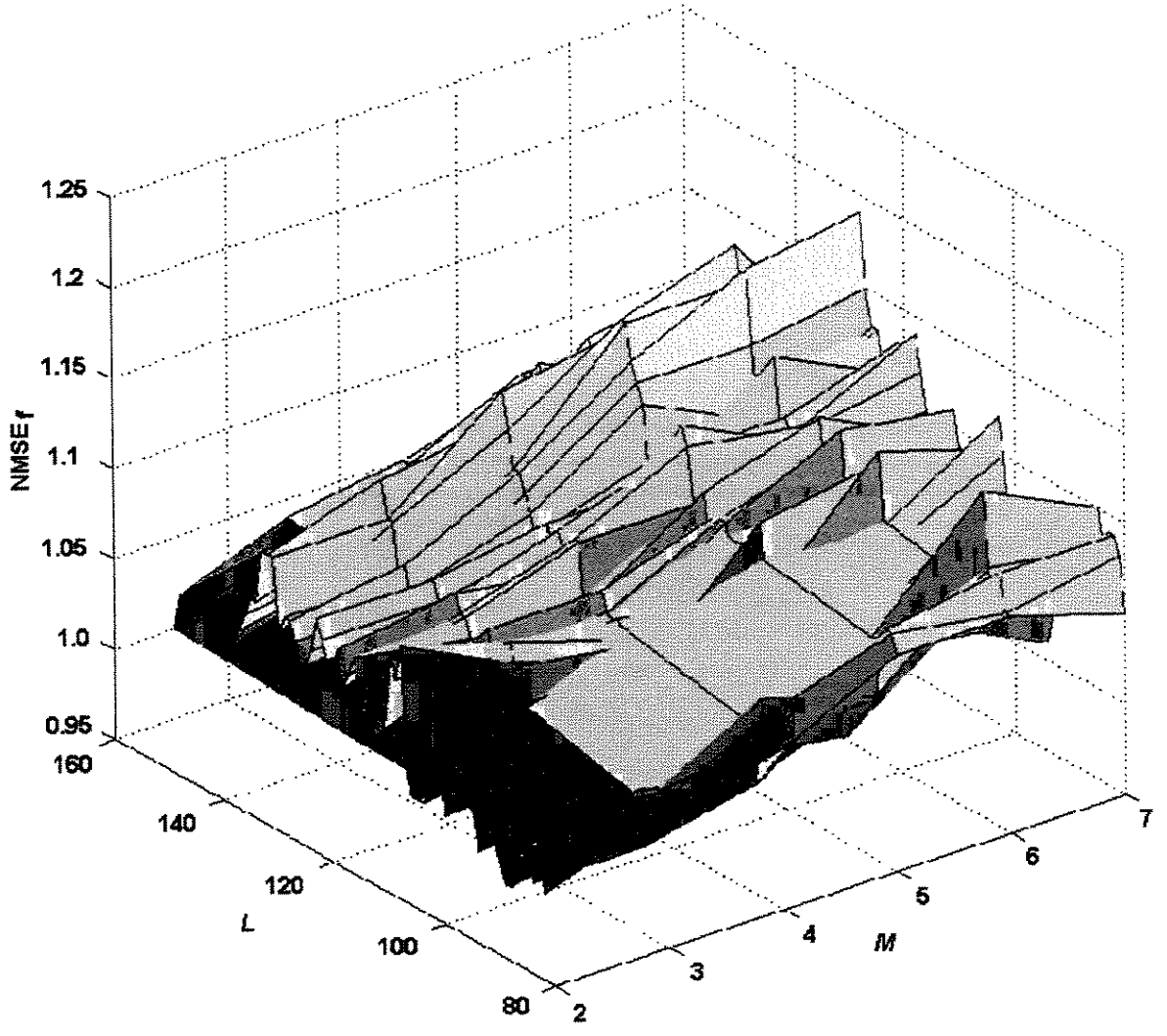

Figura 6.1: Série *Tspc6f* – Superfície $\text{NMSE}_f(M, L)$.

Em conseqüência, é razoável propor que a próxima decisão a ser tomada na heurística de definição de Ω inclua um critério seletivo que vise estabelecer quais modelos em Ω^{II} foram os menos aptos na tarefa de predição através de JPS$\{\Omega^{II}, Tspc6f\}$, para que sejam eliminados.

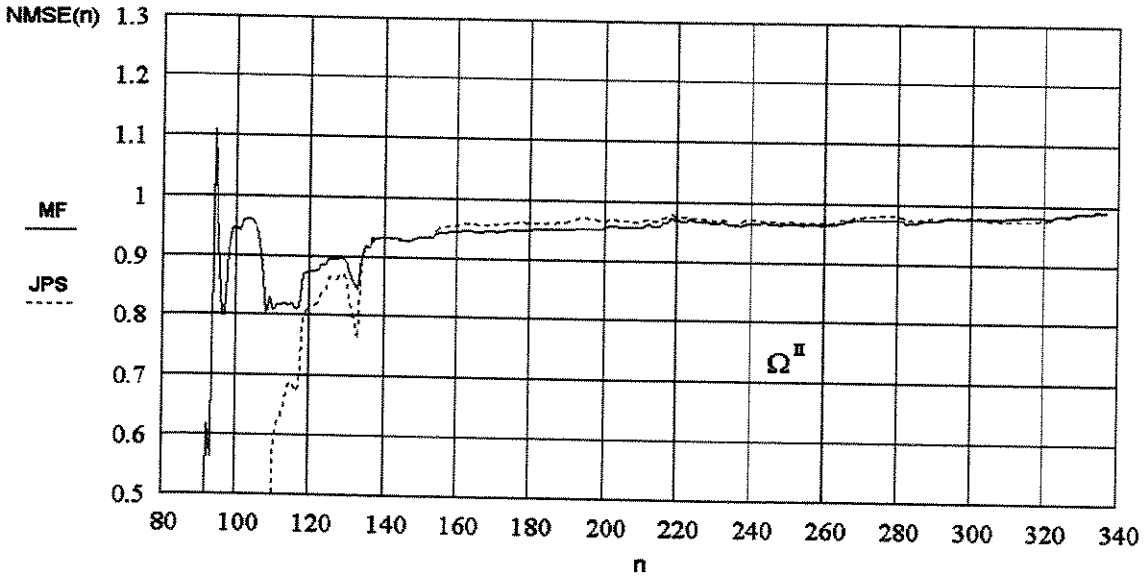

Figura 6.2: Série *Tspc6f* – Comparação entre a curva de NMSE(n) obtida com o modelo $(2,89)$ sob heurística MF – traço contínuo – e a curva NMSE^{II}(n) resultante de JPS$\{\Omega^{II}, Tspc6f\}$ – tracejado.

O novo critério seletivo a ser aplicado à População II é descrito como segue: A predição de *Tspc6f* através da heurística JPS é realizada utilizando todos os modelos em Ω^{II}, exceto um particular modelo (M_p, L_p), isto é, é efetuada a predição

$$\text{JPS}\{\Omega^{II} - \{(M_p, L_p)\}, Tspc6f\},\ p = 0,1,\cdots,\#\{\Omega^{II}\}-1 \qquad (6.57)$$

sendo $\#\{\cdot\}$ o operador que representa o número de elementos do conjunto argumento (operador cardinal) .

A cada predição p de *Tspc6f* por JPS, em que o p-ésimo modelo (M_p, L_p) não é considerado, é obtido um $\mathrm{NMSE}_{\mathrm{f}}^{p}$ resultante da p-ésima predição, operação que representaremos por

$$\mathrm{NMSE}_{\mathrm{f}}^{p} \leftarrow \mathrm{JPS}\left\{\boldsymbol{\Omega}^{\mathrm{II}} - \left\{(M_p, L_p)\right\}, Tspc6f\right\} \tag{6.58}$$

Aquele modelo (M_p, L_p) de cuja exclusão de $\boldsymbol{\Omega}^{\mathrm{II}}$ resulte o menor erro final $\mathrm{NMSE}_{\mathrm{f}}^{p}$, após as $\#\left\{\boldsymbol{\Omega}^{\mathrm{II}}\right\}$ predições pela heurística JPS, define o NMSE mínimo final $\mathrm{NMSE}_{\mathrm{f}}^{\mathrm{MIN}}$ resultante de $\#\left\{\boldsymbol{\Omega}^{\mathrm{II}}\right\}$ operações (6.58), isto é

$$\mathrm{NMSE}_{\mathrm{f}}^{\mathrm{MIN}} = \arg\min_{p}\left\{\mathrm{NMSE}_{\mathrm{f}}^{p}\right\},\ p = 0,1,\cdots,\#\left\{\boldsymbol{\Omega}^{\mathrm{II}}\right\} - 1 \tag{6.59}$$

O modelo $(M_{p\min}, L_{p\min})$, de cuja exclusão de $\boldsymbol{\Omega}^{\mathrm{II}}$ resulta o $\mathrm{NMSE}_{\mathrm{f}}^{\mathrm{MIN}}$, após as $\#\left\{\boldsymbol{\Omega}^{\mathrm{II}}\right\}$ predições pela heurística JPS, é considerado o menos apto na tarefa de predição através de $\mathrm{JPS}\left\{\boldsymbol{\Omega}^{\mathrm{II}}, Tspc6f\right\}$.

É necessário decidir se o modelo $(M_{p\min}, L_{p\min})$ considerado menos apto será excluído definitivamente ou não de $\boldsymbol{\Omega}^{\mathrm{II}}$. O critério de decisão é tal que se,

$$\mathrm{NMSE}_{\mathrm{f}}^{\mathrm{MIN}} < \mathrm{NMSE}_{\mathrm{d}} \Rightarrow \boldsymbol{\Omega}^{\mathrm{II}} = \boldsymbol{\Omega}^{\mathrm{II}} - \left\{(M_{p\min}, L_{p\min})\right\} \tag{6.60}$$

onde inicializa-se $\mathrm{NMSE}_{\mathrm{d}} = \mathrm{NMSE}_{\mathrm{f}}^{\mathrm{II}}$, e, nas demais decisões para exclusão definitiva de $(M_{p\min}, L_{p\min})$ é atribuído a $\mathrm{NMSE}_{\mathrm{d}}$ o valor de $\mathrm{NMSE}_{\mathrm{f}}^{\mathrm{MIN}}$ resultante da decisão anterior.

Tendo sido excluído o particular elemento (M_p, L_p) que estava prejudicando o desempenho da predição $\mathrm{JPS}\left\{\boldsymbol{\Omega}^{\mathrm{II}}, Tspc6f\right\}$, isto é, $(M_{p\min}, L_{p\min})$, a próxima recursão consiste na repetição de todo o processo descrito pelas Equações (6.57) a (6.60), tendo por população inicial – a cada recursão – o conjunto de modelos $\boldsymbol{\Omega}^{\mathrm{II}} = \boldsymbol{\Omega}^{\mathrm{II}} - \left\{(M_{p\min}, L_{p\min})\right\}$.

A recursão é repetida até que a exclusão definitiva de algum particular (M_p, L_p) não resulte em $\text{NMSE}_f^{\text{MIN}} < \text{NMSE}_d$, sendo NMSE_d o valor de $\text{NMSE}_f^{\text{MIN}}$ resultante da recursão anterior.

Caso a condição $\text{NMSE}_f^{\text{MIN}} < \text{NMSE}_d$ não seja satisfeita em conseqüência da exclusão de qualquer um dos modelos pertencentes à $\boldsymbol{\Omega}^{\text{II}}$, conclui-se que nenhum destes modelos esteja prejudicando o desempenho da predição $\text{JPS}\{\boldsymbol{\Omega}^{\text{II}}, Tspc6f\}$ e, portanto, nenhum deles será excluído da População II.

A Tabela 6.4 apresenta o exemplo, para a predição da série *Tspc6f*, do processo de eliminação dos indivíduos menos aptos na tarefa da predição $\text{JPS}\{\boldsymbol{\Omega}^{\text{II}}, Tspc6f\}$. As colunas 1 a 4 apresentam quatro recursões desta etapa do algoritmo. Os indivíduos assinalados nas três primeiras colunas são aqueles efetivamente eliminados da População II a cada conjunto de predições $\text{JPS}\{\boldsymbol{\Omega}^{\text{II}} - \{(M_p, L_p)\}, Tspc6f\}$, $p = 0,1,\cdots,\#\{\boldsymbol{\Omega}^{\text{II}}\}-1$, respectivo a cada coluna. Como pode ser observado na tabela, a eliminação de cada um destes indivíduos leva progressivamente à diminuição do $\text{NMSE}_f^{\text{MIN}}$, com relação ao $\text{NMSE}_f^{\text{II}}$.

Observe na coluna 1 da Tabela 6.4, que o elemento $(2,90)$ é definitivamente excluído de $\boldsymbol{\Omega}^{\text{II}}$ porque $(\text{NMSE}_f^{\text{MIN}} = 0.9874) < (\text{NMSE}_f^{\text{II}} = 1.017)$. Observe também que o modelo $(2,88)$, assinalado na coluna 4 da tabela, cuja exclusão resulta em $\text{NMSE}_f^{\text{MIN}} = 0.9664$, não será eliminado da população, pois, apesar de a sua exclusão do grupo ser a exclusão que conduz ao menor $\text{NMSE}_f^{\text{MIN}}$, o valor atingido não é inferior ao atingido pela recursão anterior (coluna 3, modelo $(2,94)$, $\text{NMSE}_f^{\text{MIN}}$=0.9661), evento que indica o final das recursões.

A Tabela 6.5 apresenta, para o mesmo exemplo, a População II após a totalidade do procedimento especificado na Tabela 6.4. Na Tabela 6.5 estão assinalados os elementos selecionados para exclusão definitiva. A exclusão dos referidos elementos na População II define a População III, associada ao conjunto $\boldsymbol{\Omega}^{\text{III}}$. Note que o NMSE_f especificado na

Tabela 6.5 é o parâmetro do indivíduo (M, L, NMSE_f). Este é o NMSE_f resultante da predição de *Tspc6f* sob a heurística MF, e não mais o $\text{NMSE}_f^{\text{MIN}}$ resultante da predição de *Tspc6f* sob a heurística JPS.

Coluna 1			Coluna 2			Coluna 3			Coluna 4		
<u>População Base:</u> População II			<u>População Base:</u> População II exceto modelo (2,90).			<u>População Base:</u> População II exceto modelos (2,90) e (4,104).			<u>População Base:</u> População II exceto modelos (2,90), (4,104) e (2,94).		
Indivíduo Excluído		NMSE_f^p	Indivíduo Excluído		NMSE_f^p	Indivíduo Excluído		NMSE_f^p	Indivíduo Excluído		NMSE_f^p
2	89	1	2	89	0.9948	2	89	0.978	2	89	0.9754
2	88	0.9933	2	88	0.9889	2	88	0.97	2	88	0.9664
2	82	0.996	2	82	0.99	2	82	0.9741	2	82	0.9705
2	90	0.9874	2	84	0.9888	2	84	0.9698	2	84	0.9674
2	84	0.9933	2	94	0.9847	2	94	0.9661	3	84	0.9784
2	94	0.99	3	84	1.002	3	84	0.9828	4	99	0.9859
3	84	1.001	4	104	0.9845	4	99	0.9888			
4	104	0.993	4	99	0.9935						
4	99	1									

Tabela 6.4: Eliminação dos indivíduos menos aptos no processo de predição $\text{JPS}\{\boldsymbol{\Omega}^{\text{II}}, Tspc6f\}$. Observe que os elementos destacados nas colunas 1 a 4 são aqueles cuja exclusão implicará no menor NMSE_f^p a cada conjunto de predições $\text{JPS}\{\boldsymbol{\Omega}^{\text{II}} - \{(M_p, L_p)\}, Tspc6f\}$, $p = 0,1,\cdots,\#\{\boldsymbol{\Omega}^{\text{II}}\}-1$, respectivo a cada coluna.

A série *Tspc6f* é, então, predita através da heurística JPS com base no recém obtido conjunto $\boldsymbol{\Omega}^{\text{III}}$ e utilizando pré-processamento ∇U com $\xi = 15$, isto é, é executada a operação definida pelo operador $\text{JPS}\{\boldsymbol{\Omega}^{\text{III}}, Tspc6f\}$. O NMSE_f resultante de $\text{JPS}\{\boldsymbol{\Omega}^{\text{III}}, Tspc6f\}$, a ser referido como $\text{NMSE}_f^{\text{III}}$, será utilizado na próxima etapa da heurística de definição de $\boldsymbol{\Omega}$.

Indivíduo (M, L, NMSE_f)			
Dimensão M dos Vetores	Número L de Vetores	NMSE_f (Heurística MF)	
2	89	0.99	
2	88	0.991	
4	104	0.997	←
2	82	0.998	
4	99	0.999	
3	84	1	
2	90	1	←
2	84	1	
2	94	1	←

Tabela 6.5: População II, na qual os indivíduos assinalados, ao serem definitivamente excluídos, darão origem à População III, associada ao conjunto $\boldsymbol{\Omega}^{\text{III}}$.

A Figura 6.3 apresenta a comparação entre a curva de NMSE(n) obtida com o melhor modelo sob heurística MF, modelo $(2,89)$ – traço contínuo – e a curva $\text{NMSE}^{\text{III}}(n)$ resultante de $\text{JPS}\{\boldsymbol{\Omega}^{\text{III}}, Tspc6f\}$ – tracejado.

Observe que nesta etapa da heurística de definição de $\boldsymbol{\Omega}$, $\text{NMSE}_f^{\text{III}} = 0.966$, resultante de $\text{JPS}\{\boldsymbol{\Omega}^{\text{III}}, Tspc6f\}$. Observe também que, nesta etapa, a curva $\text{NMSE}^{\text{III}}(n)$ se afasta mais da curva de NMSE(n) do que a curva de $\text{NMSE}^{\text{II}}(n)$ se afastava de NMSE(n) na etapa anterior. À medida que as etapas da heurística de definição de $\boldsymbol{\Omega}$ se sucederem, este comportamento é não só esperado, como também desejado.

Na etapa recém descrita do algoritmo, a população evoluiu para um conjunto de modelos mais aptos do que aqueles que compunham a população anterior, através da exclusão de alguns modelos menos adequados aos intervalos estatisticamente diferenciados que compõem a série temporal. No entanto, é razoável investigar dentre os indivíduos inicialmente excluídos da População I, associada ao conjunto $\boldsymbol{\Omega}^{\text{I}}$, a existência de algum possível modelo a ser agregado à população atual de modelos (representada pelo conjunto

Ω^{III}), caso a inclusão deste indivíduo contribua para melhorar o desempenho com relação ao da predição $\mathrm{JPS}\{\Omega^{III}, Tspc6f\}$.

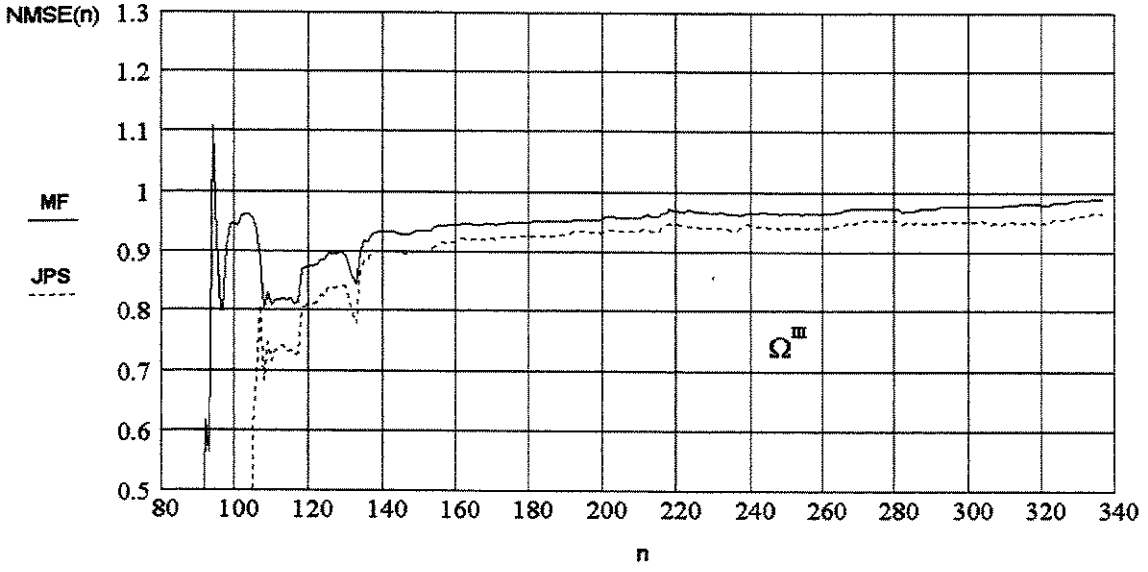

Figura 6.3: Série *Tspc6f* – Comparação entre a curva de NMSE(n) obtida com o modelo $(2,89)$ sob heurística MF – traço contínuo – e a curva $\mathrm{NMSE}^{III}(n)$ resultante de $\mathrm{JPS}\{\Omega^{III}, Tspc6f\}$ – tracejado.

Para que esta possibilidade seja considerada, um novo critério seletivo é adotado nesta etapa da heurística de definição de Ω. O critério objetiva definir possíveis adicionais indivíduos aptos a serem incluídos em Ω^{III}, para que se reduza o NMSE_f^{III} da predição $\mathrm{JPS}\{\Omega^{III}, Tspc6f\}$. Estes indivíduos (M, L, NMSE_f) adicionais – que formarão o conjunto $\Omega^{III}_{M^+}$ – serão buscados dentre os modelos (M, L) pertencentes a Ω^{I}, cujas dimensões M^+ sejam iguais ou superiores à maior dimensão M dos modelos (M, L) de Ω^{II} que foram selecionados pelo corte em Ω^{I}.

Dentro do universo de busca especificado para formar $\Omega^{III}_{M^+}$, serão selecionados para avaliação aqueles modelos (M^+, L) cuja predição de *Tspc6f* sob a heurística MF resultou

nos N_v menores valores de NMSE_f para uma dada dimensão M^+, onde N_v é um parâmetro da heurística de definição de $\boldsymbol{\Omega}$, assumindo, usualmente, os valores 1 a 4.

Para exemplificar, da Tabela 6.3[1] verifica-se que a População II, gerada pelo corte em $\boldsymbol{\Omega}^{\text{I}}$, compreende modelos (M,L) com $M=\{2,3,4\}$. Nesta situação, o conjunto das possíveis dimensões M^+ é definido por $M^+ \geq 5$, isto é, $M^+ = \{5,6,\cdots,12\}$. Portanto, para $N_v = 3$, o conjunto de modelos (M^+,L) é dado por $\boldsymbol{\Omega}^{\text{III}}_{M^+}$ = {(5,89), (5,96), (5,97), (6,95), (6,97), (6,99), (7,94), (7,95), (7,98), (8,93), (8,95), (8,100), (9,100) ,(9,99), (9,106), (10,95), (10,98), (10,93), (11,82), (11,89), (11,92), (12,22), (12,91), (12,9)}.

Seja (M_p^+,L_p) um particular modelo pertencente a $\boldsymbol{\Omega}^{\text{III}}_{M^+}$. O p-ésimo modelo $(M_p^+,L_p)\in\boldsymbol{\Omega}^{\text{III}}_{M^+}$ é adicionado ao conjunto $\boldsymbol{\Omega}^{\text{III}}$ e é realizada a p-ésima predição de *Tspc6f* através da heurística JPS, isto é, é efetuada a operação

$$\text{JPS}\{\boldsymbol{\Omega}^{\text{III}} + \{(M_p^+,L_p)\}, Tspc6f\},\ p=0,1,\cdots,\#\{\boldsymbol{\Omega}^{\text{III}}_{M^+}\}-1 \tag{6.61}$$

perfazendo um total de $\#\{\boldsymbol{\Omega}^{\text{III}}_{M^+}\}$ predições conforme (6.61).

A cada predição p de *Tspc6f* por JPS, em que o p-ésimo modelo $(M_p^+,L_p)\in\boldsymbol{\Omega}^{\text{III}}_{M^+}$ é incluído no conjunto $\boldsymbol{\Omega}^{\text{III}}$, é obtido um $\text{NMSE}_{\text{f}}^{p+}$ resultante da p-ésima predição, operação que será representada por

$$\text{NMSE}_{\text{f}}^{p+} \leftarrow \text{JPS}\{\boldsymbol{\Omega}^{\text{III}} + \{(M_p^+,L_p)\}, Tspc6f\} \tag{6.62}$$

Aqueles modelos (M_p^+,L_p) cuja inclusão individual em $\boldsymbol{\Omega}^{\text{III}}$ resulta um $\text{NMSE}_{\text{f}}^{p+} < \text{NMSE}_{\text{f}}^{\text{III}}$ na predição $\text{JPS}\{\boldsymbol{\Omega}^{\text{III}} + \{(M_p^+,L_p)\}, Tspc6f\}$ são identificados como modelos da classe $(M^+_{p\min},L_{p\min})$ – classe de modelos que contribui para melhorar o desempenho da predição $\text{JPS}\{\boldsymbol{\Omega}^{\text{III}}, Tspc6f\}$ mediante inclusão – e são selecionados para

[1] A íntegra da Tabela 6.3 encontra-se no Apêndice C.

serem posteriormente agregados à população atual de modelos (População III, representada pelo conjunto $\boldsymbol{\Omega}^{\mathrm{III}}$).

Então, é adicionado a $\boldsymbol{\Omega}^{\mathrm{III}}$ o conjunto formado por todos os modelos $\left(M^{+}_{p\min}, L_{p\min}\right)$ pertencentes a $\boldsymbol{\Omega}^{\mathrm{III}}_{\mathrm{M}^{+}}$, cuja inclusão individual em $\boldsymbol{\Omega}^{\mathrm{III}}$ resulta $\mathrm{NMSE}^{p+}_{\mathrm{f}} < \mathrm{NMSE}^{\mathrm{III}}_{\mathrm{f}}$, formando a nova População IV, representada pelo conjunto $\boldsymbol{\Omega}^{\mathrm{IV}}$, isto é

$$\boldsymbol{\Omega}^{\mathrm{IV}} = \boldsymbol{\Omega}^{\mathrm{III}} + \bigcup_{k=0}^{N^{+}_{\min}-1} \left\{\left(M^{+}_{p\min}, L_{p\min}\right)\right\}_{k} \tag{6.63}$$

onde $N^{+}_{\min}$ é o número de modelos $\left(M^{+}_{p\min}, L_{p\min}\right)$ pertencentes a $\boldsymbol{\Omega}^{\mathrm{III}}_{\mathrm{M}^{+}}$, cuja inclusão individual em $\boldsymbol{\Omega}^{\mathrm{III}}$ resulta $\mathrm{NMSE}^{p+}_{\mathrm{f}} < \mathrm{NMSE}^{\mathrm{III}}_{\mathrm{f}}$, sendo $\bigcup$ o operador que representa a união entre conjuntos.

A Tabela 6.6 apresenta, para o exemplo que está sendo acompanhado (predição da série *Tspc6f*), os resultados da operação descrita por (6.61), (6.62) e (6.63). Os modelos $(6,95)$, $(6,97)$, $(7,94)$ e $(7,98)$ salientados na Tabela 6.6 são aqueles pertencentes à classe $\left(M^{+}_{p\min}, L_{p\min}\right)$, e portanto selecionados para inclusão em $\boldsymbol{\Omega}^{\mathrm{III}}$ conforme (6.63). A Tabela 6.7 apresenta a nova População IV, representada pelo conjunto $\boldsymbol{\Omega}^{\mathrm{IV}}$, para o mesmo exemplo.

A série *Tspc6f* é, então, predita através da heurística JPS com base no recém obtido conjunto $\boldsymbol{\Omega}^{\mathrm{IV}}$ e utilizando pré-processamento $\nabla\mathbf{U}$ com $\xi = 15$, isto é, é executada a operação definida pelo operador $\mathrm{JPS}\left\{\boldsymbol{\Omega}^{\mathrm{IV}}, Tspc6f\right\}$. O $\mathrm{NMSE}_{\mathrm{f}}$ resultante de $\mathrm{JPS}\left\{\boldsymbol{\Omega}^{\mathrm{IV}}, Tspc6f\right\}$, a ser referido como $\mathrm{NMSE}^{\mathrm{IV}}_{\mathrm{f}}$, será utilizado na próxima etapa da heurística de definição de $\boldsymbol{\Omega}$.

A exemplo das etapas anteriores, a Figura 6.4 apresenta a comparação entre a curva de $\mathrm{NMSE}(n)$ obtida com o modelo $(2,89)$, melhor modelo sob heurística MF – traço contínuo – e a curva $\mathrm{NMSE}^{\mathrm{IV}}(n)$ resultante de $\mathrm{JPS}\left\{\boldsymbol{\Omega}^{\mathrm{IV}}, Tspc6f\right\}$ – tracejado.

População Envolvida	NMSE_f^{p+}
$\boldsymbol{\Omega}^{\text{III}}$ + {(5,89)}	0.967
$\boldsymbol{\Omega}^{\text{III}}$ + {(5,96)}	0.974
$\boldsymbol{\Omega}^{\text{III}}$ + {(5,97)}	0.98
$\boldsymbol{\Omega}^{\text{III}}$ + **{(6,95)}**	**0.958**
$\boldsymbol{\Omega}^{\text{III}}$ + **{(6, 97)}**	**0.954**
$\boldsymbol{\Omega}^{\text{III}}$ + {(6,99)}	0.969
$\boldsymbol{\Omega}^{\text{III}}$ + **{(7, 94)}**	**0.961**
$\boldsymbol{\Omega}^{\text{III}}$ + {(7, 95)}	0.969
$\boldsymbol{\Omega}^{\text{III}}$ + **{(7, 98)}**	**0.964**
$\boldsymbol{\Omega}^{\text{III}}$ + {(8,93)}	0.984
$\boldsymbol{\Omega}^{\text{III}}$ + {(8,95)}	0.982
$\boldsymbol{\Omega}^{\text{III}}$ + {(8,100)}	1.023
$\boldsymbol{\Omega}^{\text{III}}$ + {(9,100)}	0.998
$\boldsymbol{\Omega}^{\text{III}}$ + {(9, 99)}	1.008
$\boldsymbol{\Omega}^{\text{III}}$ + {(9,106)}	1.005
$\boldsymbol{\Omega}^{\text{III}}$ + {(10,95)}	0.996
$\boldsymbol{\Omega}^{\text{III}}$ + {(10,98)}	1.01
$\boldsymbol{\Omega}^{\text{III}}$ + {(10,93)}	1.025
$\boldsymbol{\Omega}^{\text{III}}$ + {(11,82)}	1.001
$\boldsymbol{\Omega}^{\text{III}}$ + {(11,89)}	1.026
$\boldsymbol{\Omega}^{\text{III}}$ + {(11,92)}	1.018
$\boldsymbol{\Omega}^{\text{III}}$ + {(12,22)}	1.423
$\boldsymbol{\Omega}^{\text{III}}$ + {(12,91)}	1.026
$\boldsymbol{\Omega}^{\text{III}}$ + {(12,92)}	1.015

Tabela 6.6: Resultados da predição $\text{JPS}\left\{\boldsymbol{\Omega}^{\text{III}} + \left\{\left(M_p^+, L_p\right)\right\}, Tspc6f\right\}$, $p = 0,1,\cdots,\#\left\{\boldsymbol{\Omega}_{M^+}^{\text{III}}\right\} - 1$. Os modelos salientados em negrito são aqueles cuja inclusão individual em $\boldsymbol{\Omega}^{\text{III}}$ resulta $\text{NMSE}_f^{p+} < \text{NMSE}_f^{\text{III}} = 0.966$.

Indivíduo (M, L, NMSE_f)		
Dimensão M dos Vetores	Número L de Vetores	NMSE_f (Heurística MF)
2	89	0.99
2	88	0.991
2	82	0.998
4	99	0.999
3	84	1
2	84	1
6	95	1.03
6	97	1.04
7	94	1.01
7	98	1.03

Tabela 6.7: População IV. Os indivíduos assinalados em tom cinza foram os pertencentes a $\boldsymbol{\Omega}^{\text{III}}_{M^*}$, selecionados para inclusão à População III, dando origem à População IV, associada ao conjunto $\boldsymbol{\Omega}^{\text{IV}}$.

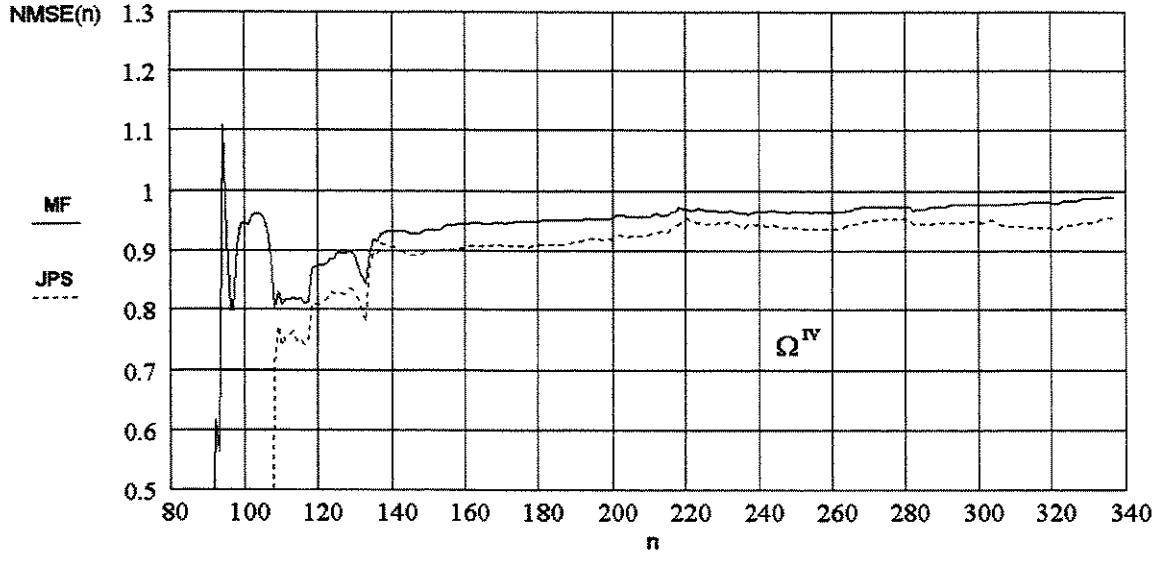

Figura 6.4: Série *Tspc6f* – Comparação entre a curva de NMSE(n) obtida com o modelo $(2,89)$ sob heurística MF – traço contínuo – e a curva $\text{NMSE}^{\text{IV}}(n)$ resultante de $\text{JPS}\{\boldsymbol{\Omega}^{\text{IV}}, Tspc6f\}$ – tracejado.

Observe que, nesta etapa da heurística de definição de $\mathbf{\Omega}$, $\text{NMSE}_{\text{f}}^{\text{IV}} = 0.96$, resultante de $\text{JPS}\{\mathbf{\Omega}^{\text{IV}}, Tspc6f\}$. Também a exemplo das etapas prévias, note que a curva $\text{NMSE}^{\text{IV}}(n)$ se afasta mais da curva de $\text{NMSE}(n)$ do que a curva de $\text{NMSE}^{\text{III}}(n)$ se afastava de $\text{NMSE}(n)$ na etapa anterior.

A predição $\text{JPS}\{\mathbf{\Omega}^{\text{IV}}, Tspc6f\}$ baseia-se no procedimento de acréscimo de diferentes modelos $\left(M^{+}_{p\min}, L_{p\min}\right)$ à $\mathbf{\Omega}^{\text{III}}$, para que o conjunto $\mathbf{\Omega}^{\text{IV}}$ resultante tenha melhores condições de representar as não-estacionariedades presentes ao longo da série. Este procedimento tem como fundamento a idéia de que modelos excluídos pelo corte em $\mathbf{\Omega}^{\text{I}}$ talvez contribuam para melhorar o desempenho de predição com relação ao desempenho da predição $\text{JPS}\{\mathbf{\Omega}^{\text{III}}, Tspc6f\}$. Assim, a nova população definida por $\mathbf{\Omega}^{\text{IV}}$ estabelece um novo cenário de relações de cooperação – ou de conflito – entre modelos (M, L) na tarefa de predição da série *Tspc6f* por JPS. No entanto, os modelos $\left(M^{+}_{p\min}, L_{p\min}\right)$ acrescentados à $\mathbf{\Omega}^{\text{III}}$ tiveram sua aptidão na predição $\text{JPS}\{\mathbf{\Omega}^{\text{III}}, Tspc6f\}$ testada apenas individualmente através de $\text{NMSE}_{\text{f}}^{p+}$. Portanto, no novo cenário de relações de cooperação entre modelos (M, L) não é sabido se, na atuação conjunta dos modelos $\mathbf{\Omega}^{\text{III}} + \bigcup_{k=0}^{N^{+}_{\min}-1} \left\{\left(M^{+}_{p\min}, L_{p\min}\right)\right\}_k$, implícita em $\text{JPS}\{\mathbf{\Omega}^{\text{IV}}, Tspc6f\}$, possam existir modelos cuja eliminação de $\mathbf{\Omega}^{\text{IV}}$ resulte em redução de $\text{NMSE}_{\text{f}}^{\text{IV}}$.

Desta maneira, nos vemos mais uma vez na situação de avaliar o novo cenário de relações sob o ponto de vista de que é possível que alguns dos modelos de $\mathbf{\Omega}^{\text{IV}}$ não sejam adequados aos intervalos estatisticamente diferenciados que compõem a série temporal e possam estar prejudicando o desempenho da predição $\text{JPS}\{\mathbf{\Omega}^{\text{IV}}, Tspc6f\}$. Assim, critério idêntico ao adotado na etapa de obtenção de $\mathbf{\Omega}^{\text{III}}$, descrito pelas Equações (6.57) a (6.60), é agora adotado com relação à $\mathbf{\Omega}^{\text{IV}}$, isto é

$$\mathrm{JPS}\left\{\boldsymbol{\Omega}^{\mathrm{IV}} - \left\{\left(M_p, L_p\right)\right\}, Tspc6f\right\},\ p = 0,1,\cdots,\#\left\{\boldsymbol{\Omega}^{\mathrm{IV}}\right\}-1 \tag{6.64}$$

$$\mathrm{NMSE}_{\mathrm{f}}^{p} \leftarrow \mathrm{JPS}\left\{\boldsymbol{\Omega}^{\mathrm{IV}} - \left\{\left(M_p, L_p\right)\right\}, Tspc6f\right\} \tag{6.65}$$

$$\mathrm{NMSE}_{\mathrm{f}}^{\mathrm{MIN}} = \arg\min_{p}\left\{\mathrm{NMSE}_{\mathrm{f}}^{p}\right\},\ p = 0,1,\cdots,\#\left\{\boldsymbol{\Omega}^{\mathrm{IV}}\right\}-1 \tag{6.66}$$

$$\mathrm{NMSE}_{\mathrm{f}}^{\mathrm{MIN}} < \mathrm{NMSE}_{\mathrm{d}} \Rightarrow \boldsymbol{\Omega}^{\mathrm{IV}} = \boldsymbol{\Omega}^{\mathrm{IV}} - \left\{\left(M_{p\min}, L_{p\min}\right)\right\} \tag{6.67}$$

A recursão definida por (6.64) a (6.67) é repetida até que a exclusão definitiva de algum particular $\left(M_p, L_p\right)$ não resulte em $\mathrm{NMSE}_{\mathrm{f}}^{\mathrm{MIN}} < \mathrm{NMSE}_{\mathrm{d}}$, sendo $\mathrm{NMSE}_{\mathrm{d}}$ o valor de $\mathrm{NMSE}_{\mathrm{f}}^{\mathrm{MIN}}$ resultante da recursão anterior.

Da mesma forma que na etapa de obtenção de $\boldsymbol{\Omega}^{\mathrm{III}}$, caso a condição $\mathrm{NMSE}_{\mathrm{f}}^{\mathrm{MIN}} < \mathrm{NMSE}_{\mathrm{d}}$ não seja satisfeita em conseqüência da exclusão de qualquer um dos modelos pertencentes à $\boldsymbol{\Omega}^{\mathrm{IV}}$, conclui-se que nenhum destes modelos esteja prejudicando o desempenho da predição $\mathrm{JPS}\left\{\boldsymbol{\Omega}^{\mathrm{IV}}, Tspc6f\right\}$ e, portanto nenhum deles será excluído da População IV.

A Tabela 6.8 apresenta os resultados, para a predição da série *Tspc6f*, do processo de eliminação dos indivíduos menos aptos na tarefa da predição $\mathrm{JPS}\left\{\boldsymbol{\Omega}^{\mathrm{IV}}, Tspc6f\right\}$. As colunas 1 a 3 apresentam três recursões desta etapa do algoritmo. Os indivíduos assinalados nas duas primeiras colunas são aqueles efetivamente eliminados da População IV a cada conjunto de predições $\mathrm{JPS}\left\{\boldsymbol{\Omega}^{\mathrm{IV}} - \left\{\left(M_p, L_p\right)\right\}, Tspc6f\right\},\ p = 0,1,\cdots,\#\left\{\boldsymbol{\Omega}^{\mathrm{IV}}\right\}-1$, respectivo a cada coluna. Como pode ser observado na tabela, a eliminação de cada um destes indivíduos leva progressivamente à diminuição do $\mathrm{NMSE}_{\mathrm{f}}^{\mathrm{MIN}}$, com relação ao $\mathrm{NMSE}_{\mathrm{f}}^{\mathrm{IV}}$.

Coluna 1			Coluna 2			Coluna 3		
População Base: População IV			População Base: População IV exceto modelo (7, 98)			População Base: População IV exceto modelos (7,98) e (7,94).		
Indivíduo Excluído		$NMSE_f^p$	Indivíduo Excluído		$NMSE_f^p$	Indivíduo Excluído		$NMSE_f^p$
2	89	0.9649	2	89	0.9652	2	89	0.9591
2	88	0.9652	2	88	0.9653	2	88	0.9514
2	82	0.9657	2	82	0.9628	2	82	0.9505
2	84	0.9582	2	84	0.9595	2	84	0.9503
3	84	0.9657	3	84	0.9675	3	84	0.9563
4	99	0.9703	4	99	0.9696	4	99	0.9595
6	95	0.9586	6	95	0.9601	6	95	0.9535
6	97	0.9638	6	97	0.9664	6	97	0.9576
7	94	0.96	7	94	0.9477			
7	98	0.9582						

Tabela 6.8: Eliminação dos indivíduos menos aptos no processo de predição $\mathrm{JPS}\{\boldsymbol{\Omega}^{\mathrm{IV}}, Tspc6f\}$. Observe que os elementos destacados nas colunas 1 a 3 são aqueles cuja exclusão implicará no menor $NMSE_f^p$ a cada conjunto de predições $\mathrm{JPS}\{\boldsymbol{\Omega}^{\mathrm{IV}} - \{(M_p, L_p)\}, Tspc6f\}$, $p = 0,1,\cdots,\#\{\boldsymbol{\Omega}^{\mathrm{IV}}\}-1$, respectivo a cada coluna.

Observe na coluna 1 da Tabela 6.8, que o elemento $(7,98)$ é definitivamente excluído de $\boldsymbol{\Omega}^{\mathrm{IV}}$ porque $\left(\mathrm{NMSE}_f^{\mathrm{MIN}} = 0.9582\right) < \left(\mathrm{NMSE}_f^{\mathrm{IV}} = 0.96\right)$. Observe também que o modelo $(2,84)$, assinalado na coluna 4 da tabela, cuja exclusão resulta em $\mathrm{NMSE}_f^{\mathrm{MIN}} = 0.9503$, não será eliminado da população, pois, apesar da sua exclusão do grupo ser a exclusão que conduz ao menor $\mathrm{NMSE}_f^{\mathrm{MIN}}$, o valor atingido não é inferior ao atingido pela recursão anterior (coluna 2, modelo $(7,94)$, $\mathrm{NMSE}_f^{\mathrm{MIN}} = 0.9477$), evento que indica o final das recursões.

A Tabela 6.9 apresenta a População IV após a totalidade do procedimento especificado na Tabela 6.8. Na Tabela 6.9 estão assinalados os elementos selecionados

para exclusão definitiva. A exclusão dos referidos elementos na População IV define a População V, associada ao conjunto $\boldsymbol{\Omega}^{V}$.

Indivíduo (M, L, NMSE_f)		
Dimensão M dos Vetores	Número L de Vetores	NMSE_f (Heurística MF)
2	89	0.99
2	88	0.991
2	82	0.998
2	84	1
3	84	1
4	99	0.999
6	95	1.03
6	97	1.04
7	98	1.03
7	94	1.01

Tabela 6.9: População IV, na qual os indivíduos assinalados em cinza, ao serem definitivamente excluídos, darão origem à População V, associada ao conjunto $\boldsymbol{\Omega}^{V}$.

A série *Tspc6f* é, então, predita através da heurística JPS com base no recém obtido conjunto $\boldsymbol{\Omega}^{V}$ e utilizando pré-processamento $\nabla\mathbf{U}$ com $\xi = 15$, isto é, é executada a operação definida pelo operador $\text{JPS}\{\boldsymbol{\Omega}^{V}, Tspc6f\}$. O NMSE_f resultante de $\text{JPS}\{\boldsymbol{\Omega}^{V}, Tspc6f\}$ será referido como NMSE_f^{V}.

A Figura 6.5 apresenta a comparação entre a curva de NMSE(n) obtida com o melhor modelo sob heurística MF, modelo $(2,89)$ – traço contínuo – e a curva $\text{NMSE}^{V}(n)$ resultante de $\text{JPS}\{\boldsymbol{\Omega}^{V}, Tspc6f\}$ – tracejado. Note que $\text{NMSE}_f^{V} = 0.948$, resultante de $\text{JPS}\{\boldsymbol{\Omega}^{V}, Tspc6f\}$. A etapa de obtenção de $\boldsymbol{\Omega}^{V}$ é a última etapa da heurística de definição de $\boldsymbol{\Omega}$. Portanto $\boldsymbol{\Omega} = \boldsymbol{\Omega}^{V}$ e $\text{NMSE}^{\Omega}(n) = \text{NMSE}^{V}(n)$.

Observe que a curva $\mathrm{NMSE}^{\Omega}(n)$ resultante de $\mathrm{JPS}\{\boldsymbol{\Omega}, Tspc6f\}$ representa uma considerável redução de erro de predição com relação à curva de $\mathrm{NMSE}(n)$ relativa à predição de *Tspc6f* pela heurística MF utilizando o modelo $(2,89)$ – melhor modelo sob esta heurística.

Tendo sido definido $\boldsymbol{\Omega}$, isto é, tendo sido definidos os indivíduos selecionados dentre a população global considerados os mais aptos para a tarefa de predição de uma série temporal $U(n)$ com N_t amostras, um procedimento adicional pode ser efetuado, objetivando diminuir ainda mais $\mathrm{NMSE}_f^{\Omega} = \mathrm{NMSE}^{\Omega}(N_t - 1)$. O procedimento consiste em executar a predição $\mathrm{JPS}\{\boldsymbol{\Omega}, U\}$ com pequenos ajustes no Fator de Variância ξ visando reduzir NMSE_f^{Ω}. No entanto, cabe ressaltar que este procedimento nem sempre implica em redução de NMSE_f^{Ω}.

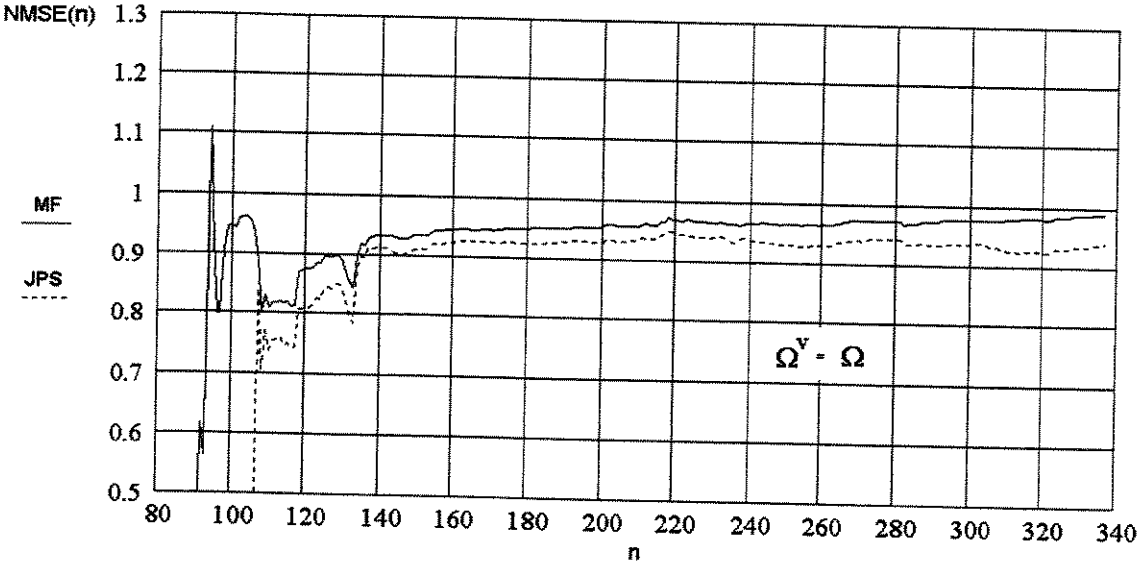

Figura 6.5: Série *Tspc6f* – Comparação entre a curva de $\mathrm{NMSE}(n)$ obtida com o modelo $(2,89)$ sob heurística MF – traço contínuo – e a curva $\mathrm{NMSE}^{\Omega}(n)$ resultante de $\mathrm{JPS}\{\boldsymbol{\Omega}, Tspc6f\}$ – tracejado.

A Tabela 6.10 apresenta os valores de $NMSE_f$ obtidos após a aplicação de cada uma das etapas da heurística de definição de Ω. Observe que, mesmo para a série *Tspc6f*, que é uma das séries temporais de maior dificuldade de predição dentre as que testamos, a adoção da heurística de definição de Ω representa um ganho significativo no desempenho da predição por DSE. O menor $NMSE_f$ obtido pela heurística de predição MF foi 0.99 enquanto que, após a definição de Ω, utilizando a heurística de predição JPS, foi possível obter uma redução no valor do $NMSE_f$ para 0.948.

População Envolvida na Predição	Heurística de Predição Adotada	$NMSE_f$
Ω^{I}	Heurística MF – melhor modelo *	0.99
Ω^{II}	$JPS\{\Omega^{II}, Tspc6f\}$	0.999
Ω^{III}	$JPS\{\Omega^{III}, Tspc6f\}$	0.966
Ω^{IV}	$JPS\{\Omega^{IV}, Tspc6f\}$	0.96
$\Omega^{V} = \Omega$	$JPS\{\Omega^{V}, Tspc6f\}$	0.948
* modelo (2,89) = melhor modelo sob a heurística de modelo de predição fixo.		

Tabela 6.10: Valores de $NMSE_f$ obtidos após a aplicação de cada uma das etapas da heurística de definição de Ω.

Heurística de Definição de $\boldsymbol{\Omega}$			
Etapa	Descrição	Conjunto Associado à População Inicial	Conjunto Associado à População Resultante
1	Eliminação dos indivíduos da População I para os quais $\mathrm{NMSE}_{\mathrm{f}}(M,L) > \mathrm{NMSE}_{\mathrm{c}}$.	$\boldsymbol{\Omega}^{\mathrm{I}}$	$\boldsymbol{\Omega}^{\mathrm{II}}$
2	Eliminação dos indivíduos menos aptos no processo de predição $\mathrm{JPS}\{\boldsymbol{\Omega}^{\mathrm{II}}, Tspc6f\}$.	$\boldsymbol{\Omega}^{\mathrm{II}}$	$\boldsymbol{\Omega}^{\mathrm{III}}$
3	Inclusão em $\boldsymbol{\Omega}^{\mathrm{III}}$ do conjunto formado por todos os modelos $(M^{+}_{p\min}, L_{p\min})$ pertencentes a $\boldsymbol{\Omega}^{\mathrm{III}}_{\mathrm{M}^{+}}$, cuja inclusão individual em $\boldsymbol{\Omega}^{\mathrm{III}}$ resulta $\mathrm{NMSE}_{\mathrm{f}}^{p+} < \mathrm{NMSE}_{\mathrm{f}}^{\mathrm{III}}$.	$\boldsymbol{\Omega}^{\mathrm{III}}$	$\boldsymbol{\Omega}^{\mathrm{IV}}$
4	Eliminação dos indivíduos menos aptos no processo de predição $\mathrm{JPS}\{\boldsymbol{\Omega}^{\mathrm{IV}}, Tspc6f\}$.	$\boldsymbol{\Omega}^{\mathrm{IV}}$	$\boldsymbol{\Omega}^{\mathrm{V}} = \boldsymbol{\Omega}$

Tabela 6.11: Descrição das etapas da heurística de definição de $\boldsymbol{\Omega}$ e respectivas populações envolvidas.

População I			População I			População II			População III			População IV			População V		
Todos os possíveis indivíduos – heurística MF.			População em ordem ascendente de $NMSE_f$ – heurística MF.			População I com a seleção dos indivíduos que serão eliminados, para os quais $NMSE_f(M,L) > NMSE_c$.			População II sem os indivíduos menos aptos no processo de predição $JPS\{\Omega^{II}, Tspc6f\}$.			População III com a inclusão do conjunto formado por todos os modelos pertencentes a $\Omega^{III}_{M^+}$, cuja inclusão individual resulta $NMSE_f^{p+} < NMSE_f^{III}$.			População IV sem os indivíduos menos aptos no processo de predição $JPS\{\Omega^{IV}, Tspc6f\}$.		
2	20	1.15	2	89	0.99	2	89	0.99	2	89	0.99	2	89	0.99	2	89	0.99
2	21	1.08	2	88	0.991	2	88	0.991	2	88	0.991	2	88	0.991	2	88	0.991
2	22	1.1	4	104	0.997	4	104	0.997	---	---	---	---	---	---	---	---	---
2	23	1.1	2	82	0.998	2	82	0.998	2	82	0.998	2	82	0.998	2	82	0.998
2	24	1.09	4	99	0.999	4	99	0.999	4	99	0.999	4	99	0.999	4	99	0.999
2	25	1.08	3	84	1	3	84	1	3	84	1	3	84	1	3	84	1
2	26	1.04	2	90	1	2	90	1	---	---	---	---	---	---	---	---	---
2	27	1.05	2	84	1	2	84	1	2	84	1	2	84	1	2	84	1
2	28	1.05	2	94	1	2	94	1	---	---	---	---	---	---	---	---	---
2	29	1.07	3	97	1.01	3	97	1.01				6	95	1.03	6	95	1.03
2	30	1.03	7	94	1.01	7	94	1.01				6	97	1.04	6	97	1.04
2	31	1.05	2	105	1.01	2	105	1.01				7	94	1.01	---	---	---
2	32	1.08	2	100	1.01	2	100	1.01				7	98	1.03	---	---	---
2	33	1.08	2	104	1.01	2	104	1.01									
2	34	1.05	3	101	1.01	3	101	1.01									
2	35	1.06	3	91	1.01	3	91	1.01									
2	36	1.07	3	88	1.01	3	88	1.01									
2	37	1.07	3	93	1.01	3	93	1.01									
2	38	1.08	3	94	1.01	3	94	1.01									
2	39	1.08	2	71	1.01	2	71	1.01									
2	40	1.09	3	99	1.01	3	99	1.01									
2	41	1.1	5	89	1.01	5	89	1.01									
.....			.....			.....											
12	143	1.18	12	23	2.14	12	23	2.14									
12	144	1.21	10	22	2.24	10	22	2.24									
12	145	1.18	9	20	2.29	9	20	2.29									
12	146	1.2	11	22	2.57	11	22	2.57									
12	147	1.14	10	20	2.6	10	20	2.6									
12	148	1.19	11	21	2.73	11	21	2.73									
12	149	1.17	12	20	3.65	12	20	3.65									

Tabela 6.12: Populações de modelos envolvidas na heurística de definição de Ω.

6.3 Resultados Experimentais

Nesta seção são apresentados os resultados experimentais da predição – por DSE e através da heurística JPS – das séries temporais *Elet6f, Petr4f, Cmig4f, Elet3f* e *Bellco*, descritas no Capítulo 2 . A heurística de definição de $\mathbf{\Omega}$, descrita na Seção 6.2, é aplicada para a predição da série U através de JPS$\{\mathbf{\Omega}, U\}$, onde o argumento U refere-se a uma das séries em questão. Os resultados são comparados àqueles obtidos para a predição da série U através da heurística MF com o melhor modelo (M, L), isto é, o modelo que resulta no menor NMSE_f sob esta heurística.

Em todos os casos é utilizado o pré-processamento $\nabla\mathbf{U}$. Melhores resultados foram obtidos com este pré-processamento se comparados com aqueles obtidos sem pré-processamento ou com pré-processamento $\nabla^2\mathbf{U}$. Na predição da série *Bellco* é utilizado $\xi = 10$ e nas demais, $\xi = 15$. É importante salientar que, na predição por heurística MF, o fator ξ é variado adaptativamente pelo algoritmo LMS, conforme Capítulo 5, Seção 5.2.1, enquanto que, na heurística JPS, ξ é mantido constante na predição efetuada por todos os modelos do conjunto $\mathbf{\Omega}$. Sob ajuste do LMS, ξ é dado pela Equação (5.46) com $\alpha = 0.1$. O limiar de ativação do ajuste de ξ é $\text{NMSE}\max = 1.0$ e o limiar de convergência é $\text{NMSE}\min = 0.9$, conforme descrito na Seção 5.2.1. O passo de adaptação adotado na Equação (5.46) é $\eta = 0.01$.

As Figuras 6.6 a 6.10 apresentam a comparação entre a curva de $\text{NMSE}(n)$ obtida da predição de U sob a heurística MF com melhor modelo (M, L) e a curva $\text{NMSE}^{\mathbf{\Omega}}(n)$ resultante de JPS$\{\mathbf{\Omega}, U\}$. As Tabelas 6.13 a 6.17 apresentam os respectivos resultados de predição.

A Figura 6.11 ilustra a capacidade de generalização resultante da heurística de definição de $\mathbf{\Omega}$ quando aplicada na predição JPS$\{\mathbf{\Omega}, U\}$. Como pode ser observado na Seção 2.5, a série *Bellco* é formada por 1000 amostras. No entanto, a heurística de determinação de $\mathbf{\Omega}$ foi executada somente com base nas primeiras 560 amostras da série

Bellco, resultando na predição JPS$\{\boldsymbol{\Omega}, Bellco^*\}$ mostrada na Figura 6.10. Para testar a capacidade de generalização do conjunto $\boldsymbol{\Omega}$ obtido com base nas primeiras 560 amostras, foi realizada a predição JPS$\{\boldsymbol{\Omega}, Bellco\}$ ao longo das $N_t = 1000$ amostras, mostrada na Figura 6.11. Observe que as curvas $\text{NMSE}^{\boldsymbol{\Omega}}(n)$ não são exatamente iguais para $n < 560$, nas Figuras 6.10 e 6.11. Isto ocorre porque a série Bellco apresenta uma faixa de variação de valores maior para $n \geq 560$, o que altera os fatores que normalizam a série para o intervalo $[-1,1]$. Mesmo com os fatores de normalização alterados, as curvas $\text{NMSE}^{\boldsymbol{\Omega}}(n)$ nas Figuras 6.10 e 6.11 não só são bastante semelhantes para $n < 560$ como, imediatamente após a amostra 620, $\text{NMSE}^{\boldsymbol{\Omega}}(n)$ apresenta redução de seu valor instantâneo, na Figura 6.11.

A heurística de definição de $\boldsymbol{\Omega}$ procura adaptar a arquitetura da rede RBF ao processo estocástico $\mathbf{U}$ associado à série U, possivelmente não-estacionário, ao longo da predição JPS$\{\boldsymbol{\Omega}, U\}$. O termo "estacionário" deve ser entendido aqui sob o ponto de vista do mapeamento $\Im : \mathbf{U} \in \Re^M \rightarrow y \in \Re$ gerado pela rede RBF. O mapeamento $\Im$ é não-linear, logo não depende exclusivamente da correlação de $\mathbf{U}$ – uma grandeza estatística de segunda ordem. Por considerar grandezas estatísticas de ordem superior, $\Im$ é potencialmente capaz de antecipar o processo estocástico $\mathbf{U}$ associado à U mesmo quando $\mathbf{U}$ não é WSS (*Wide Sense Stationary*) [57], isto é, mesmo quando sua função de correlação varia ao longo de JPS$\{\boldsymbol{\Omega}, U\}$.

Assim, somente a predição sob heurística MF, com arquitetura fixa, já é potencialmente capaz de antecipar a série U quando $\mathbf{U}$ não é WSS. Mas quando a não-estacionariedade de $\mathbf{U}$ é tal que mesmo o mapeamento $\Im$ não é capaz de antecipar eficazmente a série U, a predição JPS$\{\boldsymbol{\Omega}, U\}$ pode resultar em aumento de desempenho como conseqüência da adaptação da arquitetura da rede RBF ao longo da predição de U.

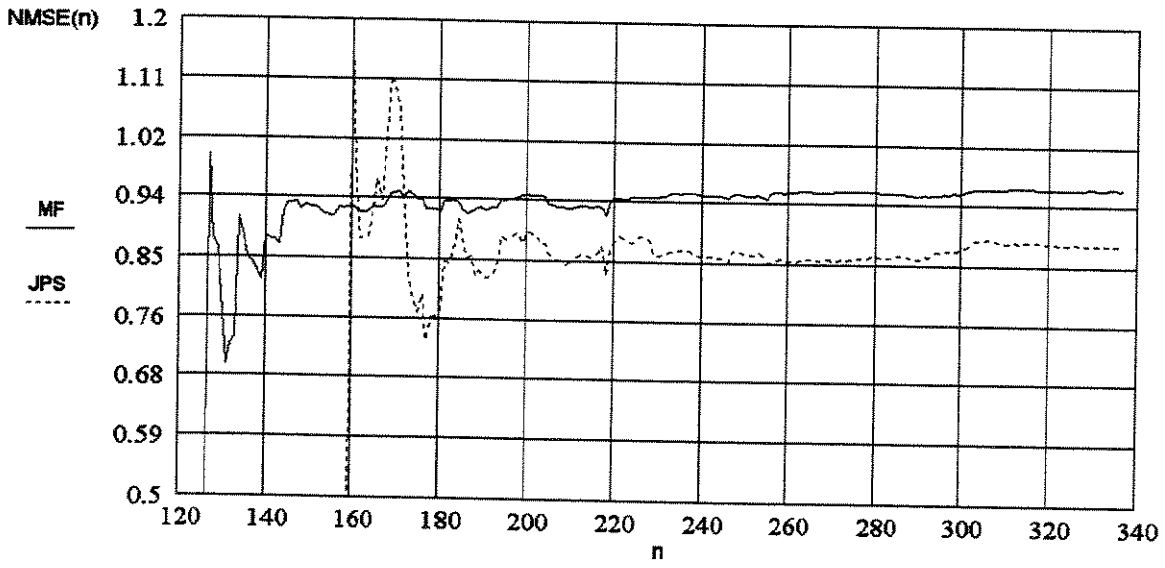

Figura 6.6: Série *Elet6f* – Comparação entre a curva de NMSE(n) obtida sob a heurística MF – traço contínuo – e a curva $\text{NMSE}^{\Omega}(n)$ resultante de JPS$\{\boldsymbol{\Omega}, Elet6f\}$ – tracejado.

Heurística de Predição	Modelos Envolvidos	NMSE_f
MF	$(4,123)$	0.963
JPS$\{\boldsymbol{\Omega}, Elet6f\}$	$\Omega = \{(3,136),(4,123),(5,141),(5,126),(5,107),$ $(5,108),(6,141),(6,96),(7,87),(7,97),(12,146)\}$	0.877

Tabela 6.13: Resultados de predição para a série *Elet6f*.

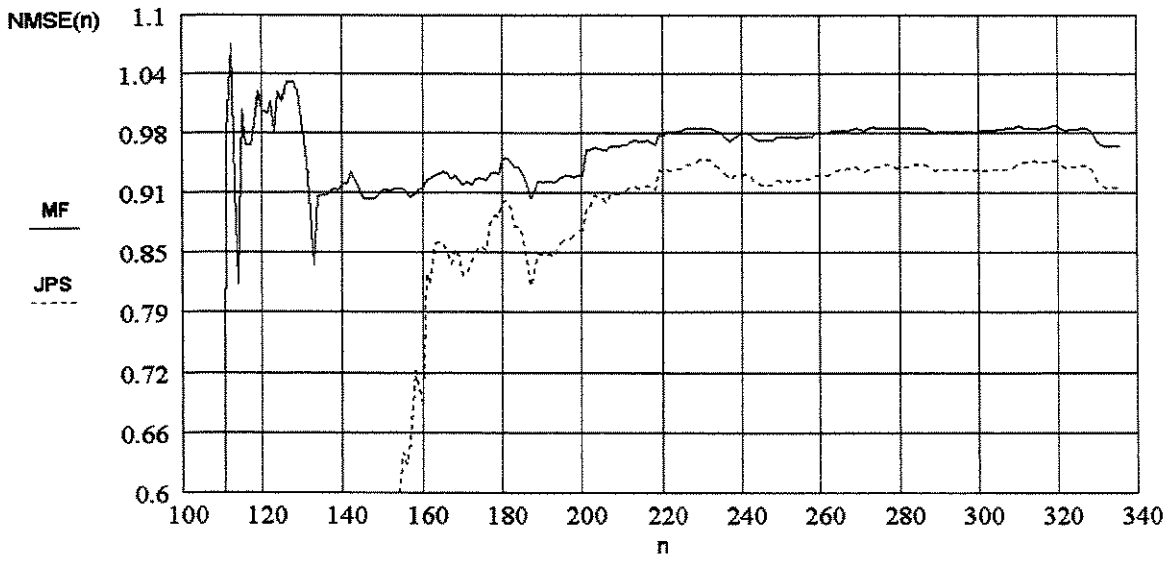

Figura 6.7: Série *Petr4f* – Comparação entre a curva de NMSE(n) obtida sob a heurística MF – traço contínuo – e a curva $\text{NMSE}^{\Omega}(n)$ resultante de JPS$\{\Omega, Petr4f\}$ – tracejado.

Heurística de Predição	Modelos Envolvidos	NMSE_f
MF	(3,108)	0.961
JPS$\{\Omega, Petr4f\}$	$\Omega = \{(2,117),(2,124),(2,141),(2,121),(2,113),$ $(2,71),(2,120),(2,116),(2,125),(2,126),(2,131),$ $(2,78),(2,95),(3,130),(3,107),(3,104),(3,119),$ $(3,148),(3,95),(3,142),(3,97),(3,102),(3,124),$ $(3,121),(3,116),(3,112),(4,118),(4,87),(4,111),$ $(4,107),(4,117),(4,109),(4,113),(4,122)\}$	0.919

Tabela 6.14: Resultados de predição para a série *Petr4f*.

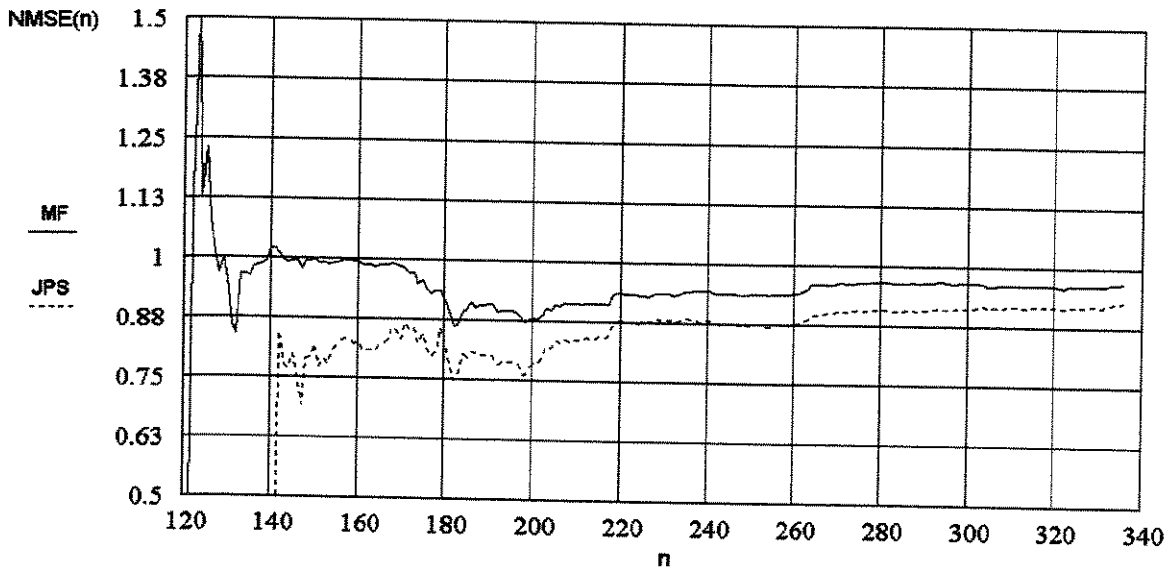

Figura 6.8: Série *Cmig4f* – Comparação entre a curva de NMSE(n) obtida sob a heurística MF – traço contínuo – e a curva $\text{NMSE}^{\Omega}(n)$ resultante de JPS$\{\boldsymbol{\Omega}, Cmig4f\}$ – tracejado.

Heurística de Predição	Modelos Envolvidos	NMSE_f
MF	$(5,117)$	0.967
JPS$\{\boldsymbol{\Omega}, Cmig4f\}$	$\Omega = \{(5,117),(5,95),(5,135)\}$	0.924

Tabela 6.15: Resultados de predição para a série *Cmig4f*.

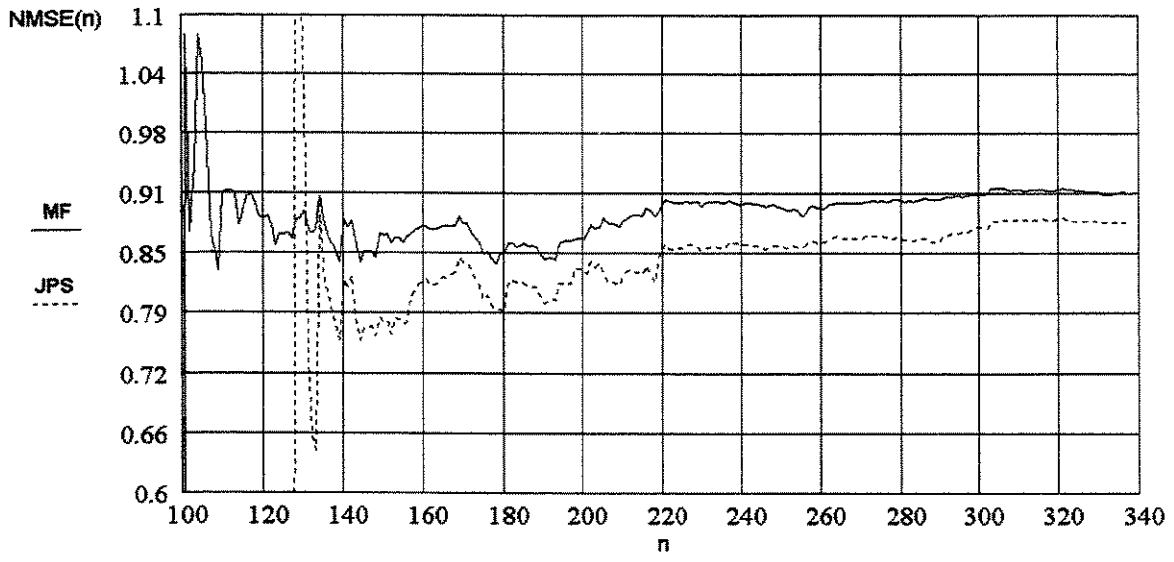

Figura 6.9: Série *Elet3f* – Comparação entre a curva de NMSE(n) obtida sob a heurística MF – traço contínuo – e a curva $\text{NMSE}^{\Omega}(n)$ resultante de JPS$\{\Omega, Elet3f\}$ – tracejado.

Heurística de Predição	Modelos Envolvidos	NMSE_f
MF	$(5,96)$	0.915
JPS$\{\Omega, Elet3f\}$	$\Omega = \{(4,88),(5,96),(5,104),$ $(6,114),(6,101),(11,110),(12,106)\}$	0.873

Tabela 6.16: Resultados de predição para a série *Elet3f*.

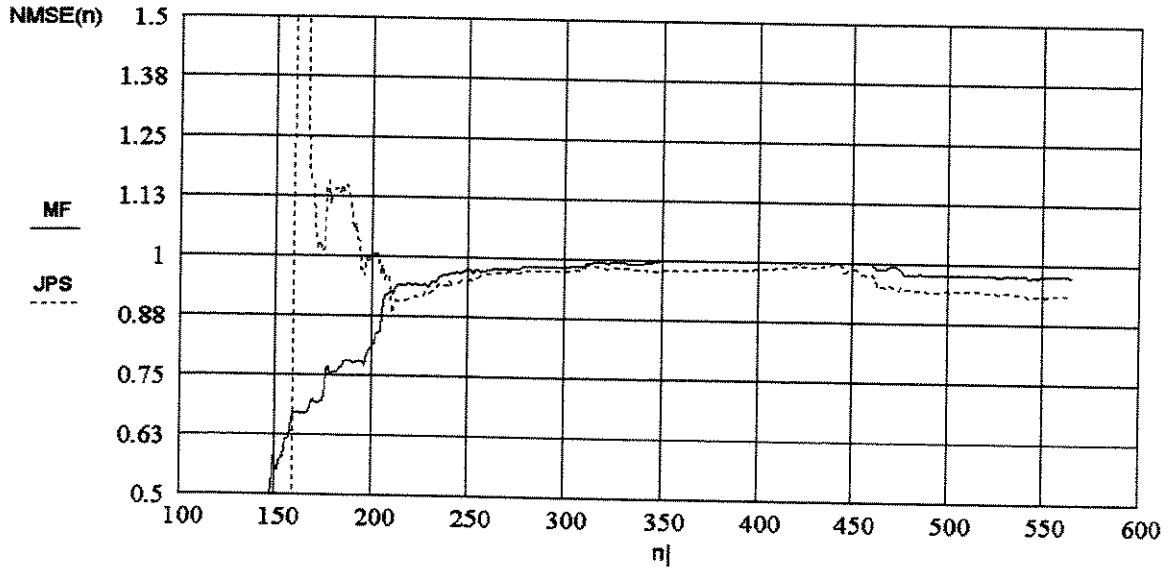

Figura 6.10: Série *Bellco* – Comparação entre a curva de NMSE(n) obtida sob a heurística MF – traço contínuo – e a curva $\text{NMSE}^{\Omega}(n)$ resultante de $\text{JPS}\{\boldsymbol{\Omega}, Bellco^{*}\}$ – tracejado. $Bellco^{*}$: somente as primeiras 560 amostras da série *Bellco*.

Heurística de Predição	Modelos Envolvidos	NMSE_f
MF	$(4,124)$	0.975
$\text{JPS}\{\boldsymbol{\Omega}, Bellco^{*}\}$	$\Omega = \{(2,135),(2,124),(2,100),(2,46),$ $(2,109),(2,44),(2,84),(4,100),(4,80),$ $(4,102),(5,123),(5,125),(6,86),(7,116),$ $(8,149),(8,148),(9,121),(9,106)\}$	0.929

Tabela 6.17: Resultados de predição para as primeiras 560 amostras da série *Bellco*.

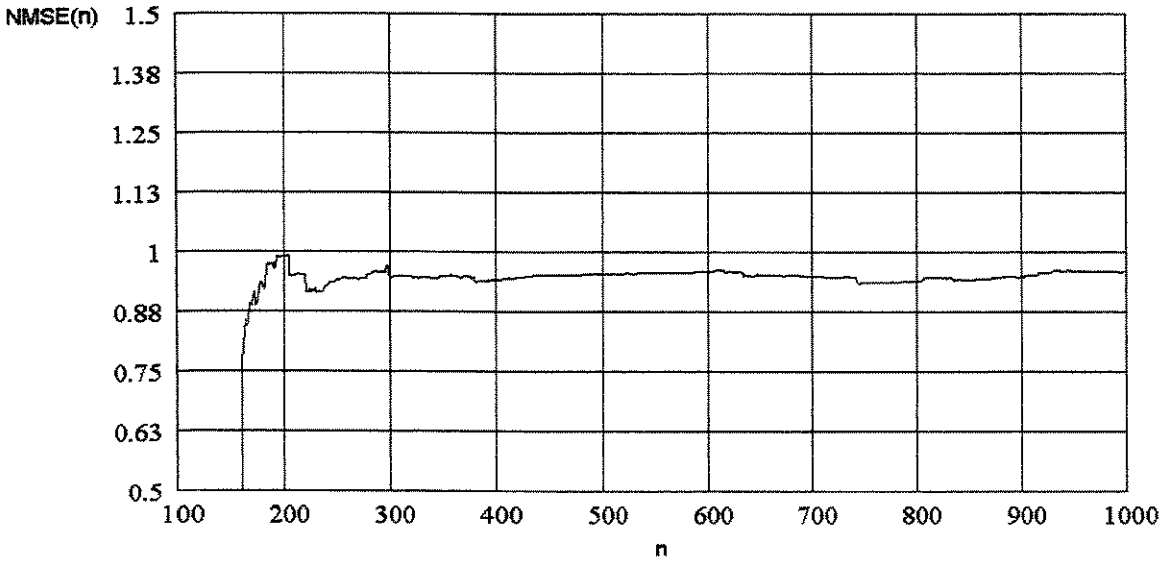

Figura 6.11: Série *Bellco* – $\mathrm{NMSE}^{\Omega}(n)$ resultante de $\mathrm{JPS}\{\boldsymbol{\Omega}, Bellco\}$ – Teste da capacidade de generalização da heurística de definição de $\boldsymbol{\Omega}$.

Capítulo 7

Conclusão

Esta tese apresentou uma heurística de predição não-linear de séries temporais através da Decomposição em Sub-Espaços – ou Decomposição em Componentes Principais – do processo estocástico vetorial $\mathbf{U}$ associado aos vetores de estado da série. A não-linearidade da heurística por Decomposição em Sub-Espaços (DSE) resulta do mapeamento não-linear $\Im : \Re^M \rightarrow \Re$ gerado por uma rede neural RBF, a qual, mediante treino, antecipa temporalmente o processo $\mathbf{U}$ com uma ordem de predição M.

Na heurística DSE, os vetores centros das funções de base radial da rede RBF são determinados a partir dos auto-valores e auto-vetores da matriz de covariância $\mathbf{C}$ do processo vetorial $\mathbf{U}$ que representa o desenrolar da série temporal S. Os vetores M-dimensionais usados para construir a matriz $\mathbf{C}$ resultam do deslizamento de uma janela $p(n)$ de M amostras sobre S. Sob tal situação, a matriz $\mathbf{C}$ pode ser interpretada como uma matriz que expressa correlação temporal entre regiões de M amostras consecutivas em S. À medida que $p(n)$ é deslizada sobre S, as matrizes de covariância formadas contêm muito da informação do desenrolar temporal de janelas anteriores, porque, a cada nova janela, apenas um elemento da série é eliminado e um é agregado.

A atribuição dos vetores de estado determinados através da heurística DSE aos centros das funções de base radial possibilita que mais informação sobre correlação temporal da série S seja passada à rede RBF do que os demais métodos de atribuição de centros até então conhecidos – a técnica APC e a técnica *k-means*, anteriormente referidas nesta tese. Na técnica APC os vetores de estado são definidos como os vetores mais recentes de $\mathbf{U}$ em $p(n)$, sem que seja aplicada qualquer transformação ao conjunto $\mathbf{U}$ de vetores no sentido de explorar as características estocásticas de $\mathbf{U}$. Na técnica *k-means* os

vetores escolhidos para vetores de estado são os vetores mais freqüentes de $\mathbf{U}$ em $p(n)$, não significando, no entanto, que sejam os mais correlacionados. Na técnica DSE apresentada nesta tese, o critério de definição dos vetores de estado a serem atribuídos aos centros das funções de base radial permite que sejam escolhidos os vetores mais correlacionados de $\mathbf{U}$ em $p(n)$.

Para o caso do processo $\mathbf{U}$ ser não-estacionário sob o ponto de vista do mapeamento $\Im$, o que ocorre freqüentemente nos processos que regem séries temporais de interesse científico, é apresentada uma extensão do método – a heurística Janela de Predição Seletiva (JPS) – inspirada em Algoritmos Genéticos, na qual a arquitetura da rede RBF é alterada ao longo do processo de predição de uma mesma série S, de modo a minimizar o efeito da não-estacionariedade de $\mathbf{U}$.

A heurística JPS não se trata de um algoritmo genético. No entanto, deixa-se aqui uma sugestão para futuros trabalhos em que o problema seja tratado sob uma abordagem clássica de algoritmos genéticos. Buscando definir o conjunto $\mathbf{\Omega}$ de soluções mais adequadas à tarefa de predição de uma determinada série temporal, uma idéia seria considerar uma população inicial de indivíduos, em que cada indivíduo seria formado pelo conjunto de algumas possíveis soluções – i.é, modelos (M, L, NMSE_f) – cada um deles sendo considerado um cromossomo do indivíduo. Operações clássicas seletivas e reprodutivas seriam aplicadas à população de indivíduos, de forma que a população resultante seja progressivamente composta de indivíduos mais aptos do que a população inicial, até que se selecione apenas o indivíduo mais apto ou, conforme chamamos nesta tese, até que seja obtido o conjunto $\mathbf{\Omega}$ de modelos (M, L) mais adequados à tarefa de predizer a série temporal.

Foi mostrado que, com um custo computacional moderado, para várias séries temporais encontradas no mundo real, o método DSE resulta em um erro de predição menor do que aquele obtido com a determinação dos centros pela técnica APC, ou do que aquele obtido com a atribuição dos centros através do algoritmo *k-means*. Quando

estendida pela heurística JPS, a predição DSE – JPS resultou em uma ulterior redução do erro de predição.

Referências Bibliográficas

[1] A. C. Harvey. Forecasting, Structural Time Series Models and the Kalman Filter, *Cambridge University Press, 1992.*

[2] A. Ossen. "Practical Tools for Derivative Instruments based on Nonlinear Time Series Prediction". *Department of Computer Science*, Technical University of Berlin. Berlin, Germany.

[3] A. S. Weigend and N. A. Gershenfeld. *Time Series Prediction: Forecasting the Future and Understanding the Past.* Addison-Wesley Publishing Company, 1994.

[4] A. S. Weigend, B. A. Huberman and D. E. Rumelhart. "Predicting the Future: a Connectionist Approach." *International Journal of Neural Systems*, vol 1, p.193-209, 1990.

[5] B. Mulgrew, "Applying Radial Basis Functions", *IEEE Signal Processing Magazine*, pp 50-65, 1996.

[6] B. Widrow and S. D. Stearns. Adaptive Signal Processing. Prentice-Hall, Englewood Cliff, 1985.

[7] B. Widrow, D. E. Rumelhart and M. A. Lehr. "Neural Networks: Applications in Industry, Business and Science". *Communications of the ACM, Vol.37, n°.3,* March, 1994.

[8] C. T. Chen, *Linear System Theory and Design*, Harcourt Brace College Publishers, 1984.

[9] C. L. Giles, S. Lawrence and A. C. Tsoi. "Rule Inference for Financial Prediction using Recurrent Neural Networks". *NEC Research Institute*, Department of Informatics, University of Wollongong, Australia.

[10] C. M. Bishop. "Mixture Density Networks". *Neural Computing Research Group.* Dept. of Computer Science and Applied Mathematics, Aston University, Birmingham, UK, February, 1994.

[11] C. M. Bishop. Neural Networks for Pattern Recognition. Oxford University Press, 1997.

[12] C. T. Lin and C. S. George Lee. *Neural Fuzzy Systems – A Neuro-Fuzzy Synergism to Intelligent Systems*, Prentice Hall, 1996.

[13] C. W. J. Granger and P. Newbold. *Forecasting Economic Time Series – Economic Theory and Mathematical Economics*, Academic Press, 1977.

[14] D. W. Pearson, N. C. Steele and R. F. Albrecht. *Artificial Neural Nets ans Genetic Algorithms – Proceedings of the International Conference in Alès, France*, Springer-Verlag Wien New York, 1995.

[15] D. Whitley. "A Genetic Algorithm Tutorial" – Technical Report CS-93-103. *Colorado State University – Department of Computer Science*, USA, 1993.

[16] E. Kreyszig, *Advanced Engineering Mathematics*, John Wiley & Sons, 1967.

[17] E. Oja and J. Karhunen. "On Stochastic Approximation of the Eigenvectors and Eigenvalues of the Expectation of a Random Matrix". *Journal of Mathematical Analysis and Applications*, 106:69-84, 1985.

[18] E. Oja. "Principal Components, Minor Components and Linear Neural Networks". *Neural Networks*, 5:927-935,1992.

[19] F. Ayres Jr., *Theory and Problems of Matrices*, Schaum's Outline Series in Mathematics, Mcraw-Hill Book Company, 1962.

[20] F. C. C. De Castro, M. C. F. De Castro e D. S. Arantes. "Controle de Taxa CBR em Sistemas MPEG com Redes Neurais Preditivas RBF". *Pub. FEEC 11/97 UNICAMP/FEEC/DECOM – Telebrás Contract 825/97 – PAPDS – Imagem – RT-002*, Dezembro, 1997.

[21] F. C. C. De Castro, M. C. F. de Castro and D. S. Arantes. "A Supervised Neural Constant Bit Rate Video Controller for MPEG2 Encoders". *1998 IEEE/SBT International Telecommunications Symposium*, pages 504-509, August, 1998.

[22] F. Fessant, S. Bengio and D. Collobert. "On the Prediction of Solar Activity Using Different Neural Network Models". *France Telecom*. Lannion Cedex, France.

[23] F. R. Gantmacher, *The Theory of Matrices*, vol.1, Chelsea Publishing Company, New York, NY, 1977.

[24] ftp://ftp.santafe.edu/pub/Time-Series

[25] G. E. P. Box, G. M. Jenkins and G. C. Reinsel. *Time Series Analysis*, Prentice Hall, 1994.

[26] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., John Hopkins University Press, 1996.

[27] http://ita.ee.lbl.gov/html/contrib/BC.html

[28] http://www.nettrade.com.br/

[29] http://www.stern.nyu.edu/~aweigend/TimeSeries/SantaFe.html

[30] J. Hertz, A. Krogh e R. G. Palmer, *Introduction to the Theory of Neural Computation*, Addison-Wesley, 1991.

[31] K. S. Shamugan and A. M. Breipohl, *Random Signals: Detection, Estimation and Data Analysis*, John Wiley & Sons, 1988.

[32] M. C. F. De Castro, F. C. C. De Castro and D. S. Arantes, "RBF Neural Networks with Centers Assignment via Karhunen-Loève Transform", *1999 IEEE International Joint Conference on Neural Networks,* Proc. CD-ROM - manuscript #0263 section 5.4, Washington DC, July 1999.

[33] M. C. F. De Castro, F. C. C. De Castro, D. S. Arantes and D. F. G. Azevedo. "A Fuzzy Neural CBR Channel Rate Controller for MPEG2 Encoders". *XVIII Simpósio Brasileiro de Telecomunicações - SBT2000,* Proc. CD-ROM Article 4120017, Gramado, RS, Brazil, September 2000.

[34] M. C. F. De Castro, F. C. C. De Castro, J. N. Amaral and P. R. G. Franco. "A Complex Valued Hebbian Learning Algorithm". *1998 IEEE World Congress on Computational Intelligence, International Joint Conference on Neural Networks,* pp. 1235-1238, Anchorage, Alaska, May, 1998.

[35] M. C. F. De Castro, F. C. C. De Castro, J. N. Amaral and P. R. G. Franco. "A New Training Algorithm to Reduce the Computational Complexity of Principal Component Analysis by Hebbian Learning". *III Congresso Brasileiro de Redes Neurais*, pages 7-11, Florianópolis, SC, July, 1997.

[36] M. C. F. De Castro, F. C. C. De Castro, J. N. Amaral and P. R. G. Franco. "Uma Formulação Complexa para o Algoritmo Hebbiano Generalizado Aplicada à Compressão de Imagens". *III Simpósio Brasileiro de Redes Neurais*, páginas 55-62, Recife, PE, Novembro, 1996.

[37] M. C. F. de Castro. "Algoritmo Hebbiano Generalizado para Extração dos Componentes Principais de um Conjunto de Dados no Domínio Complexo". Tese de Mestrado, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, RS, Brasil, 1996.

[38] M. H. Hassoun, *Fundamentals of Artificial Neural Networks*, MIT Press, 1995.

[39] "Neural Connection Application Study". *Fentress Incorporated.* SPSS P&S, 1999.

[40] "Neural Networks: The Application of Theoretical Foundations to Gilt Futures Trading". *School of Business & Management*, University College Northampton, England.

[41] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer-Verlag, 1987.

[42] P. J. McCann and B. L. Kalman. "A Neural Network Model for the Gold Market". *Department of Computer Science*. Washington University. St. Louis, Missouri.

[43] P. M. Embree e B. Kimble, *C Language Algorithms for Digital Signal Processing*, Prentice Hall, 1991.

[44] P. Sincak, M. Holecy and M. Ducai. "Computational Intelligence in Financial Cybernetics". *Computational Intelligence Group*, Laboratory of AI. Department of Cybernetics and Artiffcial Intelligence. Faculty of Electrical Engineering and Informatics. Technical University of Kosice, Slovak Republic.

[45] Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, 1965.

[46] R. Bronson, *Matrix Methods: An Introdution*, Academic Press Inc., San Diego, CA, 1991.

[47] R. D. Strum and D. E. Kirk, *First Principles of Discrete Systems and Digital Signal Processing*, Addison-Wesley, 1989.

[48] S. Bengio, F. Fessant, and D. Collobert. "A Connectionist System for Medium-Term Horizon Time Series Prediction". *France Telecom*, Centre National d'Etudes des Telecommunications, Lannion, FRANCE.

[49] S. Chen, S. McLaughlin and B. Mulgrew, "Complex-Valued Radial Basis Function Network, Part I: Network Architecture and Learning Algorithms", *Signal Processing*, vol. 35, pp 19-31, 1994.

[50] S. Haykin, *Adaptive Filter Theory*, 3rd ed., Prentice Hall, Upper Saddle River, New Jersey, 1996.

[51] S. Haykin, *Neural Networks*, 2nd ed., Prentice Hall, Upper Saddle River, New Jersey, 1999.

[52] S. Makridakis and S. C. Wheelwright. *Forecasting - Methods & Applications*, John Wiley & Sons, 1978.

[53] S. Makridakis and S. C. Wheelwright. *Forecasting – Methods for Management*, John Wiley & Sons, 1989.

[54] S. Makridakis, S. C. Wheelwright and V. E. McGee. *Forecasting - Methods and Applications*, John Wiley & Sons, 1983.

[55] U. Huebner, N. B. Abraham and C. O. Weiss. "Dimensions and entropies of chaotic intensity pulsations in a single-mode far-infrared NH3 laser." *Phys. Rev.* A 40, p. 6354, 1989.

[56] W. E. Leland and D. V. Wilson. "High Time-resolution Measurement and Analysis of LAN Traffic: Implications for LAN Interconnection." *IEEE INFOCOM'91*, pp. 1360-1366, April 1991.

[57] W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson. "On the Self-Similar Nature of Ethernet Traffic." *IEEE/ACM Transactions on Networking,* Vol.2, No.1, pp. 1-15, February 1994.

[58] W. H. Hsu. "Time Series Learning with Probabilistic Network Composites". Doctoral Thesis, Department of Computer Science, University of Illinois at Urbana-Champaign, 1998.

[59] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery. *Numerical Recipes in C*, 2nd ed., Cambridge Uniersity Press, 1992.

[60] X. Zhang and K. Thearling. "Non-Linear Time-Series Prediction by Systematic Data Exploration on a Massively Parallel Computer". *Santa Fe Institute Technical Report* 94-07-045.

[61] Y. B. Zeldovich and A. D. Myskis, *Elements of Applied Mathematics*, Mir Publishers, Moscow, USSR, 1976.

[62] Y. Saw, P. M. Grant, J. M. Hannah and B. Mulgrew. "Nonlinear predictive rate control for constant bit rate MPEG video coders". *IEEE Proceedings of ICASSP'97, 1997.*

[63] Y. Solano and H. Ikeda. "Performance Evaluation of Two Algorithms for Learning in ANN Based on a Real Financial Prediction Application". *IEICE Trans. Fundamentals*, vol. E 80 - A, n°.2, February 1997.

Apêndice A

Pseudo-Inversão Matricial por Decomposição em Valores Singulares

Este apêndice destina-se a uma explicação sucinta do método em que é baseado o algoritmo para a pseudo-inversão de Moore-Penrose [50] por Decomposição em Valores Singulares (SVD) de uma matriz, utilizado neste trabalho [59]. A SVD se aplica quando há a necessidade de inversão de matrizes singulares ou numericamente muito próximas a singulares.

Seja uma matriz $\mathbf{\Phi}$ de ordem $[N \times K]$. Há um teorema de Álgebra Linear que prova que qualquer matriz $\mathbf{\Phi}$ de ordem $[N \times K]$, cujo número de linhas N é maior ou igual ao número de colunas K, pode ser escrita como o produto de uma matriz $\mathbf{U}$ de ordem $[N \times K]$, cujas colunas são ortogonais, por uma matriz diagonal $\mathbf{S}$ de ordem $[K \times K]$ com elementos positivos ou zeros – os valores singulares – e o transposto de uma matriz ortogonal $\mathbf{V}$ de ordem $[K \times K]$ [8]. Isto é,

$$\mathrm{SVD}\{\mathbf{\Phi}_{[N\times K]}\} = \mathbf{U}_{[N\times K]} \cdot \mathbf{S}_{[K\times K]} \cdot \mathbf{V}^T{}_{[K\times K]} = \mathbf{\Phi}_{[N\times K]} \tag{A.1}$$

sendo a ortogonalidade das matrizes $\mathbf{U}$ e $\mathbf{V}$ expressa por $\mathbf{U}\mathbf{U}^T = 1$ e $\mathbf{V}\mathbf{V}^T = 1$, e onde

$$\mathbf{S} = \begin{bmatrix} s_0 & & & 0 \\ & s_1 & & \\ & & \ddots & \\ 0 & & & s_K \end{bmatrix} \tag{A.2}$$

Os elementos $s_0, s_1, \cdots, s_K$ que compõem a diagonal principal de $\mathbf{S}$, na Equação (A.2) são chamados valores singulares da matriz.

Como, devido à ortogonalidade, $\mathbf{U}^{-1} = \mathbf{U}^T$ e $\mathbf{V}^{-1} = \mathbf{V}^T$, e como $\mathbf{S}$ é uma matriz diagonal, a pseudo-inversa de $\mathbf{\Phi}$, denotada por $\mathbf{\Phi}^+$, é , então, determinada por

$$\mathbf{\Phi}^+{}_{[K\times N]} = \mathbf{V}_{[K\times K]} \cdot \begin{bmatrix} \frac{1}{s_0} & & 0 \\ & \frac{1}{s_1} & \\ & \ddots & \\ 0 & & \frac{1}{s_K} \end{bmatrix}_{[K\times K]} \cdot \mathbf{U}^T{}_{[K\times N]} \tag{A.3}$$

Apêndice B

Decomposição em Sub-Espaços através da Transformação Karhunen-Loève

A Transformação Karhunen-Loève (KLT) projeta um conjunto $\mathbf{X}$ de vetores de dados $\underline{x} \in \Re^M$ sobre uma base ortonormal em $\Re^M$ formada pelo conjunto dos M auto-vetores $\underline{e}_m \in \Re^M$, $m = 0, 1, \cdots, M-1$, da matriz de covariância de $\mathbf{X}$. A base será, portanto, orientada de acordo com as direções de maior variância de $\mathbf{X}$ em $\Re^M$ [17] [18] [37] [51].

A m-ésima projeção de $\mathbf{X}$ sobre a direção do auto-vetor $\underline{e}_m$ é chamada de m-ésimo sub-espaço (ou m-ésimo componente principal). O m-ésimo auto-valor λ_m associado ao auto-vetor $\underline{e}_m$ corresponde à variância do m-ésimo sub-espaço de $\mathbf{X}$. Ainda, a variância de cada sub-espaço é um máximo local, no universo de todas as variâncias resultantes da projeção de $\mathbf{X}$ sobre todas as possíveis direções em $\Re^M$. Embora qualquer espaço de dimensão menor do que $\Re^M$ seja um sub-espaço de $\Re^M$, esta tese refere-se muitas vezes a um sub-espaço de $\mathbf{X}$ como o conjunto dos vetores de $\mathbf{X}$ que concentram-se de forma alinhada ao longo de uma particular direção em $\Re^M$. Ainda que esta referência não seja precisa, julgou-se válido aplicá-la devido ao conceito intuitivo nela implícito.

Uma importante utilização da Transformação Karhunen-Loève, também conhecida por Análise dos Componentes Principais (PCA) é a da redução dimensional. Embora nesta tese o interesse não seja a redução dimensional, é interessante salientar que esta característica torna a KLT bastante popular em processamento digital de sinais por permitir que as informações mais significativas contidas em um sinal possam ser representadas, mediante a introdução de algum erro considerado aceitável, em um espaço de dados de

menores dimensões do que a dimensão original $\Re^M$. Como cada sub-espaço ou componente principal é orientado de acordo com as direções de maior variância de $\mathbf{X}$ em $\Re^M$, os sub-espaços de menor variância podem ser descartados – o que resultará em uma redução dimensional ótima no sentido do Erro Médio Quadrático (MSE) [37].

O desenvolvimento do método para aplicação da KLT à $\mathbf{X}$ apresentado neste apêndice segue a proposta de Haykin em [51].

Seja $\mathbf{X}$ um processo estocástico vetorial de média zero representado pelo conjunto $\mathbf{X}$ de vetores $\underline{x} \in \Re^M$. Caso $\mathbf{X}$ não possua média zero – e, portanto, caso o vetor aleatório $\underline{x}$ não possua média zero – o vetor média deverá ser subtraído dos vetores de $\mathbf{X}$ antes de iniciar a transformação.

Seja $\underline{e} \in \Re^M$ um vetor unitário e adimensional qualquer, sobre o qual será projetado um vetor $\underline{x} \in \mathbf{X}$. Sendo $\underline{e}$ um vetor unitário, a norma Euclidiana de $\underline{e}$ é dada pela Equação (B.1).

$$\| \underline{e} \| = \sqrt{\underline{e}^T \underline{e}} = 1 \tag{B.1}$$

Como $\underline{e}$ é unitário e adimensional, o tamanho (norma Euclidiana) do vetor que resulta da projeção de $\underline{x}$ sobre $\underline{e}$, denominada de projeção a, possui a mesma unidade dimensional de $\underline{x}$ e é dado pela Equação (B.2).

$$a = \underline{x}^T \underline{e} = \underline{e}^T \underline{x} \tag{B.2}$$

Portanto, a define o tamanho da projeção do vetor $\underline{x}$ na direção do vetor $\underline{e}$. A projeção a também é uma variável aleatória, com média e variância associadas à estatística do vetor aleatório $\underline{x}$.

A média da projeção a é dada por

$$E\{a\} = E\{\underline{e}^T \underline{x}\} \tag{B.3}$$

$$E\{a\} = \underline{e}^T E\{\underline{x}\} = 0 \tag{B.4}$$

onde $E\{\}$ é o operador que resulta no valor esperado da variável aleatória argumento.

A variância da projeção a é dada por

$$\sigma_a^2 = E\left\{a^2\right\} \tag{B.5}$$

$$\sigma_a^2 = E\left\{\left(\underline{e}^T \underline{x}\right)\left(\underline{x}^T \underline{e}\right)\right\} \tag{B.6}$$

$$\sigma_a^2 = \underline{e}^T E\left\{\underline{x}\,\underline{x}^T\right\}\underline{e} \tag{B.7}$$

onde $E\left\{\underline{x}\,\underline{x}^T\right\}$ é a matriz de covariância $\mathbf{C}_x$, de dimensões $M \times M$, do conjunto $\mathbf{X}$ de vetores $\underline{x} \in \Re^M$.

Portanto, a Equação (B.7) pode ser escrita como

$$\sigma_a^2 = \underline{e}^T \mathbf{C}_x \underline{e} \tag{B.8}$$

onde observa-se que a variância σ^2 da projeção a é uma função do vetor unitário $\underline{e}$. Expressando matematicamente,

$$\sigma_a^2 = f\left(\underline{e}\right) \tag{B.9}$$

A KLT objetiva determinar o conjunto de vetores $\underline{e}$ para cuja direção em $\Re^M$ a variância da projeção a é máxima. Portanto, a KLT varia o vetor $\underline{e}$ no universo de todas as possíveis direções em $\Re^M$ na busca daquela direção que resulta no máximo $f\left(\underline{e}\right)$. Para que a norma de $\underline{e}$ não influa no resultado de $f\left(\underline{e}\right)$ durante a busca, ela deve ser constante e unitária. Esta é a razão da condição preliminar definida em (B.1).

Se $\underline{e}$ alinha-se com uma direção em $\Re^M$ do conjunto $\mathbf{X}$ tal que nesta direção $f\left(\underline{e}\right)$ resulta em um máximo local dentro do universo de busca, então, devido à menor declividade de $f\left(\underline{e}\right)$ nas vizinhanças de um máximo local, para qualquer pequena variação $\delta\underline{e}$ do vetor unitário $\underline{e}$ temos que

$$f\left(\underline{e}+\delta\underline{e}\right) \approx f\left(\underline{e}\right) \tag{B.10}$$

Embora (B.10) também seja válida nas vizinhanças de um mínimo local, esta ambigüidade é resolvida quando utiliza-se a *eigen*-estrutura de $\mathbf{C}_x$ [51].

Combinando as Equações (B.8) e (B.9) obtém-se que $f(\underline{e})=\underline{e}^T\mathbf{C}_x\,\underline{e}$. Assumindo que o valor de $\delta\underline{e}$ seja suficientemente pequeno de modo que $f(\underline{e})$ não se afaste das vizinhanças do máximo local , pode-se admitir a igualdade em (B.10). Portanto,

$$f(\underline{e})=\underline{e}^T\mathbf{C}_x\,\underline{e}=f(\underline{e}+\delta\underline{e})=(\underline{e}+\delta\underline{e})^T\mathbf{C}_x(\underline{e}+\delta\underline{e}) \tag{B.11}$$

$$f(\underline{e}+\delta\underline{e})=\underline{e}^T\mathbf{C}_x\,\underline{e}+\underline{e}^T\mathbf{C}_x\delta\underline{e}+\delta\underline{e}^T\mathbf{C}_x\,\underline{e}+\delta\underline{e}^T\mathbf{C}_x\,\delta\underline{e} \tag{B.12}$$

Como a matriz de covariância $\mathbf{C}_x$ do conjunto $\mathbf{X}$ é uma matriz simétrica, tem-se que

$$\underline{e}^T\mathbf{C}_x\delta\underline{e}=\delta\underline{e}^T\mathbf{C}_x\underline{e} \tag{B.13}$$

e a Equação (B.12) pode ser reescrita sob a forma

$$f(\underline{e}+\delta\underline{e})=\underline{e}^T\mathbf{C}_x\underline{e}+2\delta\underline{e}^T\mathbf{C}_x\underline{e}+\delta\underline{e}^T\mathbf{C}_x\delta\underline{e} \tag{B.14}$$

Como $\delta\underline{e}$ é muito pequeno, o termo $(\delta\underline{e}^T\mathbf{C}_x\,\delta\underline{e})$ no lado direito de (B.12) pode ser desconsiderado. De (B.14), (B.8) e (B.9) pode-se escrever que

$$f(\underline{e}+\delta\underline{e})=f(\underline{e})+2\delta\underline{e}^T\mathbf{C}_x\underline{e} \tag{B.15}$$

Ao levar a Equação (B.10) à Equação (B.15) conclui-se que

$$\delta\underline{e}^T\mathbf{C}_x\underline{e}=0 \tag{B.16}$$

Mas, para que a norma de $\underline{e}$ não influa no resultado de $f(\underline{e})$ durante a busca, ela deve ser mantida constante e unitária. Portanto, são admissíveis apenas as perturbações $\delta\underline{e}$ para as quais a norma do vetor perturbado $\underline{e}+\delta\underline{e}$ permaneça unitária, ou seja

$$\|\underline{e}+\delta\underline{e}\|=1 \tag{B.17}$$

o que equivale a dizer que

$$(\underline{e}+\delta\underline{e})^T(\underline{e}+\delta\underline{e})=1 \tag{B.18}$$

Expandindo o produto $(\underline{e}+\delta\underline{e})^T(\underline{e}+\delta\underline{e})$, temos

$$(\underline{e}+\delta\underline{e})^T(\underline{e}+\delta\underline{e})=\underline{e}^T\underline{e}+\underline{e}^T\delta\underline{e}+\delta\underline{e}^T\underline{e}+\delta\underline{e}^T\delta\underline{e} \tag{B.19}$$

onde

$$\underline{e}^T\delta\underline{e}=\delta\underline{e}^T\underline{e} \tag{B.20}$$

e onde o termo $\delta\underline{e}^T\delta\underline{e}$ pode ser desconsiderado. Portanto,

$$(\underline{e}+\delta\underline{e})^T(\underline{e}+\delta\underline{e})=\underline{e}^T\underline{e}+2\,\delta\underline{e}^T\underline{e} \tag{B.21}$$

como, de (B.18), $(\underline{e}+\delta\underline{e})^T(\underline{e}+\delta\underline{e})=1$ então, de (B.21)

$$\underline{e}^T\underline{e}+2\,\delta\underline{e}^T\underline{e}=1 \tag{B.22}$$

mas, de (B.1), $\underline{e}^T\underline{e}=1$ em (B.22), e portanto

$$\delta\underline{e}^T\underline{e}=0 \tag{B.23}$$

Infere-se da Equação (B.23) que as variações $\delta\underline{e}$ devem ser ortogonais a $\underline{e}$ durante o processo de busca da direção de máximo $f(\underline{e})$. Este é o único tipo de variação permitida a $\underline{e}$ no espaço de busca $\Re^M$ tal que a Equação (B.17) seja obedecida.

Nas Equações (B.16) e (B.23) residem as duas condições simultâneas a serem respeitadas para a determinação dos vetores $\underline{e}$ para os quais $f(\underline{e})$ terá valor máximo. Somando-se estas equações, pode-se escrever que

$$\delta\underline{e}^T\mathbf{C}_x\underline{e}-\lambda\delta\underline{e}^T\underline{e}=\mathbf{C}_x(\delta\underline{e})^T\underline{e}-\lambda(\delta\underline{e})^T\underline{e}=0 \tag{B.24}$$

onde o fator de escala λ foi introduzido para compatibilizar a unidade dimensional do vetor unitário $\underline{e}$ (adimensional por definição) com a unidade dimensional da matriz $\mathbf{C}_x$ representativa da covariância do conjunto $\mathbf{X}$. Por exemplo, se os vetores do conjunto $\mathbf{X}$

representam uma grandeza cuja unidade dimensional é $\left[\frac{metro}{segundo}\right]$, λ terá unidade dimensional $\left[\frac{metro^2}{segundo^2}\right]$.

Reescrevendo a equação (B.24),

$$(\delta \underline{e})^T (\mathbf{C}_x \underline{e} - \lambda \underline{e}) = 0 \tag{B.25}$$

Para que a condição posta na Equação (B.25) seja satisfeita, é necessário e suficiente que

$$\mathbf{C}_x \underline{e} = \lambda \underline{e} \tag{B.26}$$

que é a equação que define os vetores $\underline{e}$ para os quais $f(\underline{e})$ tem valor máximo.

A Equação (B.26) é reconhecida como a equação dos auto-vetores da matriz $\mathbf{C}_x$ e apresenta um conjunto de soluções não-triviais para aqueles valores de λ que são denominados os auto-valores de $\mathbf{C}_x$. Seja a k-ésima solução de (B.26) tal que

$$(\mathbf{C}_x - \lambda_k \mathbf{I})\underline{e}_k = \underline{0} \tag{B.27}$$

Por serem mapeados no vetor nulo $\underline{0}$ através de $(\mathbf{C}_x - \lambda_k \mathbf{I})$, os auto-vetores $\underline{e}_k$ pertencem ao espaço nulo da transformação linear (B.27), isto é,

$$\underline{e}_k \in \Re^{\mathcal{N}} \tag{B.28}$$

onde $\Re^{\mathcal{N}}$ é denominado espaço nulo da matriz $(\mathbf{C}_x - \lambda_k \mathbf{I})$, denotado por $\mathcal{N}\{\mathbf{C}_x - \lambda_k \mathbf{I}\}$. A dimensão $\Re^{\mathcal{N}}$ do espaço nulo de uma transformação linear define o número $\mathcal{N}$ de vetores linearmente independentes que levam a transformação ao vetor nulo $\underline{0}$, os quais, portanto, definem uma base geradora de $\Re^{\mathcal{N}}$ [8][23][46].

O *rank* da matriz $(\mathbf{C}_x - \lambda_k \mathbf{I})$, denotado $\mathcal{R}\{\mathbf{C}_x - \lambda_k \mathbf{I}\}$, é o espaço $\Re^{\mathcal{R}}$ gerado pelo número $\mathcal{R}$ de vetores-colunas de $(\mathbf{C}_x - \lambda_k \mathbf{I})$ que são linearmente independentes. Um

teorema de Teoria de Matrizes prova que a dimensão M da matriz quadrada $(\mathbf{C}_x - \lambda_k \mathbf{I})_{[M\times M]}$ é igual à soma das dimensões de seu *rank* e espaço nulo [8][23][46], isto é

$$\mathcal{N} + \mathcal{R} = M = \dim\{\mathbf{C}_x - \lambda_k \mathbf{I}\} \tag{B.29}$$

Por definição, os auto-vetores $\underline{e}_k$ formam uma base ortonormal no espaço por eles gerado. Isto significa que em (B.27) o conjunto de auto-vetores $\underline{e}_k$ define a própria base para o espaço nulo $\Re^{\mathcal{N}}$. Portanto, de (B.29), para que a dimensão $\mathcal{N}$ do espaço nulo seja igual à dimensão da matriz $(\mathbf{C}_x - \lambda_k \mathbf{I})$, é necessário que o *rank* da matriz seja nulo ($\mathcal{R} = 0$). Assim $\mathcal{N} = M = \dim\{\mathbf{C}_x - \lambda_k \mathbf{I}\}$, de forma que existirão M vetores $\underline{e}_k$ linearmente independentes. Mas, uma matriz que possui *rank* nulo é uma matriz singular. Isto é, se

$$\mathcal{R}\{\mathbf{C}_x - \lambda_m \mathbf{I}\} = 0 \Rightarrow (\mathbf{C}_x - \lambda_m \mathbf{I}) \text{ é singular, } m = 0, 1, \cdots, M-1 \tag{B.30}$$

e, se a matriz $(\mathbf{C}_x - \lambda_m \mathbf{I})$ é singular, pode-se escrever que

$$\det\{\mathbf{C}_x - \lambda_m \mathbf{I}\} = 0 \tag{B.31}$$

onde $\det\{\cdot\}$ representa o determinante da matriz argumento.

A Equação (B.31) é a equação utilizada para a determinação do m-ésimo auto-valor λ_m da matriz $\mathbf{C}_x$ representativa da covariância do conjunto $\mathbf{X}$, $m = 0, 1, \cdots, M-1$. Como $\mathbf{C}_x$ é uma matriz simétrica, possui auto-valores reais e não-negativos [8][23][46].

Uma vez determinados os M auto-valores $\lambda_0, \lambda_1, \cdots, \lambda_{M-1}$ da matriz $\mathbf{C}_x$ através de (B.31), obtém-se os auto-vetores associados $\underline{e}_0, \underline{e}_1, \cdots, \underline{e}_{M-1}$ através de (B.27).

Assim, de (B.27) pode-se escrever

$$\mathbf{C}_x \, \underline{e}_m = \lambda_m \, \underline{e}_m, \; m = 0, 1, \cdots, M-1 \tag{B.32}$$

Seja $\mathbf{\Lambda}$ uma matriz diagonal, $M \times M$, composta pelos auto-valores da matriz $\mathbf{C}_x$, de acordo com a Equação (B.32),

$$\Lambda = \mathrm{diag}\left(\lambda_0, \lambda_1, \cdots, \lambda_j, \cdots, \lambda_{M-1}\right) \tag{B.33}$$

onde os M auto-valores estão ordenados de forma decrescente, de modo que λ_0 seja o maior auto-valor.

Seja $\mathbf{E}$ uma matriz $M \times M$ cujas colunas são formadas pelos M auto-vetores $\underline{e}_0, \underline{e}_1, \cdots, \underline{e}_{M-1}$ dados pela Equação (B.32) associados aos auto-valores $\lambda_0, \lambda_1, \cdots, \lambda_{M-1}$.

$$\mathbf{E} = \left[\underline{e}_0, \underline{e}_1, \cdots, \underline{e}_j, \cdots, \underline{e}_{M-1}\right] \tag{B.34}$$

De acordo com (B.33) e (B.34), a Equação (B.32) pode ser reescrita como

$$\mathbf{C}_x\mathbf{E} = \mathbf{E}\Lambda \tag{B.35}$$

Como os auto-vetores de $\mathbf{C}_x$ (vetores-coluna de $\mathbf{E}$) são ortonormais, a matriz $\mathbf{E}$ é dita ortonormal. Portanto, os vetores-coluna de $\mathbf{E}$ satisfazem a condição de ortonormalidade,

$$\underline{e}_i^T \underline{e}_j = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases} \tag{B.36}$$

A Equação (B.36) pode também ser escrita sob a forma

$$\mathbf{E}^T\mathbf{E} = \mathbf{I} \tag{B.37}$$

de onde vem que

$$\mathbf{E}^T = \mathbf{E}^{-1} \tag{B.38}$$

A Equação (B.35) pode ser reescrita, sob a forma de uma Transformação de Similaridade [8][23][46]. Pré-multiplicando ambos os lados da Equação (B.35) por $\mathbf{E}^T$,

$$\mathbf{E}^T\mathbf{C}_x\mathbf{E} = \mathbf{E}^T\mathbf{E}\Lambda \tag{B.39}$$

De (B.39) e (B.38),

$$\mathbf{E}^T\mathbf{C}_x\mathbf{E} = \mathbf{E}^{-1}\mathbf{C}_x\mathbf{E} = \mathbf{E}^{-1}\mathbf{E}\Lambda \tag{B.40}$$

ou

$$\mathbf{E}^{\mathrm{T}}\mathbf{C}_{\mathrm{x}}\mathbf{E}=\mathbf{\Lambda} \tag{B.41}$$

que, na forma expandida, resulta em

$$\underline{e}_j^T\mathbf{C}_{\mathrm{x}}\underline{e}_k=\begin{cases}\lambda_j, & k=j\\ 0, & k\neq j\end{cases} \tag{B.42}$$

Das Equações (B.8) e (B.9) tem-se

$$\sigma_a^2=f\left(\underline{e}\right)=\underline{e}^T\mathbf{C}_{\mathrm{x}}\underline{e} \tag{B.43}$$

Comparando as Equações (B.42) e (B.43), verifica-se que

$$f\left(\underline{e}_m\right)=\lambda_m,\ m=0,1,\cdots,M-1 \tag{B.44}$$

Portanto, os M auto-vetores $\underline{e}_m$ da matriz $\mathbf{C}_{\mathrm{x}}$ representam as M direções principais ao longo das quais a variância $\sigma_m^2=f\left(\underline{e}_m\right)$ é máxima e dada por λ_m.

Apêndice C

Tabela C.1: Íntegra da Tabela 6.2, Capítulo 6, Seção 6.2 (1/8)

Indivíduo (M, L, NMSE_f)

M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f
2	20	1.15	2	65	1.05	2	110	1.08	3	25	1.14
2	21	1.08	2	66	1.07	2	111	1.04	3	26	1.12
2	22	1.1	2	67	1.07	2	112	1.05	3	27	1.12
2	23	1.1	2	68	1.07	2	113	1.03	3	28	1.09
2	24	1.09	2	69	1.04	2	114	1.06	3	29	1.1
2	25	1.08	2	70	1.02	2	115	1.04	3	30	1.07
2	26	1.04	2	71	1.01	2	116	1.07	3	31	1.06
2	27	1.05	2	72	1.02	2	117	1.06	3	32	1.11
2	28	1.05	2	73	1.04	2	118	1.05	3	33	1.12
2	29	1.07	2	74	1.02	2	119	1.07	3	34	1.13
2	30	1.03	2	75	1.04	2	120	1.06	3	35	1.13
2	31	1.05	2	76	1.03	2	121	1.06	3	36	1.12
2	32	1.08	2	77	1.04	2	122	1.07	3	37	1.14
2	33	1.08	2	78	1.03	2	123	1.08	3	38	1.14
2	34	1.05	2	79	1.05	2	124	1.06	3	39	1.15
2	35	1.06	2	80	1.02	2	125	1.05	3	40	1.14
2	36	1.07	2	81	1.03	2	126	1.05	3	41	1.13
2	37	1.07	2	82	0.998	2	127	1.06	3	42	1.1
2	38	1.08	2	83	1.03	2	128	1.07	3	43	1.1
2	39	1.08	2	84	1	2	129	1.07	3	44	1.09
2	40	1.09	2	85	1.02	2	130	1.06	3	45	1.1
2	41	1.1	2	86	1.01	2	131	1.1	3	46	1.09
2	42	1.1	2	87	1.01	2	132	1.1	3	47	1.08
2	43	1.09	2	88	0.991	2	133	1.1	3	48	1.12
2	44	1.08	2	89	0.99	2	134	1.07	3	49	1.07
2	45	1.06	2	90	1	2	135	1.07	3	50	1.07
2	46	1.07	2	91	1.01	2	136	1.06	3	51	1.08
2	47	1.08	2	92	1.02	2	137	1.05	3	52	1.07
2	48	1.08	2	93	1.02	2	138	1.04	3	53	1.08
2	49	1.13	2	94	1	2	139	1.04	3	54	1.07
2	50	1.09	2	95	1.02	2	140	1.03	3	55	1.08
2	51	1.08	2	96	1.03	2	141	1.04	3	56	1.09
2	52	1.09	2	97	1.03	2	142	1.03	3	57	1.08
2	53	1.09	2	98	1.02	2	143	1.04	3	58	1.08
2	54	1.07	2	99	1.02	2	144	1.05	3	59	1.08
2	55	1.07	2	100	1.01	2	145	1.04	3	60	1.08
2	56	1.07	2	101	1.03	2	146	1.04	3	61	1.07
2	57	1.06	2	102	1.03	2	147	1.04	3	62	1.08
2	58	1.08	2	103	1.02	2	148	1.05	3	63	1.07
2	59	1.08	2	104	1.01	2	149	1.03	3	64	1.06
2	60	1.07	2	105	1.01	3	20	1.25	3	65	1.05
2	61	1.05	2	106	1.01	3	21	1.26	3	66	1.07
2	62	1.06	2	107	1.02	3	22	1.25	3	67	1.07
2	63	1.06	2	108	1.09	3	23	1.15	3	68	1.07
2	64	1.07	2	109	1.07	3	24	1.16	3	69	1.05

Tabela C.1: Íntegra da Tabela 6.2, Capítulo 6, Seção 6.2 (2/8)

Indivíduo (M, L, NMSE_f)

M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f
3	70	1.03	3	115	1.06	4	30	1.1	4	75	1.05
3	71	1.04	3	116	1.06	4	31	1.1	4	76	1.07
3	72	1.04	3	117	1.07	4	32	1.09	4	77	1.06
3	73	1.04	3	118	1.07	4	33	1.11	4	78	1.03
3	74	1.03	3	119	1.08	4	34	1.12	4	79	1.04
3	75	1.04	3	120	1.07	4	35	1.18	4	80	1.04
3	76	1.04	3	121	1.07	4	36	1.17	4	81	1.06
3	77	1.05	3	122	1.07	4	37	1.14	4	82	1.05
3	78	1.03	3	123	1.08	4	38	1.14	4	83	1.04
3	79	1.03	3	124	1.06	4	39	1.16	4	84	1.06
3	80	1.02	3	125	1.06	4	40	1.12	4	85	1.01
3	81	1.03	3	126	1.07	4	41	1.16	4	86	1.01
3	82	1.02	3	127	1.06	4	42	1.14	4	87	1.04
3	83	1.03	3	128	1.08	4	43	1.12	4	88	1.04
3	84	1	3	129	1.07	4	44	1.12	4	89	1.02
3	85	1.03	3	130	1.1	4	45	1.15	4	90	1.03
3	86	1.02	3	131	1.1	4	46	1.13	4	91	1.01
3	87	1.02	3	132	1.12	4	47	1.15	4	92	1.02
3	88	1.01	3	133	1.1	4	48	1.08	4	93	1.02
3	89	1.02	3	134	1.09	4	49	1.1	4	94	1.02
3	90	1.02	3	135	1.09	4	50	1.1	4	95	1.03
3	91	1.01	3	136	1.07	4	51	1.1	4	96	1.01
3	92	1.02	3	137	1.06	4	52	1.08	4	97	1.02
3	93	1.01	3	138	1.06	4	53	1.1	4	98	1.01
3	94	1.01	3	139	1.06	4	54	1.09	4	99	0.999
3	95	1.02	3	140	1.05	4	55	1.11	4	100	1.01
3	96	1.02	3	141	1.05	4	56	1.08	4	101	1.01
3	97	1.01	3	142	1.06	4	57	1.07	4	102	1.01
3	98	1.02	3	143	1.06	4	58	1.08	4	103	1.02
3	99	1.01	3	144	1.06	4	59	1.09	4	104	0.997
3	100	1.02	3	145	1.06	4	60	1.09	4	105	1.04
3	101	1.01	3	146	1.06	4	61	1.08	4	106	1.05
3	102	1.02	3	147	1.06	4	62	1.12	4	107	1.05
3	103	1.02	3	148	1.06	4	63	1.09	4	108	1.04
3	104	1.02	3	149	1.07	4	64	1.08	4	109	1.05
3	105	1.03	4	20	1.28	4	65	1.07	4	110	1.06
3	106	1.04	4	21	1.23	4	66	1.08	4	111	1.05
3	107	1.06	4	22	1.24	4	67	1.07	4	112	1.06
3	108	1.06	4	23	1.17	4	68	1.07	4	113	1.07
3	109	1.07	4	24	1.17	4	69	1.06	4	114	1.05
3	110	1.06	4	25	1.18	4	70	1.05	4	115	1.06
3	111	1.06	4	26	1.12	4	71	1.06	4	116	1.05
3	112	1.05	4	27	1.11	4	72	1.06	4	117	1.09
3	113	1.07	4	28	1.08	4	73	1.05	4	118	1.09
3	114	1.06	4	29	1.06	4	74	1.08	4	119	1.07

Tabela C.1: Íntegra da Tabela 6.2, Capítulo 6, Seção 6.2 (3/8)												
Indivíduo (M, L, NMSE_f)												
M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f
4	120	1.08	5	35	1.17	5	80	1.07	5	125	1.1
4	121	1.09	5	36	1.14	5	81	1.08	5	126	1.13
4	122	1.08	5	37	1.2	5	82	1.08	5	127	1.14
4	123	1.09	5	38	1.2	5	83	1.07	5	128	1.15
4	124	1.08	5	39	1.2	5	84	1.04	5	129	1.17
4	125	1.08	5	40	1.18	5	85	1.07	5	130	1.17
4	126	1.09	5	41	1.14	5	86	1.05	5	131	1.16
4	127	1.1	5	42	1.16	5	87	1.03	5	132	1.13
4	128	1.11	5	43	1.14	5	88	1.03	5	133	1.15
4	129	1.11	5	44	1.15	5	89	1.01	5	134	1.15
4	130	1.13	5	45	1.17	5	90	1.04	5	135	1.13
4	131	1.15	5	46	1.18	5	91	1.03	5	136	1.15
4	132	1.13	5	47	1.13	5	92	1.04	5	137	1.11
4	133	1.12	5	48	1.14	5	93	1.05	5	138	1.12
4	134	1.11	5	49	1.1	5	94	1.05	5	139	1.13
4	135	1.11	5	50	1.13	5	95	1.04	5	140	1.12
4	136	1.09	5	51	1.13	5	96	1.02	5	141	1.11
4	137	1.08	5	52	1.11	5	97	1.02	5	142	1.12
4	138	1.1	5	53	1.15	5	98	1.03	5	143	1.1
4	139	1.07	5	54	1.11	5	99	1.03	5	144	1.11
4	140	1.08	5	55	1.12	5	100	1.06	5	145	1.12
4	141	1.07	5	56	1.09	5	101	1.05	5	146	1.1
4	142	1.06	5	57	1.12	5	102	1.06	5	147	1.1
4	143	1.07	5	58	1.15	5	103	1.05	5	148	1.11
4	144	1.07	5	59	1.1	5	104	1.06	5	149	1.1
4	145	1.06	5	60	1.13	5	105	1.1	6	20	1.56
4	146	1.07	5	61	1.12	5	106	1.07	6	21	1.55
4	147	1.06	5	62	1.09	5	107	1.07	6	22	1.61
4	148	1.08	5	63	1.08	5	108	1.08	6	23	1.48
4	149	1.08	5	64	1.09	5	109	1.07	6	24	1.4
5	20	1.44	5	65	1.07	5	110	1.05	6	25	1.4
5	21	1.39	5	66	1.1	5	111	1.09	6	26	1.34
5	22	1.42	5	67	1.1	5	112	1.09	6	27	1.27
5	23	1.29	5	68	1.11	5	113	1.09	6	28	1.28
5	24	1.31	5	69	1.08	5	114	1.07	6	29	1.26
5	25	1.24	5	70	1.05	5	115	1.1	6	30	1.27
5	26	1.23	5	71	1.05	5	116	1.07	6	31	1.21
5	27	1.18	5	72	1.03	5	117	1.09	6	32	1.21
5	28	1.18	5	73	1.06	5	118	1.12	6	33	1.19
5	29	1.15	5	74	1.06	5	119	1.13	6	34	1.22
5	30	1.18	5	75	1.07	5	120	1.1	6	35	1.21
5	31	1.14	5	76	1.07	5	121	1.11	6	36	1.16
5	32	1.16	5	77	1.07	5	122	1.1	6	37	1.24
5	33	1.14	5	78	1.06	5	123	1.09	6	38	1.26
5	34	1.18	5	79	1.07	5	124	1.09	6	39	1.25

Tabela C.1: Íntegra da Tabela 6.2, Capítulo 6, Seção 6.2 (4/8)

Indivíduo (M, L, NMSE_f)

M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f
6	40	1.23	6	85	1.13	6	130	1.19	7	45	1.16
6	41	1.12	6	86	1.08	6	131	1.13	7	46	1.15
6	42	1.17	6	87	1.08	6	132	1.18	7	47	1.13
6	43	1.18	6	88	1.07	6	133	1.14	7	48	1.15
6	44	1.19	6	89	1.07	6	134	1.14	7	49	1.15
6	45	1.23	6	90	1.05	6	135	1.15	7	50	1.15
6	46	1.14	6	91	1.06	6	136	1.16	7	51	1.17
6	47	1.15	6	92	1.09	6	137	1.12	7	52	1.17
6	48	1.17	6	93	1.04	6	138	1.14	7	53	1.14
6	49	1.15	6	94	1.04	6	139	1.14	7	54	1.14
6	50	1.12	6	95	1.03	6	140	1.12	7	55	1.14
6	51	1.15	6	96	1.05	6	141	1.11	7	56	1.12
6	52	1.14	6	97	1.04	6	142	1.11	7	57	1.15
6	53	1.14	6	98	1.05	6	143	1.1	7	58	1.15
6	54	1.11	6	99	1.04	6	144	1.11	7	59	1.17
6	55	1.13	6	100	1.06	6	145	1.09	7	60	1.15
6	56	1.13	6	101	1.07	6	146	1.09	7	61	1.14
6	57	1.14	6	102	1.07	6	147	1.09	7	62	1.16
6	58	1.15	6	103	1.08	6	148	1.1	7	63	1.11
6	59	1.16	6	104	1.12	6	149	1.11	7	64	1.12
6	60	1.13	6	105	1.11	7	20	1.82	7	65	1.15
6	61	1.13	6	106	1.08	7	21	1.73	7	66	1.09
6	62	1.12	6	107	1.1	7	22	1.73	7	67	1.13
6	63	1.15	6	108	1.1	7	23	1.5	7	68	1.13
6	64	1.09	6	109	1.09	7	24	1.55	7	69	1.1
6	65	1.15	6	110	1.08	7	25	1.43	7	70	1.1
6	66	1.1	6	111	1.12	7	26	1.43	7	71	1.06
6	67	1.11	6	112	1.11	7	27	1.3	7	72	1.08
6	68	1.12	6	113	1.07	7	28	1.27	7	73	1.09
6	69	1.1	6	114	1.09	7	29	1.3	7	74	1.06
6	70	1.06	6	115	1.12	7	30	1.29	7	75	1.08
6	71	1.09	6	116	1.11	7	31	1.2	7	76	1.04
6	72	1.08	6	117	1.11	7	32	1.27	7	77	1.11
6	73	1.07	6	118	1.11	7	33	1.26	7	78	1.07
6	74	1.09	6	119	1.1	7	34	1.24	7	79	1.08
6	75	1.05	6	120	1.1	7	35	1.25	7	80	1.05
6	76	1.09	6	121	1.11	7	36	1.25	7	81	1.08
6	77	1.07	6	122	1.1	7	37	1.23	7	82	1.09
6	78	1.07	6	123	1.14	7	38	1.26	7	83	1.09
6	79	1.09	6	124	1.13	7	39	1.21	7	84	1.08
6	80	1.07	6	125	1.1	7	40	1.24	7	85	1.1
6	81	1.07	6	126	1.1	7	41	1.23	7	86	1.09
6	82	1.09	6	127	1.14	7	42	1.25	7	87	1.06
6	83	1.06	6	128	1.18	7	43	1.22	7	88	1.08
6	84	1.09	6	129	1.17	7	44	1.23	7	89	1.05

Tabela C.1: Íntegra da Tabela 6.2, Capítulo 6, Seção 6.2 (5/8)

Indivíduo $(M, L, \mathrm{NMSE_f})$

M	L	$\mathrm{NMSE_f}$	M	L	$\mathrm{NMSE_f}$	M	L	$\mathrm{NMSE_f}$	M	L	$\mathrm{NMSE_f}$
7	90	1.08	7	135	1.12	8	50	1.17	8	95	1.05
7	91	1.05	7	136	1.1	8	51	1.19	8	96	1.09
7	92	1.06	7	137	1.12	8	52	1.15	8	97	1.09
7	93	1.03	7	138	1.12	8	53	1.14	8	98	1.07
7	94	1.01	7	139	1.13	8	54	1.14	8	99	1.07
7	95	1.02	7	140	1.1	8	55	1.16	8	100	1.06
7	96	1.05	7	141	1.11	8	56	1.16	8	101	1.08
7	97	1.08	7	142	1.08	8	57	1.16	8	102	1.13
7	98	1.03	7	143	1.11	8	58	1.18	8	103	1.1
7	99	1.07	7	144	1.11	8	59	1.19	8	104	1.09
7	100	1.06	7	145	1.1	8	60	1.18	8	105	1.13
7	101	1.06	7	146	1.08	8	61	1.19	8	106	1.09
7	102	1.1	7	147	1.11	8	62	1.16	8	107	1.09
7	103	1.12	7	148	1.1	8	63	1.14	8	108	1.12
7	104	1.1	7	149	1.1	8	64	1.13	8	109	1.09
7	105	1.09	8	20	1.99	8	65	1.16	8	110	1.13
7	106	1.08	8	21	1.91	8	66	1.13	8	111	1.09
7	107	1.07	8	22	1.73	8	67	1.14	8	112	1.09
7	108	1.09	8	23	1.62	8	68	1.16	8	113	1.12
7	109	1.09	8	24	1.69	8	69	1.12	8	114	1.11
7	110	1.09	8	25	1.52	8	70	1.1	8	115	1.14
7	111	1.11	8	26	1.42	8	71	1.1	8	116	1.12
7	112	1.1	8	27	1.46	8	72	1.1	8	117	1.15
7	113	1.1	8	28	1.4	8	73	1.1	8	118	1.13
7	114	1.08	8	29	1.38	8	74	1.09	8	119	1.11
7	115	1.09	8	30	1.44	8	75	1.08	8	120	1.14
7	116	1.1	8	31	1.28	8	76	1.08	8	121	1.13
7	117	1.12	8	32	1.33	8	77	1.13	8	122	1.15
7	118	1.14	8	33	1.32	8	78	1.08	8	123	1.13
7	119	1.1	8	34	1.26	8	79	1.12	8	124	1.15
7	120	1.12	8	35	1.26	8	80	1.1	8	125	1.17
7	121	1.13	8	36	1.3	8	81	1.12	8	126	1.21
7	122	1.12	8	37	1.3	8	82	1.07	8	127	1.2
7	123	1.11	8	38	1.33	8	83	1.08	8	128	1.2
7	124	1.11	8	39	1.32	8	84	1.09	8	129	1.17
7	125	1.13	8	40	1.29	8	85	1.08	8	130	1.16
7	126	1.13	8	41	1.29	8	86	1.09	8	131	1.16
7	127	1.15	8	42	1.28	8	87	1.07	8	132	1.15
7	128	1.19	8	43	1.26	8	88	1.07	8	133	1.15
7	129	1.17	8	44	1.19	8	89	1.07	8	134	1.15
7	130	1.1	8	45	1.18	8	90	1.07	8	135	1.12
7	131	1.11	8	46	1.18	8	91	1.07	8	136	1.17
7	132	1.12	8	47	1.17	8	92	1.06	8	137	1.16
7	133	1.12	8	48	1.16	8	93	1.03	8	138	1.18
7	134	1.16	8	49	1.16	8	94	1.07	8	139	1.14

Tabela C.1: Íntegra da Tabela 6.2, Capítulo 6, Seção 6.2 (6/8)

Indivíduo (M, L, NMSE_f)

M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f
8	140	1.14	9	55	1.17	9	100	1.08	9	145	1.13
8	141	1.15	9	56	1.17	9	101	1.14	9	146	1.1
8	142	1.13	9	57	1.24	9	102	1.13	9	147	1.08
8	143	1.13	9	58	1.22	9	103	1.11	9	148	1.11
8	144	1.13	9	59	1.21	9	104	1.09	9	149	1.12
8	145	1.11	9	60	1.24	9	105	1.09	10	20	2.6
8	146	1.11	9	61	1.25	9	106	1.08	10	21	1.11
8	147	1.12	9	62	1.21	9	107	1.09	10	22	2.24
8	148	1.13	9	63	1.22	9	108	1.1	10	23	1.84
8	149	1.13	9	64	1.22	9	109	1.11	10	24	1.78
9	20	2.29	9	65	1.18	9	110	1.09	10	25	1.8
9	21	2.12	9	66	1.18	9	111	1.09	10	26	1.6
9	22	2	9	67	1.17	9	112	1.09	10	27	1.6
9	23	1.77	9	68	1.14	9	113	1.11	10	28	1.53
9	24	1.7	9	69	1.15	9	114	1.11	10	29	1.5
9	25	1.61	9	70	1.13	9	115	1.13	10	30	1.54
9	26	1.57	9	71	1.14	9	116	1.08	10	31	1.44
9	27	1.47	9	72	1.15	9	117	1.1	10	32	1.44
9	28	1.47	9	73	1.12	9	118	1.12	10	33	1.45
9	29	1.45	9	74	1.1	9	119	1.14	10	34	1.37
9	30	1.45	9	75	1.1	9	120	1.11	10	35	1.39
9	31	1.38	9	76	1.1	9	121	1.11	10	36	1.39
9	32	1.37	9	77	1.13	9	122	1.15	10	37	1.4
9	33	1.32	9	78	1.13	9	123	1.14	10	38	1.36
9	34	1.35	9	79	1.11	9	124	1.14	10	39	1.38
9	35	1.3	9	80	1.11	9	125	1.15	10	40	1.41
9	36	1.3	9	81	1.14	9	126	1.17	10	41	1.42
9	37	1.3	9	82	1.14	9	127	1.19	10	42	1.35
9	38	1.3	9	83	1.13	9	128	1.19	10	43	1.33
9	39	1.38	9	84	1.13	9	129	1.16	10	44	1.31
9	40	1.32	9	85	1.12	9	130	1.16	10	45	1.31
9	41	1.33	9	86	1.11	9	131	1.17	10	46	1.34
9	42	1.4	9	87	1.13	9	132	1.17	10	47	1.25
9	43	1.3	9	88	1.14	9	133	1.16	10	48	1.28
9	44	1.26	9	89	1.16	9	134	1.17	10	49	1.21
9	45	1.28	9	90	1.12	9	135	1.14	10	50	1.2
9	46	1.25	9	91	1.1	9	136	1.16	10	51	1.22
9	47	1.23	9	92	1.11	9	137	1.14	10	52	1.23
9	48	1.22	9	93	1.11	9	138	1.17	10	53	1.19
9	49	1.25	9	94	1.1	9	139	1.13	10	54	1.17
9	50	1.24	9	95	1.1	9	140	1.15	10	55	1.21
9	51	1.18	9	96	1.11	9	141	1.14	10	56	1.23
9	52	1.19	9	97	1.1	9	142	1.12	10	57	1.23
9	53	1.16	9	98	1.1	9	143	1.11	10	58	1.24
9	54	1.18	9	99	1.08	9	144	1.13	10	59	1.25

Tabela C.1: Íntegra da Tabela 6.2, Capítulo 6, Seção 6.2 (7/8)

Indivíduo (M, L, NMSE_f)											
M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f
10	60	1.29	10	105	1.16	11	20	1.12	11	65	1.22
10	61	1.22	10	106	1.12	11	21	2.73	11	66	1.24
10	62	1.25	10	107	1.11	11	22	2.57	11	67	1.18
10	63	1.2	10	108	1.14	11	23	1.98	11	68	1.17
10	64	1.21	10	109	1.15	11	24	1.85	11	69	1.17
10	65	1.22	10	110	1.13	11	25	1.81	11	70	1.16
10	66	1.21	10	111	1.17	11	26	1.7	11	71	1.15
10	67	1.21	10	112	1.14	11	27	1.68	11	72	1.19
10	68	1.18	10	113	1.18	11	28	1.59	11	73	1.17
10	69	1.19	10	114	1.16	11	29	1.63	11	74	1.15
10	70	1.19	10	115	1.16	11	30	1.6	11	75	1.13
10	71	1.16	10	116	1.19	11	31	1.48	11	76	1.16
10	72	1.17	10	117	1.16	11	32	1.52	11	77	1.15
10	73	1.15	10	118	1.17	11	33	1.46	11	78	1.11
10	74	1.14	10	119	1.15	11	34	1.39	11	79	1.13
10	75	1.15	10	120	1.17	11	35	1.38	11	80	1.15
10	76	1.15	10	121	1.14	11	36	1.37	11	81	1.15
10	77	1.12	10	122	1.13	11	37	1.41	11	82	1.1
10	78	1.14	10	123	1.19	11	38	1.42	11	83	1.12
10	79	1.13	10	124	1.15	11	39	1.45	11	84	1.14
10	80	1.13	10	125	1.18	11	40	1.51	11	85	1.15
10	81	1.14	10	126	1.25	11	41	1.42	11	86	1.13
10	82	1.14	10	127	1.25	11	42	1.41	11	87	1.14
10	83	1.18	10	128	1.19	11	43	1.37	11	88	1.13
10	84	1.11	10	129	1.19	11	44	1.34	11	89	1.1
10	85	1.11	10	130	1.19	11	45	1.29	11	90	1.12
10	86	1.14	10	131	1.2	11	46	1.3	11	91	1.15
10	87	1.14	10	132	1.21	11	47	1.34	11	92	1.11
10	88	1.13	10	133	1.21	11	48	1.32	11	93	1.14
10	89	1.14	10	134	1.24	11	49	1.21	11	94	1.12
10	90	1.12	10	135	1.2	11	50	1.23	11	95	1.13
10	91	1.11	10	136	1.19	11	51	1.21	11	96	1.13
10	92	1.11	10	137	1.2	11	52	1.23	11	97	1.12
10	93	1.09	10	138	1.19	11	53	1.17	11	98	1.14
10	94	1.11	10	139	1.17	11	54	1.22	11	99	1.16
10	95	1.09	10	140	1.15	11	55	1.23	11	100	1.16
10	96	1.12	10	141	1.14	11	56	1.22	11	101	1.18
10	97	1.11	10	142	1.16	11	57	1.24	11	102	1.15
10	98	1.09	10	143	1.17	11	58	1.24	11	103	1.18
10	99	1.15	10	144	1.17	11	59	1.23	11	104	1.15
10	100	1.14	10	145	1.18	11	60	1.27	11	105	1.21
10	101	1.18	10	146	1.18	11	61	1.23	11	106	1.18
10	102	1.11	10	147	1.17	11	62	1.24	11	107	1.19
10	103	1.13	10	148	1.14	11	63	1.23	11	108	1.19
10	104	1.11	10	149	1.15	11	64	1.21	11	109	1.17

Tabela C.1: Íntegra da Tabela 6.2, Capítulo 6, Seção 6.2 (8/8)

Indivíduo (M, L, NMSE_f)

M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f
11	110	1.15	12	25	1.97	12	70	1.19	12	115	1.19
11	111	1.17	12	26	1.75	12	71	1.17	12	116	1.2
11	112	1.19	12	27	1.62	12	72	1.2	12	117	1.22
11	113	1.18	12	28	1.75	12	73	1.17	12	118	1.22
11	114	1.15	12	29	1.7	12	74	1.17	12	119	1.19
11	115	1.18	12	30	1.77	12	75	1.17	12	120	1.22
11	116	1.18	12	31	1.66	12	76	1.17	12	121	1.25
11	117	1.19	12	32	1.59	12	77	1.19	12	122	1.28
11	118	1.18	12	33	1.51	12	78	1.17	12	123	1.28
11	119	1.19	12	34	1.5	12	79	1.18	12	124	1.27
11	120	1.18	12	35	1.48	12	80	1.14	12	125	1.23
11	121	1.21	12	36	1.49	12	81	1.16	12	126	1.26
11	122	1.25	12	37	1.49	12	82	1.17	12	127	1.27
11	123	1.24	12	38	1.45	12	83	1.16	12	128	1.25
11	124	1.25	12	39	1.59	12	84	1.14	12	129	1.23
11	125	1.27	12	40	1.46	12	85	1.14	12	130	1.19
11	126	1.26	12	41	1.49	12	86	1.16	12	131	1.25
11	127	1.24	12	42	1.48	12	87	1.16	12	132	1.24
11	128	1.21	12	43	1.38	12	88	1.17	12	133	1.25
11	129	1.24	12	44	1.45	12	89	1.14	12	134	1.29
11	130	1.19	12	45	1.39	12	90	1.14	12	135	1.23
11	131	1.22	12	46	1.41	12	91	1.11	12	136	1.23
11	132	1.25	12	47	1.32	12	92	1.13	12	137	1.23
11	133	1.24	12	48	1.28	12	93	1.16	12	138	1.26
11	134	1.26	12	49	1.26	12	94	1.15	12	139	1.23
11	135	1.22	12	50	1.22	12	95	1.13	12	140	1.17
11	136	1.21	12	51	1.3	12	96	1.16	12	141	1.18
11	137	1.21	12	52	1.22	12	97	1.15	12	142	1.14
11	138	1.18	12	53	1.2	12	98	1.18	12	143	1.18
11	139	1.22	12	54	1.19	12	99	1.15	12	144	1.21
11	140	1.21	12	55	1.27	12	100	1.14	12	145	1.18
11	141	1.21	12	56	1.25	12	101	1.16	12	146	1.2
11	142	1.17	12	57	1.25	12	102	1.14	12	147	1.14
11	143	1.2	12	58	1.25	12	103	1.19	12	148	1.19
11	144	1.17	12	59	1.25	12	104	1.18	12	149	1.17
11	145	1.19	12	60	1.32	12	105	1.17			
11	146	1.17	12	61	1.26	12	106	1.18			
11	147	1.16	12	62	1.28	12	107	1.17			
11	148	1.13	12	63	1.23	12	108	1.18			
11	149	1.18	12	64	1.25	12	109	1.17			
12	20	3.65	12	65	1.21	12	110	1.13			
12	21	1.63	12	66	1.28	12	111	1.18			
12	22	1.08	12	67	1.25	12	112	1.16			
12	23	2.14	12	68	1.21	12	113	1.18			
12	24	1.91	12	69	1.2	12	114	1.16			

Tabela C.2: Íntegra da Tabela 6.3, Capítulo 6, Seção 6.2 (1/8)

Indivíduo (M, L, NMSE_f)											
M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f
2	89	0.99	2	74	1.02	3	105	1.03	4	105	1.04
2	88	0.991	3	87	1.02	2	140	1.03	3	71	1.04
4	104	0.997	3	90	1.02	3	78	1.03	2	139	1.04
2	82	0.998	3	89	1.02	2	81	1.03	2	138	1.04
4	99	0.999	4	92	1.02	2	76	1.03	4	70	1.05
3	84	1	2	107	1.02	2	101	1.03	5	103	1.05
2	90	1	4	89	1.02	2	78	1.03	6	96	1.05
2	84	1	4	94	1.02	2	102	1.03	5	86	1.05
2	94	1	4	93	1.02	8	93	1.03	4	73	1.05
3	97	1.01	3	92	1.02	2	30	1.03	5	110	1.05
7	94	1.01	4	97	1.02	2	97	1.03	2	61	1.05
2	105	1.01	4	103	1.02	2	96	1.03	3	65	1.05
2	100	1.01	5	96	1.02	2	83	1.03	6	98	1.05
2	104	1.01	5	97	1.02	2	26	1.04	5	101	1.05
3	101	1.01	2	103	1.02	6	97	1.04	3	69	1.05
3	91	1.01	2	93	1.02	2	147	1.04	3	77	1.05
3	88	1.01	3	103	1.02	2	146	1.04	2	79	1.05
3	93	1.01	3	102	1.02	4	79	1.04	5	94	1.05
3	94	1.01	3	96	1.02	5	95	1.04	5	93	1.05
2	71	1.01	3	95	1.02	6	99	1.04	2	112	1.05
3	99	1.01	3	98	1.02	5	84	1.04	6	75	1.05
5	89	1.01	3	104	1.02	4	80	1.04	4	75	1.05
2	106	1.01	7	95	1.02	7	76	1.04	2	144	1.05
2	87	1.01	3	100	1.02	4	83	1.04	4	106	1.05
2	86	1.01	2	149	1.03	5	90	1.04	2	27	1.05
4	98	1.01	2	142	1.03	2	141	1.04	2	148	1.05
4	85	1.01	3	70	1.03	4	87	1.04	3	141	1.05
4	91	1.01	7	98	1.03	4	88	1.04	5	71	1.05
4	100	1.01	7	93	1.03	5	92	1.04	2	137	1.05
4	101	1.01	5	72	1.03	2	145	1.04	4	107	1.05
2	91	1.01	3	74	1.03	2	143	1.04	5	70	1.05
4	102	1.01	5	87	1.03	2	111	1.04	4	109	1.05
4	86	1.01	5	88	1.03	2	77	1.04	4	111	1.05
4	96	1.01	4	90	1.03	3	75	1.04	6	90	1.05
3	86	1.02	4	78	1.03	6	93	1.04	4	114	1.05
2	70	1.02	5	98	1.03	6	94	1.04	4	116	1.05
2	92	1.02	4	95	1.03	2	75	1.04	2	28	1.05
2	95	1.02	5	91	1.03	3	73	1.04	7	96	1.05
2	72	1.02	5	99	1.03	2	73	1.04	2	125	1.05
3	80	1.02	3	79	1.03	2	69	1.04	2	65	1.05
2	80	1.02	3	83	1.03	2	115	1.04	4	82	1.05
2	85	1.02	3	81	1.03	3	106	1.04	3	112	1.05
3	82	1.02	2	113	1.03	3	76	1.04	3	140	1.05
2	99	1.02	3	85	1.03	4	108	1.04	7	80	1.05
2	98	1.02	6	95	1.03	3	72	1.04	7	89	1.05

Tabela C.2: Íntegra da Tabela 6.3, Capítulo 6, Seção 6.2 (2/8)

Indivíduo (M, L, NMSE_f)

M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f
2	126	1.05	3	124	1.06	6	113	1.07	3	109	1.07
2	118	1.05	3	116	1.06	7	99	1.07	3	149	1.07
2	31	1.05	7	74	1.06	8	94	1.07	6	80	1.07
7	91	1.05	2	57	1.06	2	29	1.07	6	102	1.07
2	34	1.05	3	31	1.06	8	98	1.07	6	78	1.07
8	95	1.05	2	130	1.06	2	109	1.07	3	66	1.07
3	125	1.06	2	117	1.06	2	36	1.07	4	139	1.07
6	83	1.06	7	87	1.06	8	99	1.07	3	68	1.07
3	127	1.06	2	62	1.06	2	134	1.07	3	67	1.07
3	142	1.06	2	127	1.06	3	30	1.07	5	65	1.07
3	148	1.06	7	92	1.06	2	129	1.07	5	107	1.07
3	145	1.06	2	120	1.06	2	37	1.07	3	63	1.07
3	143	1.06	2	124	1.06	4	119	1.07	3	61	1.07
3	144	1.06	2	63	1.06	4	76	1.07	6	101	1.07
3	146	1.06	7	71	1.06	5	77	1.07	5	106	1.07
3	138	1.06	2	121	1.06	5	75	1.07	2	135	1.07
6	91	1.06	3	64	1.06	8	91	1.07	2	67	1.07
3	147	1.06	3	107	1.06	5	76	1.07	5	116	1.07
3	137	1.06	3	108	1.06	4	113	1.07	2	66	1.07
3	139	1.06	6	100	1.06	7	78	1.07	5	109	1.07
4	72	1.06	3	114	1.06	2	128	1.07	5	114	1.07
4	29	1.06	3	115	1.06	8	88	1.07	6	73	1.07
5	74	1.06	7	100	1.06	2	46	1.07	6	77	1.07
4	110	1.06	3	111	1.06	8	87	1.07	2	68	1.07
4	112	1.06	3	110	1.06	2	55	1.07	2	64	1.07
4	115	1.06	2	136	1.06	4	143	1.07	4	148	1.08
5	73	1.06	7	101	1.06	4	144	1.07	5	69	1.08
2	35	1.06	4	68	1.07	2	54	1.07	4	149	1.08
4	147	1.06	8	89	1.07	2	56	1.07	7	106	1.08
4	145	1.06	4	65	1.07	3	136	1.07	5	63	1.08
4	142	1.06	7	107	1.07	4	146	1.07	5	82	1.08
8	100	1.06	3	52	1.07	2	119	1.07	8	101	1.08
4	84	1.06	4	67	1.07	2	60	1.07	7	146	1.08
8	92	1.06	8	90	1.07	2	116	1.07	8	85	1.08
6	70	1.06	5	80	1.07	6	88	1.07	8	83	1.08
2	45	1.06	3	54	1.07	4	141	1.07	5	108	1.08
5	104	1.06	3	118	1.07	6	81	1.07	6	72	1.08
5	102	1.06	3	49	1.07	8	82	1.07	6	86	1.08
5	100	1.06	5	83	1.07	3	120	1.07	9	100	1.08
5	78	1.06	2	122	1.07	3	121	1.07	6	106	1.08
4	81	1.06	5	79	1.07	3	122	1.07	9	99	1.08
4	69	1.06	4	57	1.07	3	113	1.07	8	75	1.08
4	71	1.06	5	85	1.07	6	89	1.07	8	76	1.08
4	77	1.06	3	50	1.07	3	126	1.07	8	78	1.08
2	114	1.06	3	117	1.07	3	129	1.07	6	87	1.08

Tabela C.2: Íntegra da Tabela 6.3, Capítulo 6, Seção 6.2 (3/8)

Indivíduo (M, L, NMSE_f)

M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f
6	103	1.08	2	47	1.08	4	54	1.09	3	131	1.1
5	81	1.08	2	48	1.08	2	53	1.09	3	133	1.1
7	114	1.08	2	59	1.08	7	109	1.09	9	95	1.1
7	81	1.08	3	128	1.08	7	108	1.09	9	76	1.1
7	142	1.08	4	137	1.08	5	117	1.09	6	69	1.1
7	97	1.08	2	58	1.08	5	113	1.09	9	75	1.1
7	90	1.08	3	123	1.08	10	93	1.09	9	94	1.1
7	88	1.08	2	32	1.08	3	134	1.09	3	130	1.1
2	21	1.08	4	122	1.08	7	66	1.09	6	66	1.1
7	79	1.08	3	55	1.08	3	46	1.09	4	30	1.1
6	110	1.08	2	123	1.08	9	112	1.09	4	138	1.1
7	75	1.08	4	28	1.08	6	74	1.09	6	107	1.1
7	72	1.08	2	110	1.08	4	126	1.09	3	43	1.1
9	106	1.08	2	52	1.09	5	62	1.09	3	42	1.1
9	116	1.08	8	86	1.09	9	110	1.09	7	69	1.1
9	147	1.08	7	82	1.09	5	64	1.09	7	70	1.1
4	140	1.08	8	84	1.09	8	104	1.09	7	113	1.1
7	84	1.08	7	83	1.09	9	111	1.09	7	119	1.1
3	47	1.08	2	108	1.09	5	56	1.09	7	116	1.1
4	61	1.08	7	86	1.09	9	104	1.09	7	112	1.1
4	64	1.08	2	43	1.09	8	96	1.09	9	108	1.1
4	56	1.08	6	92	1.09	6	79	1.09	2	133	1.1
4	58	1.08	6	109	1.09	8	97	1.09	2	132	1.1
4	74	1.08	6	64	1.09	6	76	1.09	2	131	1.1
4	52	1.08	8	111	1.09	9	105	1.09	7	104	1.1
4	66	1.08	2	24	1.09	9	107	1.09	7	85	1.1
4	120	1.08	6	146	1.09	6	82	1.09	3	29	1.1
12	22	1.08	6	147	1.09	3	44	1.09	3	45	1.1
2	38	1.08	8	112	1.09	4	59	1.09	7	130	1.1
2	33	1.08	6	145	1.09	6	84	1.09	9	117	1.1
4	125	1.08	7	110	1.09	3	135	1.09	6	108	1.1
2	39	1.08	5	111	1.09	7	73	1.09	9	97	1.1
3	53	1.08	3	28	1.09	6	71	1.09	9	98	1.1
4	124	1.08	4	136	1.09	8	109	1.09	6	120	1.1
4	48	1.08	3	56	1.09	7	115	1.09	6	122	1.1
2	44	1.08	5	124	1.09	4	118	1.09	6	126	1.1
2	25	1.08	2	50	1.09	4	63	1.09	6	119	1.1
3	58	1.08	6	114	1.09	4	117	1.09	6	125	1.1
3	59	1.08	7	105	1.09	8	107	1.09	7	149	1.1
3	51	1.08	10	95	1.09	8	106	1.09	7	136	1.1
3	57	1.08	5	123	1.09	4	121	1.09	7	140	1.1
3	62	1.08	8	74	1.09	2	40	1.09	9	146	1.1
2	51	1.08	10	98	1.09	4	123	1.09	6	143	1.1
3	60	1.08	5	112	1.09	4	60	1.09	6	148	1.1
3	119	1.08	4	32	1.09	9	91	1.1	7	148	1.1

Tabela C.2: Íntegra da Tabela 6.3, Capítulo 6, Seção 6.2 (4/8)

Indivíduo (M, L, NMSE_f)

M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f
7	145	1.1	9	143	1.11	4	27	1.11	6	115	1.12
9	74	1.1	7	141	1.11	6	67	1.11	9	118	1.12
7	102	1.1	8	114	1.11	10	102	1.11	4	43	1.12
8	80	1.1	7	63	1.11	8	146	1.11	5	140	1.12
5	67	1.1	7	77	1.11	12	91	1.11	5	142	1.12
8	70	1.1	11	92	1.11	5	144	1.11	6	140	1.12
5	59	1.1	7	111	1.11	10	104	1.11	4	40	1.12
5	143	1.1	4	128	1.11	10	97	1.11	5	145	1.12
5	66	1.1	3	32	1.11	5	137	1.11	4	44	1.12
5	146	1.1	5	54	1.11	6	105	1.11	10	77	1.12
5	149	1.1	4	129	1.11	6	112	1.11	4	34	1.12
5	147	1.1	8	119	1.11	9	79	1.11	5	138	1.12
5	49	1.1	4	134	1.11	10	91	1.11	8	69	1.12
2	42	1.1	4	135	1.11	7	120	1.12	6	111	1.12
2	41	1.1	5	52	1.11	10	96	1.12	8	79	1.12
4	53	1.1	7	123	1.11	9	73	1.12	3	48	1.12
4	49	1.1	9	109	1.11	5	55	1.12	8	81	1.12
4	50	1.1	10	21	1.11	9	90	1.12	6	104	1.12
5	105	1.1	5	68	1.11	10	90	1.12	5	118	1.12
4	51	1.1	9	148	1.11	5	57	1.12	11	20	1.12
11	82	1.1	4	55	1.11	7	56	1.12	3	132	1.12
2	23	1.1	6	118	1.11	7	122	1.12	6	137	1.12
4	127	1.1	7	131	1.11	4	133	1.12	7	135	1.12
4	31	1.1	9	121	1.11	6	68	1.12	7	137	1.12
5	122	1.1	9	103	1.11	6	41	1.12	7	138	1.12
5	120	1.1	11	78	1.11	11	97	1.12	4	62	1.12
8	72	1.1	9	120	1.11	3	27	1.12	9	142	1.12
5	125	1.1	7	124	1.11	3	26	1.12	8	147	1.12
8	73	1.1	9	114	1.11	8	113	1.12	7	133	1.12
8	71	1.1	9	113	1.11	5	61	1.12	9	149	1.12
5	115	1.1	10	94	1.11	3	36	1.12	7	132	1.12
8	103	1.1	9	80	1.11	11	94	1.12	4	26	1.12
11	89	1.1	6	117	1.11	8	116	1.12	9	102	1.13
2	22	1.1	6	116	1.11	7	103	1.12	7	126	1.13
6	141	1.11	10	85	1.11	8	135	1.12	8	118	1.13
6	149	1.11	9	86	1.11	9	85	1.12	7	121	1.13
6	121	1.11	5	148	1.11	8	108	1.12	8	149	1.13
5	121	1.11	9	96	1.11	6	62	1.12	8	123	1.13
6	144	1.11	9	93	1.11	11	83	1.12	9	77	1.13
7	144	1.11	6	54	1.11	7	117	1.12	8	121	1.13
4	33	1.11	8	145	1.11	3	33	1.12	8	77	1.13
10	92	1.11	10	84	1.11	6	50	1.12	8	148	1.13
7	143	1.11	9	92	1.11	7	64	1.12	9	70	1.13
6	142	1.11	5	141	1.11	11	90	1.12	8	143	1.13
7	147	1.11	10	107	1.11	10	106	1.12	2	49	1.13

Tabela C.2: Íntegra da Tabela 6.3, Capítulo 6, Seção 6.2 (5/8)

Indivíduo (M, L, NMSE_f)

M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f
8	105	1.13	7	67	1.13	10	100	1.14	11	102	1.15
8	64	1.13	3	35	1.13	9	68	1.14	11	91	1.15
8	66	1.13	4	130	1.13	6	139	1.14	8	134	1.15
8	102	1.13	10	88	1.13	10	78	1.14	11	114	1.15
9	84	1.13	10	110	1.13	6	134	1.14	12	99	1.15
7	139	1.13	11	96	1.13	6	133	1.14	11	104	1.15
7	125	1.13	4	132	1.13	6	127	1.14	11	110	1.15
9	78	1.13	12	110	1.13	10	81	1.14	9	69	1.15
9	87	1.13	11	86	1.13	12	142	1.14	9	125	1.15
8	144	1.13	7	47	1.13	6	123	1.14	8	141	1.15
8	110	1.13	11	88	1.13	7	118	1.14	10	140	1.15
8	142	1.13	9	115	1.13	6	138	1.14	10	124	1.15
9	83	1.13	10	122	1.13	10	74	1.14	10	99	1.15
6	131	1.13	8	54	1.14	7	61	1.14	10	119	1.15
4	46	1.13	10	141	1.14	9	123	1.14	11	71	1.15
9	145	1.13	10	108	1.14	9	137	1.14	10	109	1.15
10	103	1.13	10	148	1.14	9	119	1.14	9	72	1.15
11	148	1.13	8	140	1.14	9	135	1.14	10	149	1.15
5	60	1.13	10	112	1.14	7	53	1.14	12	94	1.15
11	79	1.13	10	121	1.14	9	101	1.14	11	74	1.15
9	144	1.13	8	63	1.14	7	55	1.14	11	81	1.15
5	139	1.13	8	139	1.14	9	141	1.14	10	73	1.15
5	51	1.13	8	67	1.14	7	54	1.14	11	80	1.15
5	50	1.13	12	147	1.14	3	40	1.14	9	140	1.15
9	139	1.13	5	127	1.14	3	38	1.14	10	75	1.15
6	56	1.13	5	31	1.14	3	37	1.14	8	132	1.15
6	60	1.13	5	36	1.14	12	90	1.14	11	77	1.15
12	92	1.13	5	33	1.14	10	89	1.14	10	76	1.15
5	126	1.13	11	98	1.14	3	25	1.14	12	97	1.15
12	95	1.13	11	93	1.14	10	87	1.14	8	133	1.15
6	85	1.13	11	84	1.14	6	46	1.14	2	20	1.15
5	132	1.13	8	115	1.14	12	100	1.14	7	49	1.15
5	119	1.13	8	120	1.14	6	53	1.14	7	65	1.15
6	124	1.13	11	87	1.14	12	102	1.14	5	58	1.15
6	61	1.13	5	41	1.14	9	71	1.14	5	128	1.15
5	135	1.13	12	89	1.14	6	52	1.14	7	127	1.15
10	80	1.13	4	38	1.14	9	124	1.14	5	44	1.15
10	79	1.13	12	85	1.14	9	88	1.14	8	52	1.15
5	47	1.13	4	37	1.14	6	57	1.14	5	133	1.15
11	75	1.13	12	80	1.14	10	86	1.14	5	53	1.15
6	55	1.13	12	84	1.14	9	82	1.14	7	60	1.15
7	68	1.13	5	43	1.14	10	82	1.14	7	46	1.15
3	34	1.13	4	42	1.14	9	81	1.14	7	50	1.15
11	95	1.13	5	48	1.14	11	85	1.15	7	48	1.15
3	41	1.13	8	53	1.14	9	122	1.15	3	23	1.15

Tabela C.2: Íntegra da Tabela 6.3, Capítulo 6, Seção 6.2 (6/8)

Indivíduo (M, L, NMSE_f)

M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f
7	57	1.15	11	76	1.16	6	129	1.17	10	68	1.18
5	29	1.15	5	32	1.16	10	143	1.17	11	67	1.18
3	39	1.15	4	39	1.16	10	147	1.17	6	128	1.18
7	58	1.15	12	83	1.16	10	54	1.17	8	45	1.18
5	134	1.15	10	105	1.16	11	73	1.17	8	58	1.18
5	136	1.15	12	112	1.16	11	68	1.17	12	79	1.18
4	131	1.15	9	53	1.16	11	69	1.17	9	54	1.18
6	51	1.15	8	57	1.16	8	125	1.17	8	60	1.18
6	47	1.15	12	101	1.16	11	53	1.17	6	132	1.18
8	124	1.15	8	55	1.16	8	136	1.17	11	116	1.18
8	122	1.15	12	86	1.16	6	48	1.17	4	35	1.18
8	117	1.15	12	114	1.16	12	105	1.17	11	115	1.18
4	47	1.15	8	65	1.16	4	36	1.17	12	106	1.18
6	49	1.15	12	93	1.16	7	51	1.17	10	146	1.18
6	58	1.15	12	96	1.16	12	73	1.17	12	98	1.18
6	135	1.15	3	24	1.16	12	74	1.17	10	145	1.18
6	65	1.15	8	62	1.16	11	146	1.17	12	141	1.18
6	63	1.15	8	56	1.16	12	71	1.17	5	46	1.18
4	45	1.15	8	68	1.16	12	76	1.17	10	125	1.18
8	48	1.16	8	130	1.16	11	144	1.17	12	108	1.18
12	81	1.16	8	131	1.16	12	75	1.17	10	113	1.18
8	49	1.16	8	137	1.16	8	47	1.17	11	103	1.18
9	89	1.16	12	87	1.16	12	78	1.17	11	106	1.18
4	41	1.16	12	107	1.17	7	129	1.17	4	25	1.18
7	134	1.16	11	109	1.17	12	88	1.17	11	138	1.18
7	62	1.16	5	45	1.17	12	82	1.17	5	27	1.18
11	147	1.16	5	35	1.17	8	50	1.17	10	101	1.18
5	42	1.16	6	42	1.17	9	67	1.17	12	111	1.18
7	45	1.16	9	126	1.17	9	55	1.17	11	113	1.18
6	59	1.16	9	138	1.17	9	56	1.17	12	113	1.18
10	71	1.16	11	111	1.17	11	142	1.17	11	101	1.18
6	136	1.16	10	111	1.17	12	149	1.17	5	28	1.18
10	115	1.16	9	134	1.17	4	24	1.17	6	43	1.18
10	117	1.16	12	140	1.17	7	52	1.17	5	40	1.18
11	70	1.16	9	132	1.17	7	59	1.17	5	30	1.18
10	114	1.16	9	131	1.17	4	23	1.17	9	65	1.18
6	36	1.16	8	129	1.17	8	138	1.18	5	34	1.18
10	142	1.16	12	109	1.17	8	46	1.18	12	104	1.18
5	131	1.16	5	130	1.17	11	118	1.18	9	66	1.18
9	129	1.16	5	129	1.17	11	120	1.18	8	59	1.19
9	136	1.16	10	72	1.17	12	145	1.18	11	117	1.19
9	130	1.16	10	139	1.17	10	83	1.18	10	123	1.19
9	133	1.16	10	118	1.17	11	149	1.18	6	44	1.19
11	99	1.16	10	120	1.17	9	51	1.18	10	70	1.19
11	100	1.16	10	144	1.17	12	143	1.18	11	119	1.19

Tabela C.2: Íntegra da Tabela 6.3, Capítulo 6, Seção 6.2 (7/8)

Indivíduo (M, L, NMSE_f)										
M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f
8	61	1.19	8	128	1.2	10	61	1.22	10	134	1.24
8	51	1.19	7	31	1.2	11	56	1.22	11	57	1.24
8	44	1.19	10	50	1.2	10	65	1.22	11	58	1.24
12	148	1.19	11	51	1.21	12	120	1.22	12	64	1.25
6	130	1.19	11	105	1.21	10	51	1.22	9	61	1.25
10	116	1.19	11	64	1.21	6	40	1.23	12	59	1.25
12	103	1.19	10	49	1.21	11	55	1.23	12	58	1.25
11	130	1.19	11	49	1.21	5	26	1.23	12	67	1.25
7	128	1.19	7	39	1.21	6	45	1.23	12	56	1.25
10	69	1.19	10	55	1.21	7	41	1.23	12	57	1.25
10	128	1.19	10	133	1.21	7	44	1.23	11	122	1.25
12	119	1.19	6	32	1.21	11	52	1.23	7	42	1.25
11	72	1.19	11	128	1.21	11	50	1.23	11	132	1.25
9	52	1.19	11	121	1.21	11	59	1.23	11	124	1.25
12	130	1.19	12	68	1.21	7	37	1.23	9	49	1.25
9	128	1.19	12	144	1.21	12	63	1.23	6	39	1.25
11	145	1.19	9	59	1.21	11	63	1.23	7	35	1.25
12	54	1.19	9	62	1.21	12	129	1.23	7	36	1.25
12	115	1.19	12	65	1.21	10	57	1.23	9	46	1.25
11	107	1.19	11	136	1.21	12	125	1.23	10	47	1.25
11	108	1.19	11	141	1.21	12	136	1.23	12	128	1.25
11	112	1.19	11	140	1.21	12	135	1.23	10	127	1.25
10	129	1.19	11	137	1.21	10	52	1.23	3	20	1.25
6	33	1.19	10	132	1.21	12	137	1.23	10	62	1.25
10	53	1.19	8	126	1.21	12	139	1.23	10	59	1.25
10	136	1.19	10	64	1.21	10	56	1.23	12	133	1.25
12	77	1.19	10	66	1.21	11	61	1.23	12	131	1.25
9	127	1.19	6	35	1.21	4	21	1.23	3	22	1.25
10	130	1.19	10	67	1.21	9	47	1.23	12	121	1.25
12	70	1.19	6	31	1.21	9	50	1.24	10	126	1.25
10	138	1.19	11	65	1.22	9	60	1.24	12	61	1.26
10	135	1.2	11	131	1.22	9	57	1.24	8	43	1.26
12	69	1.2	6	34	1.22	11	66	1.24	7	38	1.26
10	131	1.2	9	48	1.22	11	129	1.24	12	138	1.26
11	143	1.2	9	58	1.22	11	127	1.24	3	21	1.26
5	39	1.2	9	64	1.22	12	132	1.24	6	38	1.26
8	127	1.2	9	63	1.22	11	133	1.24	11	134	1.26
5	37	1.2	12	118	1.22	5	25	1.24	8	35	1.26
5	38	1.2	11	135	1.22	7	34	1.24	12	126	1.26
12	53	1.2	12	52	1.22	11	123	1.24	12	49	1.26
12	116	1.2	12	50	1.22	4	22	1.24	8	34	1.26
12	72	1.2	7	43	1.22	7	40	1.24	9	44	1.26
12	146	1.2	12	117	1.22	6	37	1.24	6	29	1.26
10	137	1.2	11	139	1.22	11	62	1.24	11	126	1.26
10	63	1.2	11	54	1.22	10	58	1.24	7	33	1.26

Tabela C.2: Íntegra da Tabela 6.3, Capítulo 6, Seção 6.2 (8/8)

Indivíduo (M, L, NMSE_f)

M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f	M	L	NMSE_f
12	127	1.27	8	33	1.32	10	31	1.44	12	21	1.63
12	55	1.27	9	33	1.32	5	20	1.44	11	29	1.63
11	125	1.27	12	60	1.32	10	32	1.44	12	31	1.66
7	28	1.27	10	43	1.33	8	30	1.44	11	27	1.68
6	30	1.27	9	41	1.33	12	38	1.45	8	24	1.69
12	124	1.27	8	32	1.33	11	39	1.45	11	26	1.7
11	60	1.27	8	38	1.33	9	30	1.45	12	29	1.7
6	27	1.27	6	26	1.34	12	44	1.45	9	24	1.7
7	32	1.27	10	46	1.34	10	33	1.45	7	21	1.73
10	48	1.28	11	47	1.34	9	29	1.45	8	22	1.73
12	123	1.28	11	44	1.34	12	40	1.46	7	22	1.73
12	122	1.28	10	42	1.35	8	27	1.46	12	26	1.75
8	42	1.28	9	34	1.35	11	33	1.46	12	28	1.75
9	45	1.28	10	38	1.36	9	28	1.47	12	30	1.77
6	28	1.28	11	43	1.37	9	27	1.47	9	23	1.77
12	62	1.28	10	34	1.37	12	35	1.48	10	24	1.78
12	48	1.28	9	32	1.37	11	31	1.48	10	25	1.8
8	31	1.28	11	36	1.37	12	42	1.48	11	25	1.81
12	66	1.28	12	43	1.38	6	23	1.48	7	20	1.82
4	20	1.28	10	39	1.38	12	41	1.49	10	23	1.84
8	41	1.29	9	31	1.38	12	37	1.49	11	24	1.85
7	30	1.29	8	29	1.38	12	36	1.49	12	24	1.91
12	134	1.29	11	35	1.38	12	34	1.5	8	21	1.91
8	40	1.29	9	39	1.38	7	23	1.5	12	25	1.97
5	23	1.29	5	21	1.39	10	29	1.5	11	23	1.98
11	45	1.29	12	45	1.39	12	33	1.51	8	20	1.99
10	60	1.29	11	34	1.39	11	40	1.51	9	22	2
11	46	1.3	10	36	1.39	11	32	1.52	9	21	2.12
9	43	1.3	10	35	1.39	8	25	1.52	12	23	2.14
9	38	1.3	10	37	1.4	10	28	1.53	10	22	2.24
9	36	1.3	8	28	1.4	10	30	1.54	9	20	2.29
7	29	1.3	9	42	1.4	6	21	1.55	11	22	2.57
12	51	1.3	6	25	1.4	7	24	1.55	10	20	2.6
8	36	1.3	6	24	1.4	6	20	1.56	11	21	2.73
8	37	1.3	10	40	1.41	9	26	1.57	12	20	3.65
9	37	1.3	11	37	1.41	12	32	1.59			
7	27	1.3	12	46	1.41	12	39	1.59			
9	35	1.3	11	42	1.41	11	28	1.59			
10	45	1.31	8	26	1.42	10	27	1.6			
5	24	1.31	11	38	1.42	10	26	1.6			
10	44	1.31	10	41	1.42	11	30	1.6			
9	40	1.32	11	41	1.42	6	22	1.61			
8	39	1.32	5	22	1.42	9	25	1.61			
12	47	1.32	7	26	1.43	8	23	1.62			
11	48	1.32	7	25	1.43	12	27	1.62			