

UNIVERSIDADE FEDERAL DE SANTA CATARINA
Programa de Pós-Graduação em Engenharia de Produção
Área de concentração : Mídia e Conhecimento

APLICAÇÃO DE REDES NEURAIS PARA PROGNÓSTICO
COM BASE EM SÉRIES TEMPORAIS

Claudio Martin

Dissertação apresentada ao Curso de Pós-Graduação em Engenharia de Produção, Área de Concentração : Mídia e Conhecimento, da Universidade Federal de Santa Catarina – UFSC, como requisito parcial para obtenção do título de Mestre em Engenharia de Produção.

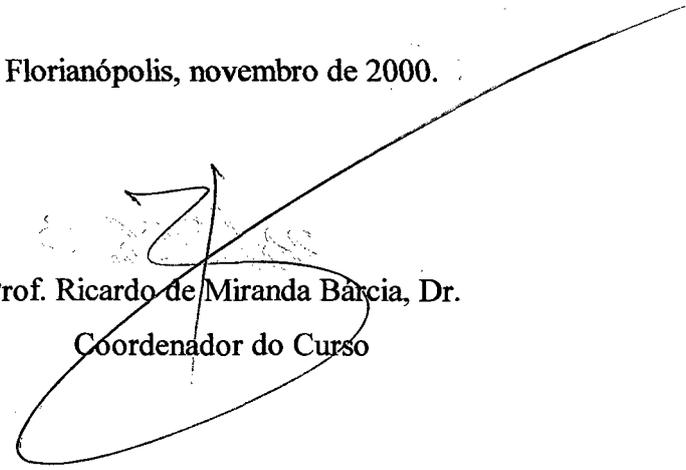
Florianópolis
Novembro-2000

Claudio Martin

**APLICAÇÃO DE REDES NEURAIIS PARA PROGNÓSTICO
COM BASE EM SÉRIES TEMPORAIS**

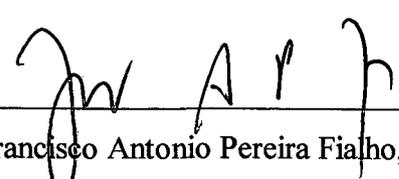
Esta Dissertação foi julgada adequada para a obtenção do título de **Mestre em Engenharia de Produção (Área de Concentração:Mídia e Conhecimento)** e aprovada em sua forma final pelo **Programa de Pós-Graduação em Engenharia de Produção** da Universidade Federal de Santa Catarina.

Florianópolis, novembro de 2000.



Prof. Ricardo de Miranda Bârcia, Dr.
Coordenador do Curso

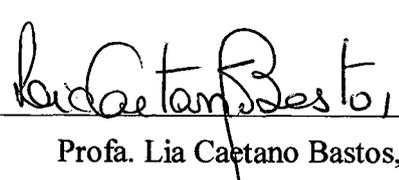
BANCA EXAMINADORA



Prof. Francisco Antonio Pereira Fialho, Dr.Eng.
Orientador



Prof. Luís Gómez, Dr.



Profa. Lia Caetano Bastos, Dra.

Para AMANDA, companheira de uma vida inteira, que com sua maneira incansável de conduzir os afazeres, é fonte inesgotável de inspiração e motivação.

Para CHARLES, FERNANDO e ROBERTO que compartilham conosco a juventude, fornecendo a energia e vitalidade necessárias, para sempre dar um passo adiante.

AGRADECIMENTOS

Ao Prof. Dr. **FRANCISCO ANTONIO PEREIRA FIALHO**, com o qual tive o privilégio de conviver, por duas vezes, nas disciplinas Ergonomia Cognitiva e Engenharia do Conhecimento, cujos conhecimentos ministrados, na ocasião, permitiram-me vislumbrar um mundo totalmente novo e desconhecido, conduzindo-me inclusive para a escolha deste tema fascinante, que é o da Inteligência Artificial. Na fase de desenvolvimento da dissertação, o Prof. Fialho foi o grande incentivador e motivador, fazendo-se presente em todas as dificuldades, permitindo, com o seu aconselhamento, a superação dos momentos de incerteza e desânimo.

Aos professores do Programa de Pós-Graduação em Engenharia de Produção/Laboratório de Ensino a Distância, da Universidade Federal de Santa Catarina (UFSC), que ministraram as disciplinas da fase acadêmica, aqui nominados, **CARLOS RAUL BORENSTEIN, FERNANDO O. GAUTHIER, MÁRCIO VIEIRA DE SOUZA, ROSÂNGELA RODRIGUES** e **SILVANA BERNARDES ROSA**, que compartilharam conosco os seus conhecimentos, e contribuíram para o nosso crescimento intelectual.

Aos colegas do Curso de Mestrado em Engenharia de Produção do CEFET-PR – interlocutores indispensáveis nas situações de ordem intelectual, e principalmente nas questões de ordem afetiva – com os quais tive a oportunidade de conviver, conhecendo-os melhor a partir desta experiência.

SUMÁRIO

LISTA DE FIGURAS	vii
RESUMO	x
ABSTRACT	xi
CAPÍTULO I. Introdução	1
1.1 Justificativa	2
1.2 Estabelecimento do problema	3
1.3 Subordinação do tema a outras áreas do conhecimento científico	4
1.4 Definição das variáveis a serem trabalhadas	5
1.5 Limitações do tema	6
1.6 Motivação pessoal para o tema	6
1.7 Objetivos gerais e específicos	7
1.8 Hipóteses gerais e específicas	8
1.9 Descrição dos capítulos	8
1.10 Conclusão – A importância do tema para a logística	10
CAPÍTULO II. Séries Temporais	12
2.1 Introdução	12
2.2 Utilização de regressão para o prognóstico de séries temporais	14
2.2.1 Regressão – Modelagem constante	17
2.2.2 Regressão – Modelagem linear.....	19
2.2.3 Regressão – Modelagem quadrática	20
2.2.4 Regressão – Modelagem exponencial.....	22
2.3 Análise de séries temporais utilizando a técnica de ajuste exponencial	24
2.3.1 Ajuste exponencial simples	26
2.3.2 Ajuste exponencial duplo	28
2.3.3 Ajuste exponencial triplo	33
2.4 Prognóstico com sazonalidade	36
2.5 Análise de erro na determinação do prognóstico	37
2.6 A técnica de Box-Jenkins	39
2.6.1 Séries temporais multivariáveis pela técnica Box-Jenkins.....	44

CAPÍTULO III. REDES NEURAIIS	47
3.1 Introdução	47
3.2 Aplicações das Redes Neurais	50
3.2.1 Redes neurais utilizadas para classificação	50
3.2.2 Redes neurais utilizadas para auto-associação	52
3.2.3 Redes neurais utilizadas para predições de séries temporais	52
3.2.4 Redes neurais utilizadas para aproximações de funções	54
3.3 Redes neurais de múltiplas camadas sem retroalimentação	56
3.4 Redes de função de base radial	60
3.5 Redes Neurais e os Métodos Tradicionais	62
CAPÍTULO IV. APLICAÇÃO DE REDES NEURAIIS EM SÉRIES TEMPORAIS	74
4.1 Introdução	74
4.2 Metodologia	74
4.3 Modelo	75
4.4 Resultados obtidos	82
4.4.1 Comportamento das redes neurais com a alteração da camada oculta	84
4.4.1.1 Rede com conjunto de entrada composto por dois elementos	84
4.4.1.2 Rede com conjunto de entrada composto por quatro elementos	88
4.4.1.3 Rede com conjunto de entrada composto por seis valores	91
4.4.1.4 Rede com conjunto de entrada composto por oito valores	94
4.4.2 Estudo comparativo dos casos processados com base no valor de MSE	97
CAPÍTULO V. CONCLUSÕES E SUGESTÕES PARA FUTUROS TRABALHOS	103
5.1 Conclusões	103
5.2 Sugestões para futuros trabalhos	107
REFERÊNCIAS BIBLIOGRÁFICAS	109

LISTA DE FIGURAS

Figura 1 – “Representação gráfica do ajuste de série temporal utilizando um processo de regressão – modelagem constante”	18
Figura 2 – “Representação gráfica do ajuste de série temporal utilizando um processo de regressão – modelagem linear”	20
Figura 3 – “Representação gráfica do ajuste de série temporal utilizando um processo de regressão – modelagem quadrática”	22
Figura 4 – “Representação gráfica do ajuste de série temporal utilizando um processo de regressão – modelagem exponencial”	24
Figura 5 – “Representação gráfica que mostra o grau de participação de valores históricos da série temporal na formação do prognóstico de valores atuais, utilizando a técnica de ajuste exponencial”	26
Figura 6 – “Representação gráfica do ajuste exponencial simples aplicado a série temporal com característica constante”	27
Figura 7 – “Representação gráfica do ajuste exponencial simples aplicado a série temporal com característica crescente”	27
Figura 8 – “Representação gráfica do ajuste exponencial duplo-caso linear- aplicado a série temporal com característica crescente”	30
Figura 9 – “Representação gráfica da comparação entre o ajuste exponencial simples e duplo aplicado a série temporal com característica crescente”	32
Figura 10 – “Representação gráfica da comparação entre o ajuste exponencial duplo e a modelagem efetuada com regressão”	32
Figura 11 – “Representação gráfica do ajuste exponencial quadrático aplicado a série temporal com característica crescente”	34
Figura 12 – “Representação gráfica do ajuste exponencial triplo aplicado a série temporal com característica crescente e sazonalidade”	35
Figura 13 – “Representação gráfica da função erro do ajuste da Figura 12”	39
Figura 14 – “Representação gráfica da função autocorrelação dos valores históricos de série temporal para determinação do modelo Box-Jenkins a ser utilizado”	42
Figura 15 – “Representação gráfica de ajuste Box-Jenkins, com intervalo de confiança de 90%, utilizando o modelo autoregressivo(AR)”	43

Figura 16 – “Representação gráfica da análise comparativa de testes de autocorrelação dos valores históricos de série temporal, com e sem os efeitos de sazonalidade”	43
Figura 17 – “Representação gráfica de ajuste Box-Jenkins aplicado a série temporal com característica crescente e sazonalidade”	44
Figura 18 – “Representação gráfica de ajuste Box-Jenkins aplicado a séries temporais multivariáveis – ajuste da variável 1”	46
Figura 19 – “Representação gráfica de ajuste Box-Jenkins aplicado a séries temporais multivariáveis – ajuste da variável 2”	46
Figura 20 – “Exemplo de rede neural”	47
Figura 21 – “Neurônio artificial”	48
Figura 22 – “Exemplo simples de rede neural”	49
Figura 23 – “Rede classificadora binária”	51
Figura 24 – “Exemplo de uma rede simples para predição”	53
Figura 25 – “Uma rede neural para previsão do preço do dólar”	54
Figura 26 – “Rede neural para aproximação de funções”	55
Figura 27 – “Rede neural de quatro camadas sem retroalimentação”	56
Figura 28 – “Rede de função de base radial”	61
Figura 29 – “Rede neural para prognóstico de séries temporais caóticas”	71
Figura 30 – “Rede neural utilizada no caso exemplo da pesquisa”	75
Figura 31 – “Processamento do neurônio j de uma rede neural”	76
Figura 32 – “Função sigmoideal”	77
Figura 33 – “Derivada da função sigmoideal”	79
Figura 34 – “Representação gráfica da série temporal em estudo-índice BOVESPA”	82
Figura 35 – “Representação gráfica da série temporal com valores normalizados”	83
Figura 36 – “Resultado do treinamento da rede $2 \times 8 \times 1$ ”	84
Figura 37 – “Resultado de trecho do treinamento da rede $2 \times 8 \times 1$ ”	85
Figura 38 – “Valores de MSE para as redes com conjunto de dois valores na camada de entrada”	86
Figura 39 – “Valores de prognóstico para redes com dois valores na camada de entrada”	87
Figura 40 – “Desempenho das redes com dois componentes na camada de entrada”	87
Figura 41 – “Resultado do treinamento da rede $4 \times 8 \times 1$ ”	88
Figura 42 – “Resultado de trecho do treinamento da rede $4 \times 8 \times 1$ ”	89

Figura 43 – “Valores de MSE para as redes com conjunto de quatro valores na camada de entrada”	89
Figura 44 – “Valores de prognóstico para redes com quatro valores na camada de entrada”	90
Figura 45 – “Desempenho das redes com quatro componentes na camada de entrada”	90
Figura 46 – “Resultado do treinamento da rede $6 \times 4 \times 1$ ”	91
Figura 47 – “Resultado de trecho do treinamento da rede $6 \times 4 \times 1$ ”	92
Figura 48 – “Valores de MSE para as redes com conjunto de seis valores na camada de entrada”	92
Figura 49 – “Valores de prognóstico para redes com seis valores na camada de entrada”	93
Figura 50 – “Desempenho das redes com seis componentes na camada de entrada”	94
Figura 51 – “Resultado do treinamento da rede $8 \times 8 \times 1$ ”	94
Figura 52 – “Resultado de trecho do treinamento da rede $8 \times 8 \times 1$ ”	95
Figura 53 – “Valores de MSE para as redes com conjunto de oito valores na camada de entrada”	96
Figura 54 – “Valores de prognóstico para redes com oito valores na camada de entrada”	96
Figura 55 – “Desempenho das redes com oito componentes na camada de entrada”	97
Figura 56 – “Desempenho da rede com dois elementos na camada de entrada”	98
Figura 57 – “Desempenho da rede com quatro elementos na camada de entrada”	98
Figura 58 – “Desempenho da rede com seis elementos na camada de entrada”	99
Figura 59 – “Desempenho da rede com oito elementos na camada de entrada”	99
Figura 60 – “Desempenho da fase de treinamento das redes do estudo de caso”	100
Figura 61 – “Desempenho da fase de prognóstico das redes do estudo de caso”	101
Figura 62 – “Diferenças percentuais de valores prognosticados para as redes com dois elementos da entrada”	101
Figura 63 – “Diferenças percentuais entre valores históricos e simulados da rede $2 \times 4 \times 1$ ”	102
Figura 64 – “Desempenho global da rede $2 \times 4 \times 1$ ”	102

RESUMO

MARTIN, Claudio. Aplicação de redes neurais para prognóstico com base em séries temporais. Florianópolis, 2000. 116p. Dissertação (mestrado em Engenharia de Produção) – curso de pós-graduação em Engenharia de Produção, Universidade Federal de Santa Catarina.

Essa pesquisa visa apresentar as técnicas utilizadas para o prognóstico de demandas, a partir de séries temporais, e foi motivada pela necessidade de definir valores futuros com base em valores históricos, altamente influenciados pelas variáveis ambientais, e portanto difíceis de serem precisadas.

As técnicas apresentadas se reportam aos trabalhos desenvolvidos por **Makridakis et al (1982)** em que foram avaliadas as diversas metodologias existentes na época, através do conhecido trabalho “*M – Competition*”, em que foram testadas 1001 séries temporais reais.

O objetivo da pesquisa, no entanto, é encontrar uma ferramenta matemática que tenha a capacidade de dar respostas a séries temporais não-lineares, considerando que, cada vez mais, as séries históricas, sujeitas às variações ambientais, não apresentam um comportamento compatível com a utilização de modelos lineares.

Alguns desenvolvimentos efetuados utilizando modelagem não-linear, e descritos no trabalho de **Gooijer e Kumar (1992)**, permitem avaliar o grau de complexidade envolvido no prognóstico de séries temporais com características de não-linearidade.

A partir dos anos 80, com o trabalho de **Rumelhart e McClelland (1986)**, foram desenvolvidas as redes neurais que são modelos inspirados no funcionamento dos neurônios biológicos, os quais permitem inúmeras variações de arquitetura, podendo apresentar muitas vantagens sobre os métodos estatísticos tradicionais.

A pesquisa efetuada, além de apresentar os métodos tradicionais, utiliza como ferramenta, redes neurais para o prognóstico de séries temporais, e apresenta um estudo de caso da série temporal real do índice BOVESPA..

Palavras-chave: Séries Temporais, Redes Neurais, Prognóstico, Backpropagation.

ABSTRACT

MARTIN, Cláudio. Aplicação de redes neurais para prognóstico com base em séries temporais. Florianópolis, 2000. 116p. Dissertação (mestrado em Engenharia da Produção) – curso de pós-graduação em Engenharia da Produção, Universidade Federal de Santa Catarina.

This research presents the techniques used to forecasting the demands from time series, and it was motivated by the demand to set future values based on historical ones, which are influenced by the environmental variables and consequently very difficult of being precised.

Those techniques search the studies that were developed by **Makridakis et al** (1982), in which the different methodologies – existing at this time – were evaluated through the well known work “*M – Competition*”, where 1001 real time series were tested.

The research aim is to find a mathematical instrument that has the capacity to give responses to non-linear time series, considering that, more and more, the historical series subject to environmental variation do not present behaviour compatible with the utilisation of linear standard.

Some developments that used the non-linear standard described in the **Gooijer and Kumar's** studies (1992), allow to evaluate the degree of complexity involved in forecasting of the time series where the characteristic is non-linear.

Since 1980's, from **Rumelhart and McClelland's** studies (1986), the neural networks - standards inspired by the functioning of biological neurons that allow several architectures variations – have been developed, presenting more advantages than the traditional statistics methodologies do.

This research, besides presenting the traditional methods, uses neural networks as mathematical instrument for forecasting time series, and it also introduces a case study of the real time series from BOVESPA.

Key-Words: Time Series, Neural Networks, Forecasting, Backpropagation.

Capítulo I. Introdução

Muitos gerentes nos dias atuais, deparam-se com novas realidades e incertezas, e conseqüentemente estão em permanente confronto com questões críticas ligadas às suas companhias, onde a competitividade vislumbrada no futuro exige a capacidade de mudanças rápidas em função do meio ambiente que apresenta mutações também frequentes. Em síntese, eles devem decidir em um ambiente com uma quantidade bastante expressiva de variáveis que podem influenciar o julgamento.

Portanto, ao longo das últimas décadas, tem havido um grande número de pesquisas visando entender e prognosticar o futuro. Essas pesquisas ajudaram teoricamente os gerentes a efetuarem os julgamentos em função dos diversos fatores associados aos métodos, assim como as técnicas proliferaram, gerenciar e prognosticar o comportamento econômico e projetar cenários tornaram-se atividades com incertezas menores. Então equipado apenas com um pequeno histórico, poucos dados questionáveis e ferramentas teóricas frágeis, o gerente deve, contudo, tomar decisões críticas com respeito a um futuro em constante mutação e portanto incerto.

Apesar dessas dificuldades, os executivos responsáveis diretamente pelas decisões de futuro das empresas têm se utilizado de produtos desenvolvidos pelo mercado para avaliação periódica da tecnologia, com o fim de guiar o desenvolvimento de produtos e programas para acomodar o fluxo industrial.

Pelo fato de que cada técnica apresenta vantagens e fragilidades, cada situação de prognóstico é limitada por restrições, por exemplo, de tempo, custo, competências ou dados.

Efetuando-se o balanço entre vantagens e desvantagens das técnicas desenvolvidas, apesar das limitações e requisitos, são ferramentas indispensáveis à disposição do gerenciamento de decisões.

O crescimento das empresas como conseqüência do processo acentuado de globalização, nas quais a produção está sendo executada onde melhor possa trazer resultados compensadores, e onde o processo produtivo – ou mesmo a prestação de serviço ou desenvolvimento de uma atividade complexa – é executado em forma de parceria, está a exigir mecanismos de acompanhamento e controle cada vez mais sofisticados.

Esses mecanismos de acompanhamento, que sempre estiveram presentes no dia a dia do processo produtivo, nos dias atuais dada a complexidade das operações envolvidas, necessitam apresentar resultados cada vez mais confiáveis com respeito às demandas de produtos.

O conhecimento das demandas de produtos influenciam o desenvolvimento da logística industrial, cuja atividade é responsável por um custo, em algumas empresas, entre 4% e 30% do faturamento (LaLonde e Zinszer,1976), e mais recentemente estimado em 10,5% (Ballou,1999; Davis e Drumm,1996), podendo vir representar uma economia significativa para as empresas que têm uma preocupação com o fluxo de materiais e de informações.

A pesquisa efetuada pretende apresentar as técnicas utilizadas para a solução dos problemas relacionados com o prognóstico de demanda, para atender às necessidades da logística industrial, e mais particularmente a utilização de redes neurais para determinar, a partir de séries temporais, prognósticos de demanda com horizonte de curto prazo.

1.1 - Justificativa

Em incursões pelo setor automotivo , o tema que sempre tem se mostrado importante, além dos aspectos de produção, é aquele voltado para a definição da logística industrial e que é em última instância o planejamento que definirá o sucesso empresarial e conseqüentemente a própria sobrevivência da organização. Logo métodos de prognóstico de demandas são extremamente úteis para definir o planejamento logístico a médio e longo prazos - além de 6 meses.

O ajuste para o planejamento de curto prazo, efetuado com as metodologias tradicionais utilizadas para o prognóstico de médio e longo prazo, podem conduzir a resultados satisfatórios, desde que a demanda em estudo apresente um comportamento uniforme.

Para responder a comportamentos irregulares de demanda, é necessário obter outro ferramental matemático que possa prognosticar o comportamento a partir de premissas que venham a apresentar descontinuidades. Esse ferramental está relacionado com a utilização de técnicas de Inteligência Artificial mais precisamente com a utilização de redes neurais, as quais podem responder com maior precisão às necessidades dos requisitos da logística empresarial (Asakawa *et al.*1994).

Essas técnicas poderão servir para qualquer setor que trabalhe com demandas irregulares, não se atendo somente ao setor de produção, mas também à rede de consumo, assim como por exemplo, na definição da estratégia de aquisição de produtos pelas redes de supermercados onde uma grande parcela de produtos apresenta uma série temporal de demanda irregular.

A relevância social e econômica do trabalho que nos propomos realizar está:

- a) na necessidade de se encontrar melhores soluções para os problemas de logística;
- b) na dificuldade em se resolver problemas de programação não linear multi-dimensionais.

1.2 - Estabelecimento do problema

Artigos apresentados em publicações internacionais (Fisher et al. 1994, Lin 1993) apresentam a necessidade de se efetuar o prognóstico de demandas como uma ferramenta indispensável para o planejamento.

De acordo com (Tanwari e Betts,1999), *“Prognosticar corretamente é uma ferramenta de auxílio para planejar eficientemente e eficazmente, bem como controlar as atividades de logística. O prognóstico de demandas de produtos industriais providencia os dados necessários para o planejamento e controle de todas as áreas funcionais, incluindo a logística, o marketing, a produção e as finanças. Os níveis de demanda necessários associados à cadência requerida de produção, afetam consideravelmente a capacidade produtiva, a necessidade de recursos financeiros e a estrutura geral da corporação.”*

Prognosticar, incorretamente, demandas pode conduzir a um acréscimo nos custos finais dos produtos, afetando a participação no mercado, gerando situações indesejáveis de demanda reprimida e/ou estoques elevados.

Quanto mais precisa a capacidade de prognosticar, tanto menor será o custo debitado ao produto deste prognóstico, conforme [Fisher et al,1994], *“ Companhias que industrializam produtos com duração de vida longa e preços estabilizados não necessitam efetuar mudanças nos seus critérios de prognóstico e planejamento. O prognóstico destes produtos geralmente são bastante consistentes com a demanda real, além de que uma eventual falha no processo de prognóstico é perfeitamente incorporado ao sistema de custos, por tratar-se de produto com vida longa. No entanto para companhias que trabalham com produtos novos ou de grande sazonalidade, ou mesmo com duração de vida curta, a resposta do sistema de prognóstico deve ser extremamente confiável e correta. Cada sistema de manufatura cuja*

capacidade está exaurida durante os períodos de pico de produção, pode beneficiar-se produzindo melhor com a utilização dos períodos de baixa produção, fora do período de pico. Da mesma forma, o comércio varejista, que tem dificuldades para prognosticar demandas, pode beneficiar-se, aprendendo quais produtos e a quantidade a ser encomendada do setor industrial antes do início da estações (verão, inverno), e quais os incrementos necessários nas encomendas ao longo da estação.”

O problema consiste em desenvolver um método geral para solução de problemas de logística, especificamente no que diz respeito à necessidade de estabelecer quantidades para as áreas produtivas, tendo como elementos para efetuar o prognóstico, as demandas ocorridas em períodos anteriores e disponibilizadas na forma de séries temporais.

A determinação de demandas futuras é um problema que vem sendo trabalhado desde longa data, uma vez que representa um item fundamental para o estabelecimento da logística de setores produtivos, e tem sido objeto de inúmeras estudos metodológicos, visando obter resultados confiáveis para o planejamento industrial.

1.3 - Subordinação do tema a outras áreas do conhecimento científico

A busca de resultados, a partir de observações do comportamento temporal das variáveis em estudo, tem estimulado os pesquisadores no sentido de encontrar métodos capazes de dar respostas a partir de processos matemáticos, que apresentem um grau de aderência à série de dados compilados no tempo, de forma a satisfazer as necessidades de obter valores confiáveis para as demandas futuras.

Os métodos convencionais- regressão e técnicas exponenciais - têm apresentado respostas nem sempre satisfatórias às necessidades dos usuários, além de que as necessidades estão cada vez mais baseadas em séries temporais com um espaço amostral de dados com descontinuidades e para horizontes de curta e média duração.

A presença de descontinuidades e de efeitos de sazonalidade não são contemplados em muitos dos processos convencionais e podem conduzir a erros de interpretação e análise de resultados (Hill *et al.*1996)

Outra metodologia para a solução de problemas de demandas industriais prognosticadas a partir de series temporais está relacionada com a utilização de uma ferramenta da inteligência artificial, as redes neurais artificiais. A aplicação de redes neurais utiliza uma metodologia que permite, a partir de valores observáveis do passado da variável em estudo e com a arquitetura computacional na forma de neurônios, a exemplo do funcionamento do cérebro humano, estabelecer uma resposta adequada para valores que extrapolem o tempo presente.

Para resolver esses problemas o emprego de técnicas de redes neurais aplicadas a séries temporais apresenta resultados bastante satisfatórios para horizontes de curto prazo, permitindo, a partir de um processo de aprendizagem, ensinar a rede utilizando os valores históricos para dar as respostas de prognósticos futuros da série.

As redes neurais tipicamente apresentam-se de duas formas: a) Se não houver laços de realimentação, o sinal produzido a partir de um conjunto de entrada mover-se-á somente em uma direção, e a saída será exatamente o conjunto de valores que se seguem na série temporal em estudo. Nesse caso, a rede funciona como uma função matemática não-linear dos valores da entrada. Esse tipo de rede é mais utilizado em prognóstico de séries temporais, com valores passados da série temporal utilizados como valores de entrada, e os valores esperados futuros, como saída dos dados; b) o segundo tipo de rede utiliza a realimentação dos caminhos ao longo da rede, funcionando como um sistema dinâmico, convergindo o processo para as saídas desejadas e, portanto, variáveis com o tempo. Os valores de saída podem oscilar ou estabilizar em determinado valor, ou ainda, desde que a função de entrada introduza não-linearidades no sistema, este poderá vir a tornar-se caótico.

1.4 - Definições das variáveis a serem trabalhadas

Considerar-se-ão no problema uma sucessão de valores, obtidos a partir de medições efetuadas ao longo do tempo, constituindo uma base de dados e, portanto, representativos de uma série temporal.

Os valores a serem trabalhados foram obtidos de uma base de dados , e representam o índice médio do valor **BOVESPA**, que mede o desempenho da bolsa de valores de São Paulo. Os valores desta base de dados apresentam as variações do índice a cada dia de pregão, refletindo as variações próprias das bolsas de valores, mas também as influências introduzidas pelos elementos políticos da esfera nacional e internacional.

Na definição da variável, está sendo considerado que não existe nenhuma outra variável ambiental que provoque variações no índice em exame, portanto o problema será tratado como sendo influenciado única e exclusivamente pelo desempenho temporal observado em períodos anteriores.

1.5 - Limitações do tema

O desenvolvimento do tema não aborda, de forma exaustiva, os métodos convencionais utilizados para a determinação de prognósticos de demandas, mas apresenta os mais empregados nesse tipo de problema, como o equacionamento básico e algumas características.

A proposta do tema é a aplicação de metodologia que utiliza como ferramenta as redes neurais, sem no entanto entrar no mérito comparativo com algoritmos que utilizam métodos convencionais

Da mesma forma aplicar-se-á uma determinada arquitetura de rede neural, não tendo como finalidade obter aquela que melhor possa representar a série temporal adotada no estudo, mas obter uma metodologia consistente para a solução de séries temporais com a utilização dessa ferramenta.

1.6 - Motivação pessoal para o tema

No aspecto pessoal vários motivos conduziram-me para a escolha do tema, no entanto o motivo principal relaciona-se à importância atribuída pelo setor produtivo de bens de consumo ao planejamento da logística, cujo objetivo final é produzir somente o necessário, transportar e distribuir os produtos, de forma a garantir a sobrevivência das corporações (Courtney *et al.* 97).

Em quase todos os assuntos sobre os quais os administradores devem adotar atitudes de decisão, estas não mais são tomadas de forma intuitiva, mas a partir cada vez mais da análise de índices que possam representar as variáveis ambientais envolvidas no problema. Essas variáveis, que podem ser de qualquer natureza, são transformadas em números e contribuem para a redução das componentes aleatórias da tomada de decisão.

Dentro do planejamento da logística, um dos pontos balizadores para o bom desempenho industrial como um todo é a determinação, com a maior exatidão possível, das quantidades das demandas a serem produzidas, de forma que o mercado existente possa apresentar condições de absorver.

O tema tem sido estudado de longa data pelos pesquisadores e, com a introdução dos algoritmos que utilizam técnicas de redes neurais, está permeando todas as áreas do conhecimento, tomou novo impulso, tornando-se em algumas situações, a exemplo do mercado de ações, uma ferramenta indispensável para a tomada de decisões em situações de risco.

Apesar de estar sendo aplicado em uma base de dados específica, o estudo deste problema permitirá o desenvolvimento de pesquisas, nos mais diversos campos do conhecimento, considerando-se que o assunto permite a generalização de procedimentos para outros temas.

1.7 - Objetivos gerais e específicos

O objetivo geral da pesquisa é apresentar as metodologias de prognóstico de demanda existentes, suas características, vantagens e desvantagens, e suas limitações, permitindo o correto entendimento do problema de logística industrial .

Os objetivos específicos consistem em:

1-Propor um ferramental, utilizando teorias de inteligência artificial, para a determinação de prognósticos de demandas mercadológicas capazes de reconhecer e de manipular dados com características de séries temporais de demanda irregulares, atendendo necessidades de logística do setor produtivo.

2-Apresentar resultados provenientes da utilização de arquiteturas diferentes de redes neurais, mostrando sua influência sobre o desempenho esperado.

1.8 - Hipóteses gerais e específicas

Como hipótese geral podemos postular que os resultados obtidos com a aplicação de redes neurais, para representar uma sucessão de valores obtidos de uma série temporal, apresentam uma aderência à curva, mesmo quando esta série apresenta descontinuidades, podendo, portanto representar a variação com o tempo, da variável em estudo.

A hipótese específica a ser levantada está relacionada com a capacidade apresentada pelos métodos com redes neurais de permitir o prognóstico com horizonte de curto prazo, com bastante exatidão.

Outra questão a ser verificada é a diferença de desempenho quando da utilização de arquiteturas diferentes para as redes neurais, mostrando que existem redes apropriadas para cada tipo de estudo, como consequência do comportamento diferenciado das variáveis ambientais do problema.

1.9 - Descrição dos capítulos

A dissertação está estruturada em cinco capítulos, com os seguintes conteúdos:

O capítulo I – **Introdução**, conceitua em termos gerais a problemática relacionada com a necessidade de se efetuar o planejamento logístico de setores produtivos, visando à obtenção de resultados satisfatórios no desempenho industrial.

Uma das dificuldades apontadas para a execução do planejamento desejado é a falta do conhecimento preciso do comportamento do mercado consumidor, ou mais precisamente, qual é a demanda a ser exigida no futuro próximo, considerando que uma grande parcela dos produtos apresentam flutuações significativas de consumo.

Os objetivos e hipóteses, gerais e específicos, bem como as limitações do tema igualmente estão propostos e em concordância com a motivação pessoal para a escolha do tema, metodologias para a determinação de demandas industriais com base em séries históricas, para subsidiar o planejamento da logística industrial.

O capítulo II – **Séries Temporais**, apresenta as metodologias existentes na literatura técnica e que dão suporte para a determinação de valores necessários ao planejamento logístico de demandas.

Assim, os métodos matemáticos convencionais, como os de regressão, ajuste exponencial, e a técnica de Box-Jenkins, dentre os mais utilizados, são apresentados, e representam as metodologias atualmente de maior uso entre os planejadores e, portanto, de grande importância para o estudo.

O capítulo III – **Redes Neurais**, introduz o estudo de redes neurais como ferramenta para o prognóstico de demandas futuras, consistindo em resposta às necessidades de solucionar problemas que apresentam séries temporais irregulares.

A teoria de redes neurais está apresentada de forma sucinta, mostrando a classificação em redes com aprendizagem supervisionada e não supervisionada, e as inúmeras possibilidades de aplicações nos mais diversos campos do conhecimento.

Os desenvolvimentos mais recentes para modelagem, teste e prognóstico de séries temporais não-lineares são apresentados, de forma a permitir comparações qualitativas com a utilização de redes neurais, e embasar a escolha do método com redes neurais para as séries irregulares.

O capítulo IV – **Aplicação de Redes Neurais em Séries Temporais**, apresenta metodologia a ser utilizada na obtenção dos dados necessários para responder à questão levantada no estabelecimento do problema, bem como para justificar os objetivos propostos e corroborar as hipóteses levantadas sobre a modelagem a ser utilizada.

Apresenta o modelo de rede neural a ser utilizada, mostrando os passos a serem executados até a obtenção dos resultados. E, além de apresentar os resultados, efetua a análise, seguindo a metodologia escolhida e a modelagem matemática adotada, verificando a identificação com os objetivos propostos e as hipóteses levantadas.

O capítulo V – **Conclusões e sugestões para trabalhos futuros**, apresenta as conclusões gerais e específicas da pesquisa efetuada, indicando temas para a continuidade do trabalho dentro da mesma linha.

1.10 – Conclusão - A importância do tema para a logística.

As atividades a serem gerenciadas pela logística incluem as seguintes partes: transporte, manutenção de inventário, processamento de ordens, compras, armazenamento, entrega de materiais, embalagem de materiais, prestação de serviços ao consumidor e planejamento da produção (Ballou,1999).

Dentre essas atividades enumeradas, e devido ao aumento da complexidade das operações envolvidas para o seu gerenciamento, novos mecanismos capazes de prever o comportamento do mercado - e conseqüentemente o desenvolvimento de toda a cadeia produtiva - fazem com que as premissas utilizadas para gerenciar processos venham agregadas de novos algoritmos capazes de fornecer as respostas para o planejamento da logística.

A necessidade de projeção da demanda permeia todo o processo produtivo afetando significativamente os níveis de capacidade produtiva, as necessidades financeiras e a estrutura do negócio fazendo com que cada setor funcional tenha que resolver os seus próprios problemas de prognóstico.

Prognóstico logístico consiste na definição espacial e quase sempre na definição temporal da natureza da demanda, na definição da extensão da sua variabilidade e do seu grau de incerteza (Ballou,1999).

As variações da demanda com o tempo são oriundas do decréscimo de faturamento, efeitos de sazonalidade no modelo de demanda e outras causas gerais geradas por múltiplos fatores.

A maioria dos prognósticos que utilizam métodos para determinação de demanda para um horizonte de curto prazo (até 6 meses), trabalham com variações temporais, geralmente denominadas séries temporais.

A logística utiliza ambas as dimensões espacial e temporal , isto é, a logística deve conhecer onde o volume de demanda acontecerá e quando isto ocorrerá.

Nesse contexto, os requisitos necessários para prognosticar a logística do comportamento de mercado receberam um importante aliado com a teoria das redes neurais que podem responder às oscilações de mercado para um horizonte de curto prazo e, portanto, de difícil estimativa.

Os prognósticos com a utilização de redes neurais são mais precisos e podem responder mais adequadamente nas situações em que as séries temporais apresentam descontinuidades, considerando que os algoritmos apresentam soluções baseadas em processos de aprendizagem das ocorrências anteriores, que é a característica da rede neural.

As oscilações do mercado consumidor se refletem em toda a cadeia produtiva, vindo a exigir a necessidade de ferramentas capazes de prognosticar as demandas em todos os níveis do percurso da linha, e conseqüentemente toda uma reordenação logística para atender economicamente à nova situação, obtendo o menor custo/benefício, isto é, minimizando a variável custo e maximizando a variável lucro.

Esse tipo de problema já apresenta soluções advindas de algoritmos desenvolvidos a partir das necessidades que eram apresentadas, sendo que, no entanto, a maioria deles são mais adequados para planejamento logístico com horizonte de médio e longo prazos. Neste contexto é entendido como tempo curto um período de planejamento inferior à 6 meses; tempo médio, entre 6 meses e alguns anos; e tempo longo, mais do que alguns anos.

As demandas intermitentes, que apresentam séries temporais irregulares, são de difícil prognóstico a partir de técnicas já estabelecidas para as demandas com variações uniformes, no entanto representam mais de 50% dos produtos manuseados nas empresas (Ballou,1999), representando, portanto, um problema de prognóstico de demanda especial para os profissionais da logística.

Assim, para a situação apontada hoje já estão sendo utilizadas teorias de Inteligência Artificial notadamente as redes neurais, as quais podem responder adequadamente às necessidades de prever as demandas a partir de banco de dados que não apresentam uma série temporal uniforme.

Capítulo II. Séries Temporais

2.1 - Introdução

As técnicas de prognóstico podem ser classificadas em dois tipos: qualitativas e quantitativas.

A metodologia que utiliza pesquisas de comportamento das preferências apontadas pelas características do mercado consumidor, utiliza técnicas qualitativas de prognóstico, apesar de a análise estar baseada em dados de acúmulo de valores quantitativos.

A aceitação de prognósticos qualitativos geralmente baseia-se na premissa de que a fonte de dados é representativa e constante. As técnicas quantitativas estão orientadas para prognosticar em situações futuras em que os valores futuros representam uma projeção das situações ocorridas no passado, ou seja, são métodos de extrapolação de dados.

Desse modo, podemos pensar que um modelo subjacente existe e que pode predizer o futuro e que os dados históricos podem ser utilizados para estimar os parâmetros desse modelo, porque existem outros dados históricos que podem testar a veracidade e exatidão do modelo adotado.

Considere-se que técnicas de prognóstico utilizando dados históricos referem-se à análise de séries temporais.

A primeira questão a ser colocada quando da utilização de um método de prognóstico é para que função está sendo utilizado. A resposta a esta pergunta indicará qual o procedimento a ser desenvolvido.

A utilização pode estar sendo necessária para determinar demandas novas ou talvez para ajustar demandas de períodos com sazonalidade, ou mesmo para determinar quantitativos para períodos de extremas flutuações do mercado consumidor.

Cada necessidade apresenta soluções específicas e, dependendo da quantidade de variáveis independentes que intervêm no valor da variável dependente, a modelagem matemática pode assumir uma complexidade tal que venha a exigir programas computacionais.

Brown (1963) sugere que os métodos de prognóstico devam apresentar critérios estabelecidos em três categorias: *exatidão*, *simplicidade computacional* e *flexibilidade de ajuste do padrão de resposta*. O significado de exatidão em prognosticar diz respeito à medida de quanto é precisa a resposta do sistema proposto, sempre que um evento que apresenta as condições atuais venha a ocorrer. Quanto maior a exatidão requerida, tanto maior será o esforço computacional necessário, o que contradiz o segundo requisito apresentado de simplicidade computacional, no entanto os dois critérios devem ser perseguidos, balanceando a exatidão com a minimização do tempo computacional necessário.

A flexibilidade de ajuste do padrão de resposta é a medida da habilidade, apresentada pelo modelo de prognóstico adotado, em adaptar-se às condições de mudanças apresentadas pelos dados.

Outra variável a considerar é o tempo de espera, definido como sendo o tempo decorrido entre a decisão do prognóstico e a implementação do resultado; um problema apresentado por um tempo de espera longo é de que o erro na aplicação do prognóstico é tanto maior à medida que este tempo aumenta.

Outra observação diz respeito ao horizonte de prognóstico, definido como sendo o tempo decorrido entre o instante em que as ações do prognóstico são iniciadas e o momento em que os efeitos dessas ações são observáveis, caso, por exemplo, no setor automotivo, em que uma grande gama de ações são desenvolvidas antes do início efetivo da produção e venda de um novo modelo de automóvel.

Prognosticar é geralmente definido como sendo um processo inteligente que pode fazer a variável a ser determinada assumir determinados valores esperados.

Algumas vezes, alguns processos de causa-efeito podem ser modelados, por exemplo a demanda de combustível em um determinado estado estar relacionada às mudanças na população e turismo. E essas mudanças, que podem projetar o consumo de combustível, são também bons prognósticos futuros da população e de turistas. Por outro lado existem processos desconhecidos, mas apresentam dados consistentes com o tempo, de forma que um modelo construído com base em dados históricos poderá estabelecer valores futuros, permitindo a tomada de decisões em planejamentos futuros.

No entanto, em uma grande quantidade de situações que ocorrem, quando da necessidade de planejamento de ações para correção de percursos, inexistem ou não existem dados suficientes para estabelecer uma modelagem de prognóstico adequada para estabelecer a necessária capacidade de tomada de decisão. Nessas situações em que dados quantitativos não podem ser avaliados com segurança, os analistas devem ser realmente pessoas reconhecidamente “experts”, pois terão que desenvolver resultados a partir do desenvolvimento de modelos de prognóstico qualitativos.

A metodologia utilizada para determinar a melhor solução a ser dada para determinada questão, na ausência de processo quantitativo que possa dar suporte à decisão, é obter junto de um grupo de “experts”, a partir de um processo de adesão a cada argumento proposto, o grau de consenso em torno de cada proposição.

O painel de “experts”, conduz à estimativa de valores para as variáveis desconhecidas, desenvolvendo cenários para o melhor caso, o caso esperado e o pior caso, a partir dos quais por

um processo de obtenção de opinião da média do grupo, procura-se obter o consenso e conseqüentemente efetuar o prognóstico.

O presente trabalho explorará a tomada de decisão com base em valores obtidos a partir de modelagem matemática, portanto a decisão será baseada em prognóstico quantitativo.

Segundo Georgoff e Murdick (1986), embora muitos planejadores tenham incorporado alguma subjetividade ao seu trabalho, os gerentes acham realmente que prognosticar para o futuro é muito mais fácil a partir da utilização de dados quantitativos do que a partir de seus próprios julgamentos.

Pesquisas efetuadas (Hogarth e Makridakis, 1981) concluíram que mesmo a técnica quantitativa mais simples desautoriza a utilização de avaliações intuitivas desestruturadas de “experts” e seu julgamento para o ajuste de valores obtidos de prognósticos quantitativos, uma vez que sempre são acompanhados de uma perda de exatidão no processo.

Isto é dessa forma porque prognósticos intuitivos são suscetíveis a tendências, e os gerentes são limitados nas habilidades de processar informações e manter relações consistentes entre variáveis (Sjoberg, 1982).

2.2 – Utilização de regressão para o prognóstico de séries temporais

Vários métodos associados à utilização de séries temporais foram desenvolvidos, sendo que o prognóstico a partir de séries temporais tem por objetivos (Bedworth e Bailey, 1986):

- a) avaliar tendências significativas nos dados históricos disponíveis;
- b) retirar os valores inconsistentes da base de dados;
- c) responder dinamicamente às mudanças na demanda se elas vierem a ocorrer;
- d) projetar tendências requeridas para o futuro pesquisado, para o qual está sendo necessário a definição de prognósticos;
- e) determinar parâmetros de prognóstico, ótimos, para atender aos objetivos anteriores.

As duas primeiras premissas definem a curva de ajuste necessária a ser obtida da série histórica temporal a ser utilizada.

A resposta para as mudanças que possam vir a ocorrer no meio ambiente pode ser obtida através de um procedimento de cálculo dinâmico do processo, a cada novo valor experimentado pelo meio. A técnica de prognóstico de média móvel utiliza esse processo, adicionando os valores

mais recentes dos dados históricos, abandonando os valores mais antigos, sem antes calcular as médias de N valores que possam conduzir para um novo prognóstico.

Se existir uma relação causa-efeito entre uma variável independente e uma variável a ser prognosticada dependente, então existe a possibilidade de utilizar-se um modelo de prognóstico para a determinação futura desta variável.

O modelo de prognóstico leva em consideração os dados históricos da variável independente, e por meio da utilização de tratamento estatístico de dados, determina os melhores parâmetros para o estabelecimento das equações da variável dependente, que irá determinar os valores a serem utilizados no prognóstico futuro.

Os modelos que podem ser obtidos dessa análise apresentam a modelagem matemática:

$$S_t = a + b x_{t-2} \quad [2.2-1]$$

$$S_t = a + b x_{t-3} \quad [2.2-2]$$

$$S_t = a + b x_{t-2} + c x_{t-3} \quad [2.2-3]$$

S_t = variável a ser prognosticada no período t

x_{t-2} ; x_{t-3} = valor indicado da variável independente no período (t-2) e (t-3)

a ; b ; c = parâmetros a serem determinados.

Como exemplo tome-se o caso de uma empresa que produz unidades de ar condicionado para novas construções em uma grande cidade. A empresa necessita saber, para efetuar o planejamento de curto prazo (próximos meses), qual a quantidade a ser fabricada de modo a atender da melhor forma possível esse mercado consumidor. Para tanto poderá tomar como referência para a sua variável independente, a quantidade de permissões para construção emitidas pela Prefeitura local, a qual poderá fornecer uma importante base de dados da forma “ Unidades

de ar condicionados fabricados x permissões concedidas “ , a partir das quais é possível estabelecer os parâmetros a , b e c das equações [2.2 -1,2,3] .

Esses parâmetros podem ser obtidos a partir da aplicação de regressão linear à tabela de dados existentes, reservando os dados mais recentes para efetuar a necessária verificação de comportamento da modelagem matemática. Quanto mais completa for a representação matemática do modelo em uso, tanto menor será o erro percentual quando de aplicação a novos dados. Assim para o exemplo indicado, a modelagem matemática representada pela equação [2.2-3], apresentará o menor erro entre os três modelos propostos.

Na maioria das situações encontradas no ambiente, não é possível estabelecer uma variável independente para determinar o equacionamento para prognóstico futuro, considerando que nem sempre uma variável independente escolhida está correlacionada com a variável a ser determinada. No entanto, geralmente as variáveis em estudo apresentam padrões de comportamento temporal satisfatórios permitindo a utilização do tempo como variável independente do problema a ser resolvido.

Como prognosticar significa determinar o padrão histórico e projetar este padrão para o futuro; o engenheiro olha para os dados históricos e estabelece a hipótese para modelagem matemática que possa descrever o passado. Os dados históricos são utilizados para a determinação dos parâmetros do modelo, para finalmente avaliar valores para períodos futuros. A modelagem matemática de maior utilização para a determinação de curvas de prognóstico, é da forma polinomial de ordem n , conforme equação [2.2-4]:

$$S_t = a + bt + ct^2 + \dots + gt^{n-1} + ht^n \quad [2.2-4]$$

onde S_t representa o valor estimado a partir dos valores verdadeiros da série histórica, no tempo t , e a , b , c h são os coeficientes ajustados do polinômio.

Truncando os elementos de maior grau, podemos obter equações de ajuste bastante familiares, e que podem resolver alguns problemas relacionados com a determinação de demandas.

$$S_t = a \quad \text{modelo constante} \quad [2.2-5]$$

$$S_t = a + bt \quad \text{modelo linear com inclinação } b. \quad [2.2-6]$$

$$S_t = a + bt + ct^2 \quad \text{modelo quadrático com inclinação } b \quad [2.2-7] \\ \text{e variação da inclinação } c.$$

2.2.1 – Regressão - Modelagem constante

Pela observação da série temporal, é possível avaliar se a flutuação dos dados permite a utilização de um valor constante representativo da evolução histórica dos dados.

A equação representativa da evolução do prognóstico de demanda pode ser escrita conforme [2.2-5], e aqui rerepresentada:

$$S_t = a$$

Determinando o erro no valor prognosticado, efetuando a diferença entre o valor atual y_t determinado a partir dos dados históricos e o valor prognosticado S_t , obtém-se

$$e_t = y_t - S_t = y_t - a$$

Na análise de regressão, o valor de a é determinado a partir do valor que minimiza o erro, o qual pode ser obtido efetuando a soma do quadrado do erro para N períodos da série histórica, obtendo-se

$$\sum_{t=1}^N e_t^2 = \sum_{t=1}^N [y_t - a]^2 \quad [2.2.1-1]$$

Para determinar-se o valor apropriado para **a** diferencia-se a expressão do erro relativamente ao valor **a**, iguala-se a zero , de onde obtém-se:

$$a = \frac{\sum_{t=1}^N y_t}{N} = y_{\text{medio}} \quad [2.2.1-2]$$

Como era esperado, o valor obtido para parâmetro **a** determinar para a equação de prognóstico é a média dos últimos N valores da série temporal, o qual sofre alterações a cada novo valor da série histórica agregada aos dados, razão pela qual esta modelagem também é denominada de “técnica de prognóstico com média móvel”.

A modelagem constante de prognóstico é extremamente simples de ser implementada , e tem a vantagem de remover, com a utilização do processo de determinação da média, eventuais flutuações nos valores históricos da demanda, conforme está mostrado graficamente na Figura 1, onde mesmo uma perturbação grande nos valores de demanda, refletem com pequenas variações na determinação do prognóstico.

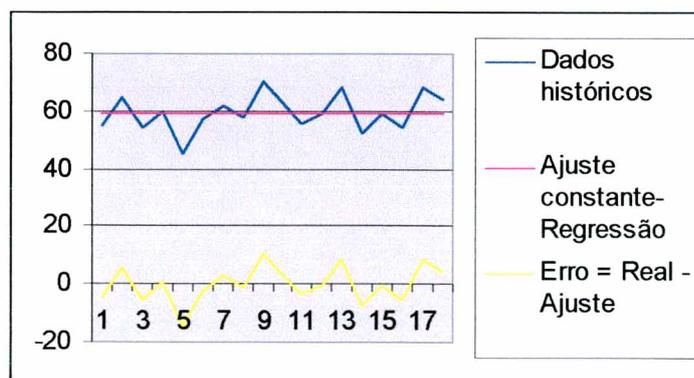


Figura 1

Para prognóstico de demanda de situações estabilizadas esta é uma característica desejada, no entanto para a grande maioria dos casos em que as flutuações são importantes termômetros de comportamento mercadológico, e que não podem ser desprezadas, esta abordagem não satisfaz as necessidades de determinação de valores para a tomada de decisões futuras de demanda.

2.2.2 – Regressão - Modelagem linear

Considere-se o caso em que os dados históricos apresentem valores que flutuam no entorno de uma determinada direção não constante, que é o caso anterior, podendo esta direção tanto indicar crescimento como declínio nos valores.

A modelagem que se adequa a esta situação é a apresentada na equação [2.2-6], cuja função erro a ser minimizada é da forma:

$$\sum_{t=1}^N e_t^2 = \sum_{t=1}^N [y_t - a - bt]^2 \quad [2.2.2-1]$$

Tomando-se a derivada da função relativamente a **a**, e na sequência a derivada em relação a **b**, Ter-se-ão duas equações nas incógnitas **a** e **b**, que resolvidas simultaneamente nos conduzem ao valor dos coeficientes, cujos valores podem ser obtidos pelas equações [2.2.2-2] e [2.2.2-3].

$$a = \frac{\sum_{t=1}^N y_t - b \sum_{t=1}^N t}{N} \quad [2.2.2-2]$$

$$b = \frac{N \sum_{t=1}^N t y_t - \sum_{t=1}^N y_t \sum_{t=1}^N t}{N \sum_{t=1}^N t^2 - \left(\sum_{t=1}^N t \right)^2} \quad [2.2.2-3]$$

O exemplo demonstrado na Figura 2 apresenta em um mesmo gráfico os valores históricos, o ajuste a partir de modelagem linear, considerando o valor assumido pela variável y_t até o ponto considerado e a reta de ajuste linear global, e que corresponde ao último intervalo do ajuste parcial.

Observa-se que mesmo com flutuações na base de dados, da mesma forma que no caso anterior, o modelo mantém a direção do conjunto de valores da base histórica.

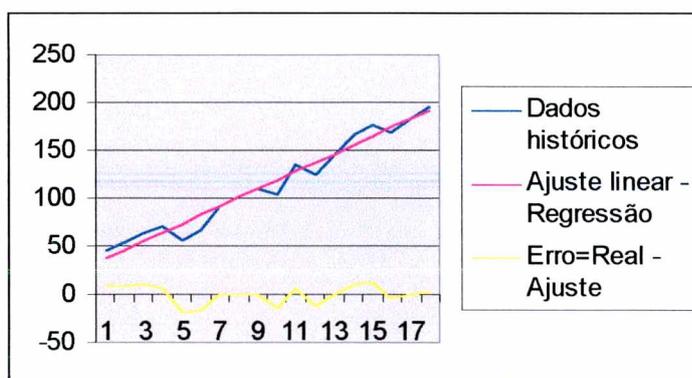


Figura 2

2.2.3 – Regressão - Modelagem quadrática

A modelagem pode sempre ser utilizada, independente da tendência do conjunto de valores históricos, pois os parâmetros a serem determinados para a equação de prognóstico, ajustar-se-ão de forma a refletir o melhor resultado com respeito à série temporal.

Assim, se a base de dados apresentar valores com pouca variação no tempo, onde o modelo constante se adequa melhor, os parâmetros b e c da equação quadrática a serem obtidos no processo, refletirão esta tendência, apresentando valores muito pequenos e que podem ser desprezados.

Aplicando o processo da soma dos quadrados do erro, conforme [2.2.3-1], pode-se obter os parâmetros que compõem a equação de prognóstico.

$$\sum_{t=1}^N e_t^2 = \sum_{t=1}^N [y_t - a - bt - ct^2]^2 \quad [2.2.3-1]$$

Derivando a equação do erro com respeito aos parâmetros a serem determinados, obtém-se um conjunto de três equações, que resolvidas simultaneamente apresentam os valores de **a**, **b** e **c**, que podem ser obtidos com o conjunto de expressões.

$$\alpha = \sum_{t=1}^N t \sum_{t=1}^N t^2 - N \sum_{t=1}^N t^3$$

$$\gamma = \left(\sum_{t=1}^N t^2 \right)^2 - N \sum_{t=1}^N t^4$$

$$\delta = \sum_{t=1}^N t \sum_{t=1}^N y_t - N \sum_{t=1}^N t y_t$$

$$\beta = \left(\sum_{t=1}^N t \right)^2 - N \sum_{t=1}^N t^2$$

$$\theta = \sum_{t=1}^N t^2 \sum_{t=1}^N y_t - N \sum_{t=1}^N t^2 y_t$$

$$b = \frac{\gamma\delta - \theta\alpha}{\gamma\beta - \alpha^2} \quad [2.2.3-2]$$

$$c = \frac{\theta - b\alpha}{\gamma} \quad [2.2.3-3]$$

$$a = \frac{\sum_{t=1}^N y_t - b \sum_{t=1}^N t - c \sum_{t=1}^N t^2}{N} \quad [2.2.3-4]$$

Na Figura 3 está apresentada uma série temporal cuja base histórica permite a utilização de um ajuste quadrático. Pode-se observar que o ajuste acompanha a evolução dos dados históricos, mas não corresponde a um processo que acompanha as variações dos dados históricos, o que pode ser visto na curva representativa do erro, que corresponde à diferença entre o valor ajustado e o valor histórico correspondente.

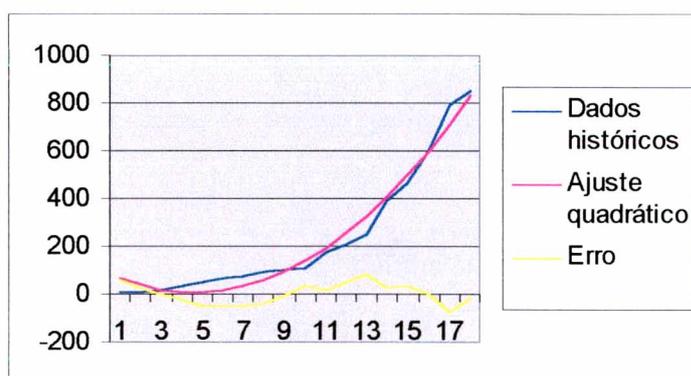


Figura 3

2.2.4 – Regressão - Modelagem exponencial

Pelo fato de que não existe uma aderência adequada dos modelos de ajuste, outras formas de modelagem são estudadas para permitir melhorar a performance do modelo de prognóstico.

A modelagem exponencial, apesar de não corresponder a uma modelagem na forma polinomial, que é o escopo do processo de regressão, também pode representar dados provenientes de séries temporais, apresentando uma equação de modelagem da forma mostrada em [2.2.4-1].

$$S_t = a e^{bt} \quad [2.2.4-1]$$

Para solucionar o problema para encontrar os parâmetros a e b , é necessário efetuar a transformação logarítmica, conforme:

$$\ln[S_t] = \ln a + \ln(e^{bt}) = \ln a + bt$$

O resultado encontrado é um modelo linear, de onde podemos obter o valor dos parâmetros a partir das equações

$$\ln a = \frac{\sum_{t=1}^N \ln(y_t) - b \sum_{t=1}^N t}{N} \quad [2.2.4-2]$$

$$b = \frac{N \sum_{t=1}^N t \ln(y_t) - \sum_{t=1}^N \ln(y_t) \sum_{t=1}^N t}{N \sum_{t=1}^N t^2 - \left(\sum_{t=1}^N t \right)^2} \quad [2.2.4-3]$$

A Figura 4 apresenta a aplicação da modelagem exponencial utilizando a mesma série temporal plotada na Figura 3 onde o ajuste foi efetuado com modelagem quadrática.

Observa-se que o ajuste exponencial aplicado a este problema estabelece um equacionamento para prognóstico com um perfil melhor que no caso da situação anterior.

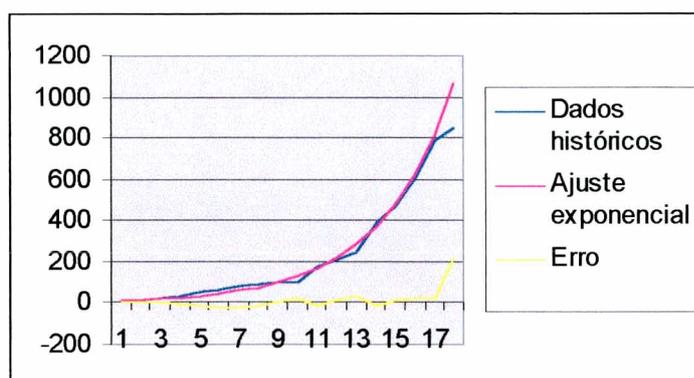


Figura 4

2.3 - Análise de séries temporais utilizando a técnica de ajuste exponencial

Trata-se, talvez, da técnica mais importante aplicada à determinação de prognósticos.

Os procedimentos adotados seguem os mesmos atributos da técnica de média móvel.

Da mesma forma que na utilização da técnica de regressão, a técnica de ajuste exponencial suave é aplicada para a determinação dos parâmetros no mesmo modelo polinomial, ou seja: constante, linear, quadrático e exponencial desde que se trabalhe na forma logarítmica.

A equação geral para este método se apresenta na forma indicada em [2.3-1].

$$S_t = \alpha y + (1 - \alpha) S_{t-1} \quad [2.3-1]$$

Na equação, a variável S_t representa o ajuste exponencial de todos os valores prévios da série temporal, anteriores ao período t , e é utilizada como constante do modelo para todos os valores a serem prognosticados nos períodos futuros.

A variável S_t é denominada como sendo o valor no período t do ajuste médio de primeira ordem do primeiro ajuste dos dados históricos obtidos da série temporal.

A variável α é a constante de ajuste e situa-se entre os valores $0 \leq \alpha \leq 1,0$.

Uma característica de S_t é que esta variável contém parcelas de todos os dados que tenham sido ajustados mesmo que não tenha havido dados anteriores.

Isto pode ser observado a partir da expansão da equação [2.3-1] para termos anteriores ao período t, onde fazendo $\beta = (1 - \alpha)$ tem-se:

$$S_t = \alpha y_t + \beta S_{t-1}$$

$$\beta S_{t-1} = \alpha \beta y_{t-1} + \beta^2 S_{t-2}$$

$$\beta^2 S_{t-1} = \alpha \beta^2 y_{t-2} + \beta^3 S_{t-3}$$

que pode ser reduzido para:

$$S_t = \alpha \sum_{i=0}^{t-1} \beta^i y_{t-i} + \beta^t S_0 = \alpha \sum_{i=0}^{t-1} (1-\alpha)^i y_{t-i} + (1-\alpha)^t S_0 \quad [2.3-2]$$

Pela equação [2.3-2] é possível observar-se que todos os dados históricos da série temporal estão representados na média, e considerando que tanto α como β são valores fracionários, os valores mais recentes participam com um peso maior, decrescendo monotonicamente à medida que os dados ficam mais distantes do ponto de estudo.

Assim, por exemplo, a equação expandida para o valor ajustado S_4 tem a forma:

$$S_4 = \alpha \left[(1-\alpha)^0 y_{4-0} + (1-\alpha)^1 y_{4-1} + (1-\alpha)^2 y_{4-2} + (1-\alpha)^3 y_{4-3} \right] + (1-\alpha)^4 S_0$$

Na equação que determina o valor a ser prognosticado S_4 , para o período $t = 4$, participam todos os valores anteriores, sendo que os mais recentes participam com um peso maior, podendo-se expressar graficamente a influência dos valores anteriores conforme indicado na Figura 5, onde no eixo das abcissas está representada a variável i da equação [2.3-2], e no eixo das ordenadas

o valor $(1 - \alpha)^i$, podendo-se aferir o grau de participação dos termos anteriores na formação do prognóstico dos valores atuais.

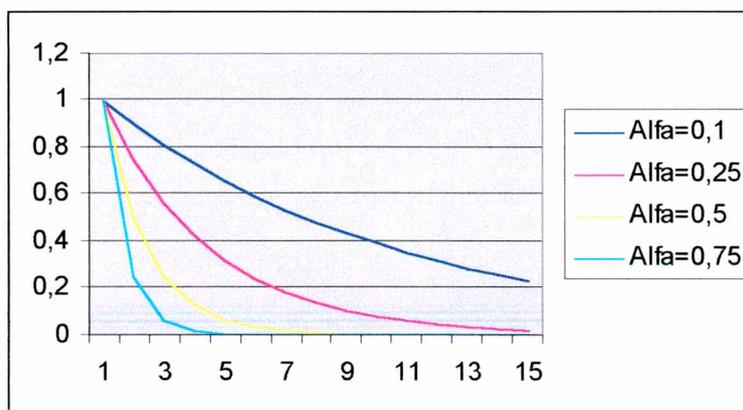


Figura 5

2.3.1 – Ajuste exponencial simples

No modelo constante de ajuste exponencial participam na formação da equação de prognóstico somente o primeiro termo da somatória da equação [2.3-2] e o termo independente.

$$S_{t+1} = \alpha y_t + (1 - \alpha) S_t = S_t + \alpha (y_t - S_t) = S_t + \alpha \varepsilon_t$$

A técnica consiste em determinar sucessivamente valores para S_t a partir do valor S_1 , adotando para valor inicial S_0 , por exemplo, a média dos três primeiros valores da série temporal, para evitar a inicialização do processo com um valor que não seja representativo da variação experimentada pelos dados históricos.

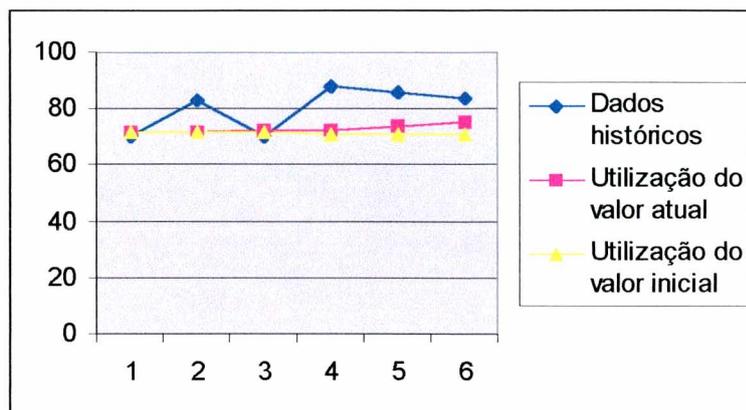


Figura 6

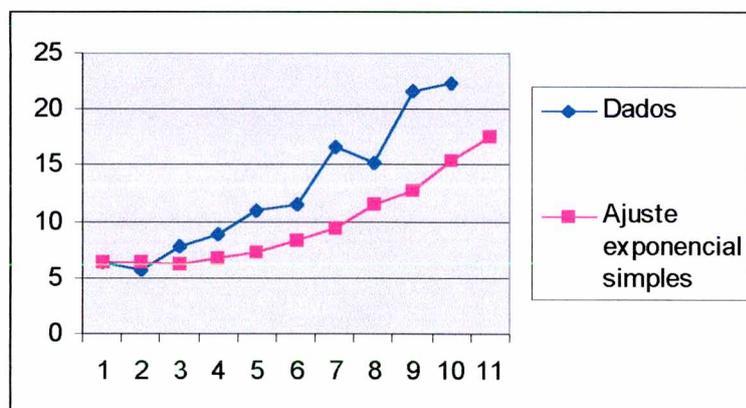


Figura 7

A Figura 6 apresenta um ajuste exponencial com modelagem constante, efetuado para uma série temporal com os dados históricos apresentados no mesmo gráfico.

O ajuste efetuado com a utilização do valor atual considera que existem dados históricos suficientes para processar o prognóstico da série temporal.

O ajuste efetuado com a utilização do valor inicial da série está pressupondo que não há dados históricos conhecidos para efetuar o ajuste, e portanto o valor atual da série é adotado em todo o percurso de prognóstico para efetuar o ajuste. Observa-se que o ajuste com a presença da série temporal tende a ser melhor que aquele onde inexistem dados para o ajuste.

Para uma grande massa de dados, em uma série que não apresente uma tendência de crescimento, a utilização de uma aproximação exponencial simples se aproxima bastante da modelagem utilizando regressão do mesmo tipo, ou seja, constante.

Isto está a mostrar que existe um valor α , equivalente a N , utilizado no método de regressão.

Na Figura 7 está representado graficamente o ajuste exponencial simples aplicado a uma série temporal com uma tendência de crescimento, onde se pode observar que o ajuste não apresenta uma aderência muito boa, demonstrando que tal processo não se ajusta a uma série com estas características.

2.3.2 – Ajuste exponencial duplo

O caso do ajuste exponencial duplo pressupõe que os dados históricos apresentados guardam um crescimento que possa ser aproximado na modelagem, e o modelo de equação prognóstico se apresenta na forma $S_t = a + bt$, e a técnica exponencial deve ser utilizada para a determinação dos parâmetros a e b .

Tomando-se a equação [2.3-2] e reescrevendo os termos para o caso linear tem-se:

$$S_t = \alpha y_t + \beta S_{t-1}$$

Manipulando convenientemente a expressão, pode-se chegar a uma equação da forma :

$$y_t - S_t = (\beta / \alpha) [S_t - S_{t-1}] \quad [2.3.2-1]$$

onde y_t é o valor no final do período t da série histórica real;

S_t é o valor prognosticado pela técnica exponencial ao fim do período t ;

S_{t-1} é o valor prognosticado pela técnica exponencial ao fim do período $t-1$;

β e α são as constantes de ajuste utilizadas pela técnica.

Pela observação da equação, pode-se concluir que a diferença entre os valores prognosticados S_t e S_{t-1} representa o acréscimo sofrido pela variável prognosticada S , para uma variação unitária da unidade de tempo que está sendo utilizada, é, portanto, a inclinação do ajuste linear que corresponde à variável b da equação $S_t = a + bt$.

Pode-se então escrever:

$$y_t - S_t = (\beta / \alpha)b \quad [2.3.2-2]$$

Suponha-se que será utilizada a técnica exponencial não somente para estimar y_t a partir de S_t , mas também S_t a partir de S_t^2 . Desta forma obteremos a equação [2.3.2-3] onde o termo S_t^2 é denominado de valor de ajuste de segunda ordem e representa o ajuste médio do ajuste médio.

$$S_t^2 = \alpha S_t + (1 - \alpha) S_{t-1}^2 \quad [2.3.2.- 3]$$

Como a obtenção de S_t^2 ocorre com os mesmos α e β que S_t , pode-se estabelecer considerando a equação [2.3.2.-2] que:

$$(\beta / \alpha)b = [y_t - S_t] = [S_t - S_t^2] \quad [2.3.2-4]$$

de onde pode-se obter:

$$b_t = (\alpha / \beta)[S_t - S_t^2] \quad [2.3.2-5]$$

Definido a_t como sendo o valor mais recente de y_t pode-se estabelecer considerando a equação [2.3.2-4]:

$$a_t = y_t = S_t + (\beta / \alpha)b = 2S_t - S_t^2 \quad [2.3.2-6]$$

O exemplo trabalhado neste caso, apresentado na Figura 8, é o mesmo que já foi abordado para exemplificar o ajuste linear quando da utilização da técnica de regressão.

Observe-se que o erro cometido em ambas as situações situa-se dentre os mesmos valores, não apresentando diferenças quanto à aderência aos valores históricos.

No entanto os métodos apresentam valores diferentes para extrapolações para além do período atual, sendo que a técnica exponencial privilegia mais os dados ocorridos em períodos mais recentes considerando em escala menor os valores apresentados em passado mais remoto.

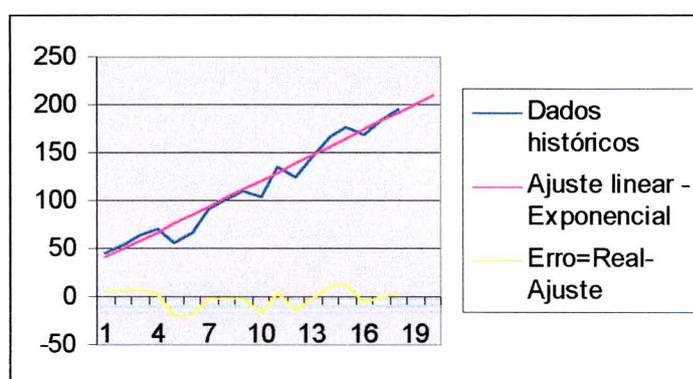


Figura 8

Uma outra forma de efetuar o ajuste utilizando a técnica de ajuste exponencial duplo, é introduzindo uma segunda equação com uma segunda constante de ajuste γ , que não tem necessariamente relação direta com a constante α , conforme era o caso da modelagem anterior, em que existe a relação entre constantes de ajuste $\beta=(1-\alpha)$.

Nesta modelagem existe um par de valores das constantes α e γ , que podem ser obtidos a partir de uma técnica de otimização não-linear, a exemplo do algoritmo de Marquardt, que efetuam o ajuste de forma que o quadrado da soma dos erros (MSE), dado pela equação [2.3.2-7], seja mínimo.

$$MSE = \frac{\sum_{t=1}^N (y_t - S_t)^2}{N} \quad [2.3.2-7]$$

As duas equações associadas a esta modelagem alternativa estão apresentadas em [2.3.2-8] e [2.3.2-9] :

$$S_t = \alpha y_t + (1-\alpha)(S_{t-1} + b_{t-1}) \quad 0 < \alpha \leq 1 \quad [2.3.2-8]$$

$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1} \quad 0 < \gamma \leq 1 \quad [2.3.2-9]$$

A equação [2.3.2-8] ajusta o novo valor a ser prognosticado S_t , diretamente da tendência do período anterior, b_{t-1} , adicionando-o ao último valor ajustado S_{t-1} .

Este procedimento ajuda a eliminar o “gap” entre valores consecutivos, trazendo o novo valor ajustado S_t , mais próximo do valor corrente da série histórica.

A equação [2.3.2-9] determina o valor da tendência da série histórica obtida a partir da diferença entre os dois últimos valores prognosticados S_t e S_{t-1} .

O valor inicial para a variável S_1 normalmente é adotado igual a y_1 e para a variável b_1 podem ser adotados os valores obtidos das fórmulas abaixo:

$$b_1 = y_2 - y_1$$

$$b_1 = \frac{y_n - y_1}{2n}$$

$$b_1 = \frac{(y_2 - y_1) + (y_3 - y_2) + (y_4 - y_3)}{3}$$

Na Figura 9 está representado o ajuste exponencial simples e duplo de uma série temporal, onde se pode observar que o ajuste exponencial duplo apresenta um valor de MSE menor que aquele obtido do ajuste simples.

A projeção de valores além da série histórica apresenta uma maior coerência para o ajuste exponencial duplo, sendo que o ajuste simples projetado conduz a um valor constante igual ao último valor dos dados da série.

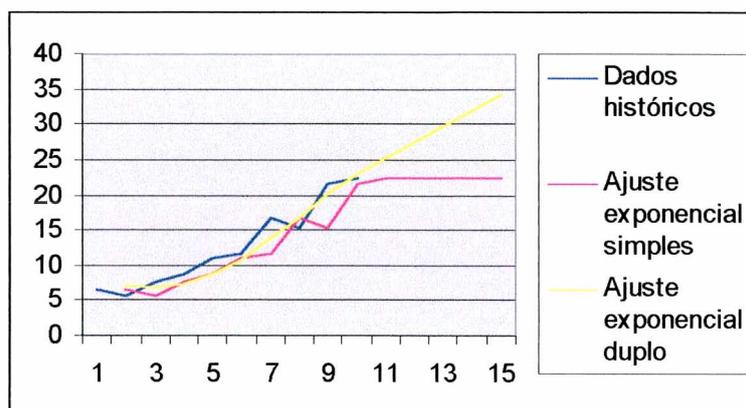


Figura 9

Comparando os valores obtidos, utilizando-se a modelagem de regressão linear, conforme Figura 10, com aqueles calculados utilizando a modelagem exponencial dupla, observa-se que no primeiro caso o crescimento do ajuste se apresenta mais lento demonstrando tratar-se de uma metodologia mais conservadora.

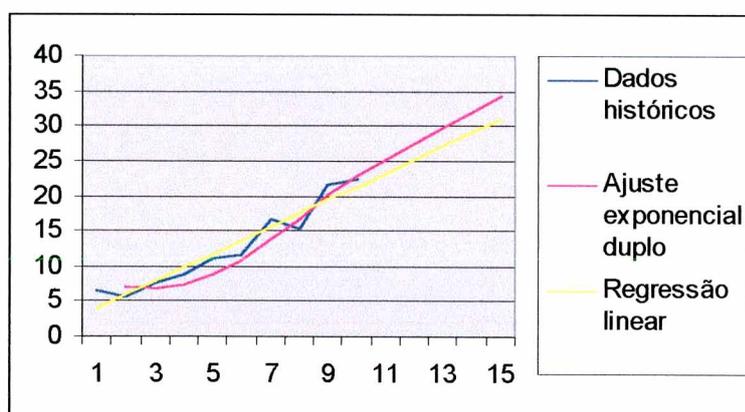


Figura 10

2.3.3 – Ajuste exponencial triplo

Polinômios de qualquer ordem podem ser ajustados pelo ajuste exponencial, enquanto o ajuste linear requer a aplicação sucessiva de dois ajustes, no caso quadrático é necessário um ajuste de terceira ordem juntamente com as duas primeiras ordens do caso linear.

A equação [2.3.3-1] estabelece a terceira ordem a ser utilizada no ajuste quadrático.

$$S_t^3 = \alpha S_t^2 + \beta S_{t-1}^3 \quad [2.3.3-1]$$

O modelo quadrático a ser ajustado toma a forma da equação [2.3.3-2]

$$S_t = a + bt + ct^2 \quad [2.3.3-2]$$

e os coeficientes podem ser estimados pelas equações:

$$a_t = 3S_t - 3S_t^2 + S_t^3 \quad [2.3.3-3]$$

$$b_t = \frac{\alpha}{2\beta^2} \left[(1+5\beta)(S_t) - (2+8\beta)(S_t^2) + (1+3\beta)(S_t^3) \right] \quad [2.3.3-4]$$

$$c_t = \frac{\alpha^2}{2\beta^2} \left[(S_t) - 2(S_t^2) + (S_t^3) \right] \quad [2.3.3-5]$$

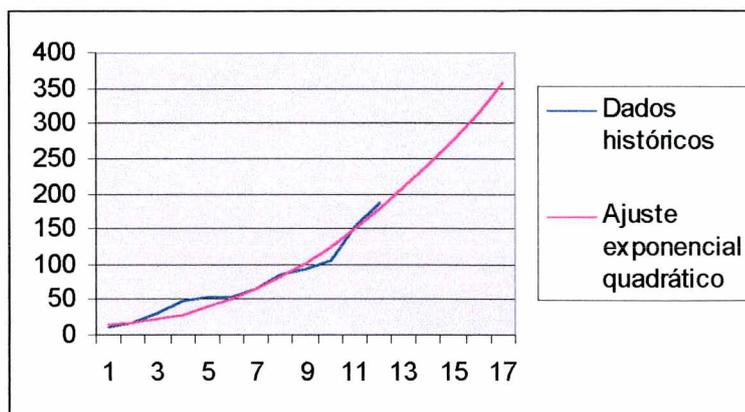


Figura 11

Observando a representação gráfica do ajuste exponencial com modelagem quadrática na Figura 11, verifica-se que esta técnica utiliza todos os dados históricos, mas privilegia os valores mais recentes, sendo que as constantes de ajuste e a equação correspondente somente são adequadas para valores que extrapolam o último período que contribuiu para a formação da modelagem.

A representação para períodos anteriores ao último período participante da modelagem não deve ser utilizada, pois o equacionamento de todo o processo resulta em uma equação de prognóstico para além do último período.

Aparentemente o processo de regressão pode responder adequadamente às necessidades de prognóstico de demanda futura, considerando que pode vir a apresentar resultados satisfatórios na extrapolação dos valores, no entanto deve-se considerar que se está tratando de prognósticos que devem atender aos seguintes requisitos:

- Responder a variações de direção rapidamente.
- Desconsiderar as variações, na série temporal, que não são representativas do fenômeno.
- Utilizar base de dados com uma quantidade muito pequena de amostras.

O que acontece quando os dados históricos apresentam na série valores que indicam uma tendência de crescimento e acompanhados com efeitos de sazonalidade?

Nesse caso o ajuste duplo não apresenta resultados satisfatórios, e o método exponencial com modelagem quadrática, visto anteriormente, não permite representar os efeitos de sazonalidade. Portanto, o ajuste não satisfaz às necessidades do prognóstico.

Nesse caso pode ser aplicado o método denominado de “Holt-Winters”, que basicamente é representado pelas equações:

$$S_t = \alpha \frac{y_t}{I_{t-L}} + (1-\alpha)(S_{t-1} + b_{t-1}) \quad [2.3.3-6]$$

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1} \quad [2.3.3-7]$$

$$I_t = \beta \frac{y_t}{S_t} + (1-\beta)I_{t-L} \quad [2.3.3-8]$$

$$F_{t+m} = (S_t + mb_t)I_{t-L+m} \quad [2.3.3-9]$$

A equação [2.3.3-6] efetua o ajuste exponencial de todos os dados históricos, a equação [2.3.3-7] efetua o ajuste para a tendência de crescimento dos dados, a equação [2.3.3-8] considera os efeitos de sazonalidade da série de dados.

A metodologia utiliza três constantes de ajuste, α , γ e β , que devem ser ajustadas de modo que o ajuste procurado apresente um valor de MSE mínimo.

A equação [2.3.3-9] é a equação que determina o futuro da variável em estudo, ou seja, o prognóstico para um período futuro m , onde está sendo considerado o índice de sazonalidade, retirado da série histórica para efetuar o ajuste.

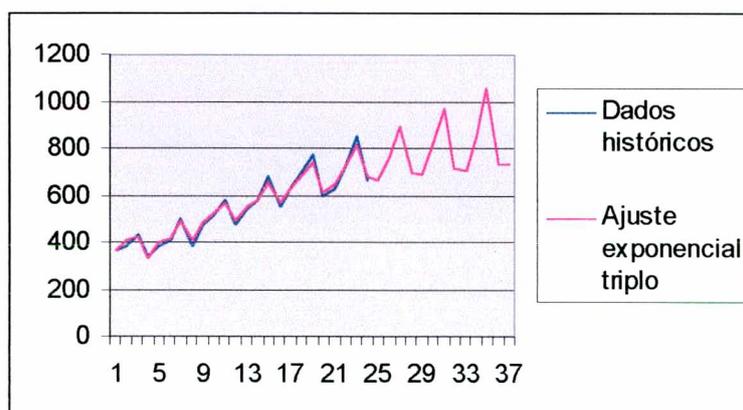


Figura 12

A Figura 12 representa graficamente o ajuste exponencial triplo efetuado em uma série de dados históricos, que apresentam uma tendência de crescimento e experimentam variações sazonais.

A metodologia empregada permite efetuar com bastante precisão o prognóstico de dados futuros, no entanto isto se deve à regularidade dos dados históricos, o que não seria possível, caso estas condições mencionadas não fossem satisfeitas.

2.4 – Prognóstico com sazonalidade

Existem variáveis que apresentam os dados históricos com variações cíclicas e que ocorrem com regularidade mantendo o mesmo espaço de tempo entre ocorrências simultâneas.

A sazonalidade dos dados pressupõe a existência de períodos iguais, dentro dos quais ocorrem as mesmas sequências de variações de valores da série histórica, podendo ou não apresentar tendência de crescimento.

Pela simples observação dos dados tabelados ou apresentados graficamente, é possível verificar a existência da ocorrência de dados sazonais.

Uma sequência lógica para prognosticar séries históricas com valores que apresentam tendência de crescimento, combinados com valores sazonais, pode ser a seguinte:

1. Plotar os dados e determinar visualmente alguma característica evidenciada na série.
2. Se algum fator de crescimento estiver aparente, como por exemplo polinomial ou exponencial, removê-lo matematicamente dos dados.
3. Verificar se nos dados remanescentes está evidenciada a presença de efeitos sazonais significativos.
4. Matematicamente dessazonalize os dados originais verificando o fator de sazonalidade, o que também pode ser observado a partir da plotagem dos dados dessazonalizados.
5. Ajustar os dados utilizando uma metodologia adequada-regressão ou ajuste exponencial por exemplo.
6. O prognóstico de valores futuros consiste na combinação das tendências de crescimento da série e dos valores de sazonalidade, utilizando os valores adequados dos fatores de crescimento e de sazonalidade.

O exemplo mostrado na Figura 12 apresenta uma série histórica com tendência de crescimento e dados sazonais, tratados segundo a sequência indicada anteriormente, apresentando um resultado bastante satisfatório para prognóstico de valores com fator sazonal presente.

Em algumas situações, o fator de sazonalidade pode ser modelado a partir de funções trigonométricas. A tendência de crescimento dos dados históricos podem ser ajustados utilizando o ajuste exponencial, e a análise dos dados cíclicos é efetuada a partir da análise de Fourier.

A equação que ajusta a tendência de crescimento da série, combinada com os dados cíclicos tem a forma da equação [2.4-1].

$$S_t = a + bt + ct^2 + \sum_{i=1}^k \left[(A_i + B_i) \sin \frac{2\pi t}{p_i} + (C_i + D_i) \cos \frac{2\pi t}{p_i} \right] \quad [2.4-1]$$

2.5 – Análise de erro na determinação do prognóstico

Quando são avaliados os diversos métodos para prognósticos de valores de séries históricas, torna-se extremamente necessário o estabelecimento de procedimentos de avaliação de desempenho da metodologia empregada, visando determinar aquele que melhor possa vir a representar os valores obtidos no passado, e conseqüentemente melhores resultados poderá apresentar para valores que extrapolem o tempo presente.

Assim pode-se definir o erro no período t , como a diferença entre o valor atual y_t e o valor ajustado S_t , conforme apresentado em [2.5-1].

$$e_t = y_t - S_t \quad [2.5-1]$$

A soma dos erros [2.5-2]

$$\sum_{t=1}^N e_t = \sum_{t=1}^N (y_t - S_t) \quad [2.5-2]$$

não é uma medida válida da efetividade do método de prognóstico, mas pode ser utilizada como medida de tendência do ajuste.

Normalmente são utilizados os valores indicados nas expressões [2.5-3], [2.5-4] e [2.5-5]

Desvio absoluto médio - MAD

$$MAD = \frac{\sum_{t=1}^N (y_t - S_t)}{N} \quad [2.5-3]$$

Erro médio quadrático – MSE

$$MSE = \frac{\sum_{t=1}^N (y_t - S_t)^2}{N} \quad [2.5-4]$$

Erro médio percentual absoluto - MAP

$$MAP = \frac{100}{N} \sum_{t=1}^N \left| \frac{y_t - S_t}{y_t} \right| \quad [2.5-5]$$

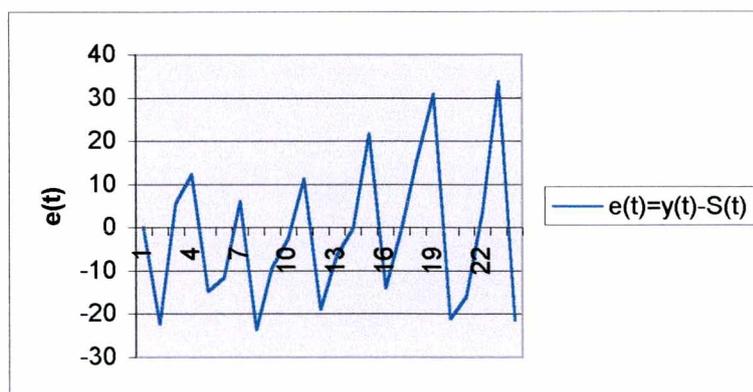


Figura 13

A Figura 13 mostra graficamente o erro apresentado para cada valor da série histórica relativamente ao valor prognosticado pela metodologia de ajuste utilizada, para os valores do exemplo apresentado na Figura 12.

Pode-se observar que apesar de visualmente o ajuste apresentar-se bastante aderente, os valores dos erros indicam discrepâncias significativas, tendo-se obtido para $MAD=13,51$ - $MAP=0,135$ e o erro médio quadrático $MSE=268,03$.

2.6 - A técnica de Box - Jenkins

Uma série temporal de valores históricos que se apresenta na forma de uma sequência de observações do fenômeno, com uma tomada de dados para incrementos iguais no tempo, consistir-se-á de uma série temporal invariável.

Para este tipo de série que não apresenta uma forma definida de crescimento, onde fatores de sazonalidade não podem ser observados, os métodos vistos até agora não apresentam resultados satisfatórios, e que possam ser utilizados para extrapolação de dados futuros.

O método de Box-Jenkins, apresentado em 1970, poderá representar uma metodologia que permita prognosticar séries temporais desta natureza, sendo que a sua abordagem é de efetuar um ajuste com modelagem estocástico (probabilístico) em séries estacionárias (sem tendência definida), constituindo-se em três estágios principais:

Identificação: Neste estágio é investigada a estacionaridade por meio de uma função de autocorrelação, e é efetuada uma tentativa de avaliação para a construção de ordem de um modelo.

Estimação : Este estágio estima os parâmetros do modelo. A melhor maneira de estimar é através da utilização de procedimentos de otimização não linear.

Diagnóstico: Esta fase examina os resíduos ocasionais e a periodicidade. Os erros padronizados dos estimadores são estudados para evitar valores sobreestimados.

Se tudo estiver bem, ao final da aplicação dos três estágios, o modelo de equação resultante pode ser utilizado para prognóstico.

O prognóstico apresenta uma qualidade, seus erros médios quadráticos são mínimos, e estes erros são utilizados para produzir intervalos de confiança para o prognóstico.

A forma geral destas séries é dada por:

$$y_t = K + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad [2.6-1]$$

onde :

ϕ_i = são parâmetros de ajuste;

K = constante

ε_t = componente de erro aleatório

O modelo apresentado na equação [2.6-1] é denominado de *auto-regressivo (AR)*, pelo fato de utilizar-se de valores para o ajuste, anteriores ao valor a ser prognosticado.

Se houver necessidade de predizer p valores de Φ_i , o modelo auto-regressivo é classificado como sendo modelo AR(p).

A partir da equação do erro médio quadrático [2.6-2], aplicado à equação [2.6-1], podem ser obtidos os coeficientes K e Φ_i , efetuando-se a derivada parcial da equação resultante em relação a cada variável a determinar.

$$\sum_{t=1}^N e_t = \sum_{t=1}^N \left(y_t - K - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p} \right)^2 \quad [2.6-2]$$

Uma segunda categoria da modelagem Box-Jenkins é da forma da equação [2.6-3]:

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad [2.6-3]$$

onde:

μ = Média do modelo.

θ_i = Série de parâmetros de ajuste.

ε_t = Componente de erro ocasional no tempo t , também denominados de *impacto randômico* ou *ruído branco*, tipicamente originados da distribuição normal, apresentando média= 0 e variância σ^2 finita.

Esta modelagem é denominada de *média móvel* MA e o valor de y_t é baseado em uma sequência de componentes randômicos, certamente uma violação da hipótese de regressão.

O objetivo da análise pela metodologia de Box-Jenkins é estimar os valores de ϕ 's e θ 's, sendo que o número necessário de estimadores são definidos durante o estágio de identificação.

Quando na modelagem estão presentes ambos os parâmetros de ajuste ϕ 's e θ 's, conforme indicado na equação geral [2.6-4], o modelo é denominado de modelo ARMA- modelo *auto-regressivo-média móvel*, de ordem p, q , onde p significa que existem $p\phi$ termos e q significa que existem $q\theta$.

$$y_t = K + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad [2.6-4]$$

A metodologia de Box-Jenkins é essencialmente utilizada para determinar o modelo probabilístico da série temporal. Dessa forma, é também denominada de modelagem estocástica, e em particular tem a capacidade de analisar processos estacionários, que têm a propriedade de apresentarem variações na média μ e da variância σ^2 com o tempo.

A prova efetuada através de uma função de autocorrelação tem importância fundamental para o estágio de identificação de procedimentos da modelagem Box-Jenkins.

A função de autocorrelação calcula a correlação entre uma sucessão de valores observados e que foram medidos ao longo do tempo, e uma mesma sucessão de valores observados em períodos anteriores.

A forma da função de autocorrelação define o modelo a ser utilizado para o ajuste através da modelagem Box-Jenkins, além de determinar sazonalidade na série temporal e o período de ocorrência.

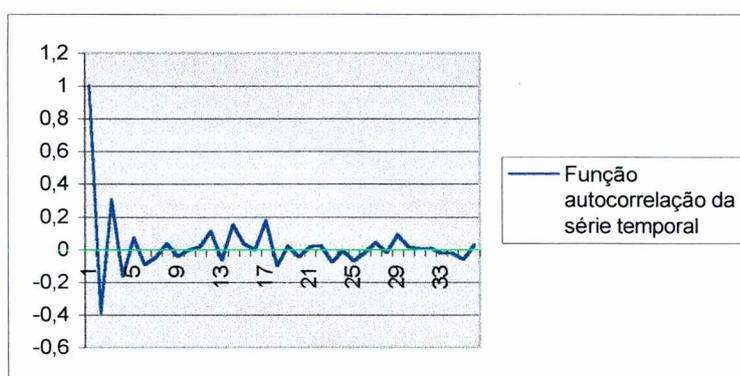


Figura 14

A Figura 14 apresenta graficamente a função autocorrelação para os 35 períodos iniciais da série histórica da Figura 15, onde a forma predominante é alternada com tendência ao valor zero, indicando a utilização de um ajuste na modelagem Box-Jenkins na forma do modelo autoregressivo (AR).

A Figura 15, corresponde ao ajuste efetuado com ordem $p=2$, e limites inferior e superior, entre os quais é esperada a ocorrência do evento com um grau de confiança de 90%.

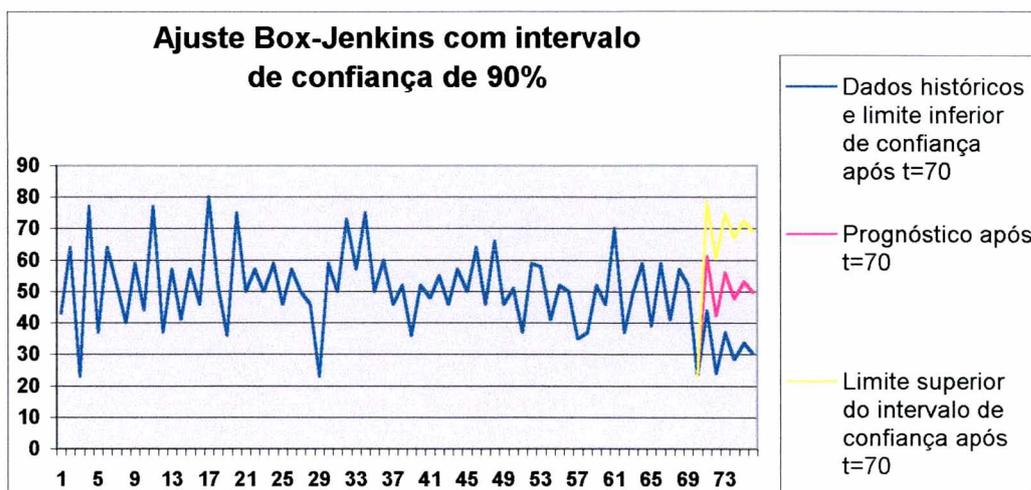


Figura 15

A análise de séries temporais em presença de valores crescentes, indicando uma tendência, além de componente de sazonalidade, exige a transformação da modelagem dos valores históricos a partir da utilização de uma função, por exemplo logaritmos naturais adotados no caso das Figuras 16 e 17, de forma a tornar o resultado do teste de autocorrelação mais próximo do valor nulo.

No caso exemplo, na Figura 16, o teste de autocorrelação com sazonalidade foi efetuado após a transformação por logaritmos naturais e está a indicar a existência de componentes de sazonalidade, que devem ser identificados para a análise Box-Jenkins.

Com a remoção das componentes de sazonalidade, o aspecto gráfico da análise de autocorrelação, indicado na Figura 16, mostra uma redução significativa na correlação entre eventos, identificando a ordem do modelo para a transformação em funções matemáticas definidas, bem como a definição de termos correspondentes à sazonalidade.

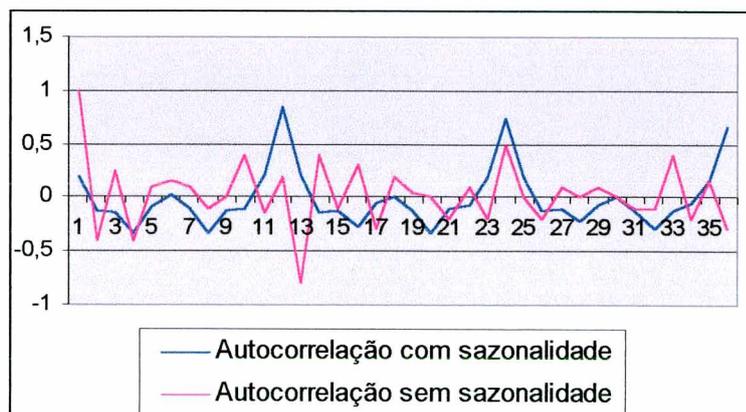


Figura 16

Com a identificação da modelagem a ser utilizada, a partir do teste de autocorrelação, o resultado da análise Box-Jenkins está indicado na Figura 17, onde está representado o prognóstico para tempos futuros e os limites inferior e superior do intervalo de confiança de 90%.

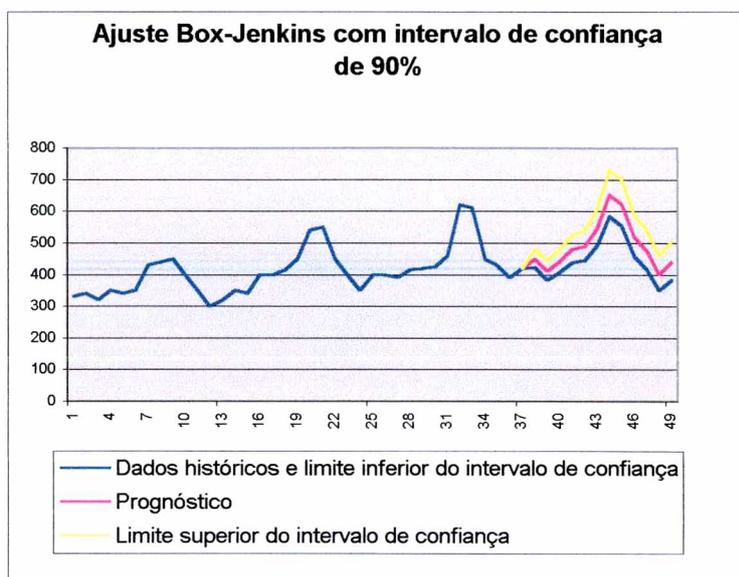


Figura 17

2.6.1-Séries temporais multivariáveis pela técnica Box-Jenkins

Se cada série temporal em observação é um vetor de números, a solução pode ser obtida a partir da modelagem Box-Jenkins na forma de multivariáveis.

A equação [2.6.1-1] apresenta a forma matricial da modelagem Box-Jenkins multivariável, onde os termos têm o mesmo significado já descritos quando da apresentação da modelagem com variável única.

$$\begin{aligned}
 [y_t] = & [K] + [\phi_1][y_{t-1}] + [\phi_2][y_{t-2}] + \dots + [\phi_p][y_{t-p}] + \\
 & + [\mu] + [\varepsilon_t] - [\theta_1][\varepsilon_{t-1}] - [\theta_2][\varepsilon_{t-2}] - \dots - [\theta_q][\varepsilon_{t-q}]
 \end{aligned}
 \tag{2.6.1-1}$$

Na equação [2.6.1-1], $[y_t]$, $[K]$, $[\mu]$ e $[\varepsilon_t]$, são vetores coluna de ordem $n \times 1$, e os vetores $[\phi_k]$ e $[\theta_k]$ são de ordem $n \times n$, correspondendo aos parâmetros do ajuste auto-regressivo e do ajuste de média móvel, conforme equações [2.6.1-2] e [2.6.1-3].

$$[\phi_k]_1^p = \begin{bmatrix} \phi_{k,11} & \phi_{k,12} & \dots & \phi_{k,1j} \\ \phi_{k,21} & \phi_{k,22} & \dots & \phi_{k,2j} \\ \dots & \dots & \dots & \dots \\ \phi_{k,i1} & \phi_{k,i2} & \dots & \phi_{k,ij} \end{bmatrix} \quad [2.6.1-2]$$

$$[\theta_k]_1^q = \begin{bmatrix} \theta_{k,11} & \theta_{k,12} & \dots & \theta_{k,1j} \\ \theta_{k,21} & \theta_{k,22} & \dots & \theta_{k,2j} \\ \dots & \dots & \dots & \dots \\ \theta_{k,i1} & \theta_{k,i2} & \dots & \theta_{k,ij} \end{bmatrix} \quad [2.6.1-3]$$

Seja o caso exemplo, em que está sendo considerada a composição de ar com gás metano em um forno com a finalidade de obter a mistura dos gases, onde existe um conteúdo de dióxido de carbono (CO^2). A alimentação com gás metano constitui a série temporal de entrada $X(t)$, e a concentração de CO^2 foi considerada como a série temporal de saída $Y(t)$.

Sucessivos valores de $X(t)$ e $Y(t)$ foram observados e anotados, constituindo as séries históricas para utilização da metodologia de ajuste com 2 variáveis a prognosticar.

O equacionamento com duas séries temporais e com ordem dois apresenta-se na forma da equação [2.6.1-4], onde x_t e y_t são os valores a serem obtidos do ajuste, para prognósticos futuros, das variáveis como definidas anteriormente.

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \phi_{1.11} & \phi_{1.12} \\ \phi_{1.21} & \phi_{1.22} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \phi_{2.11} & \phi_{2.12} \\ \phi_{2.21} & \phi_{2.22} \end{bmatrix} \begin{bmatrix} x_{t-2} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad [2.6.1-4]$$

Com a determinação dos coeficientes de ajuste, é necessária a análise estatística do resíduo ε_t , com a finalidade de verificar o grau de afinidade do modelo, com os dados das séries históricas.

As matrizes a serem analisadas da estatística do resíduo estão relacionadas com a média – que deve ser nula para os resíduos, a covariância – que indica o grau de dispersão dos resíduos, e a

correlação – que determina a influência de uma variável sobre a outra, devendo apresentar valores unitários na diagonal principal.

Para o caso teste apresentado, a representação gráfica das variáveis e o prognóstico com limites de confiança de 90% estão indicados nas Figuras 18 e 19.

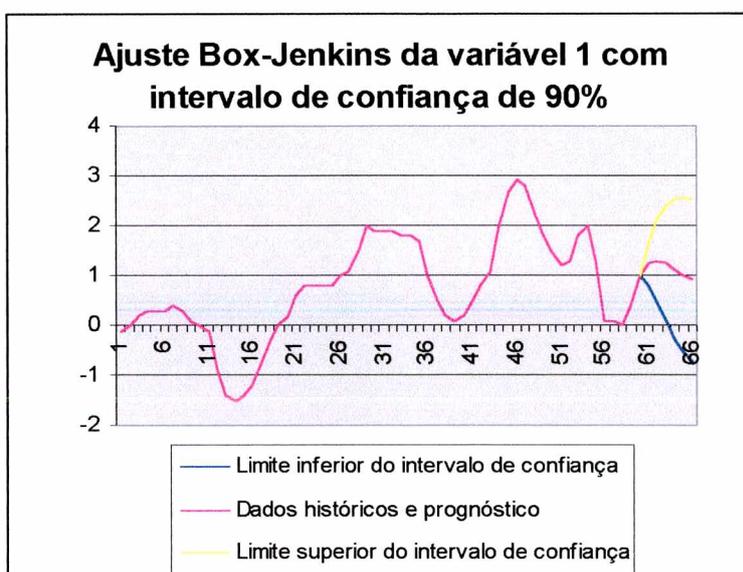


Figura 18

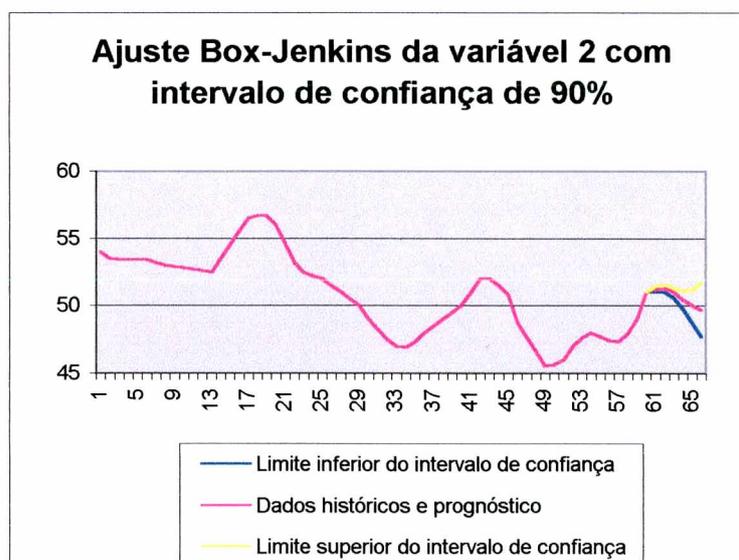


Figura 19

Capítulo III. Redes Neurais

3.1 - Introdução

A arquitetura geral de uma rede neural é mostrada na figura 20. Existem uma ou mais entradas que são denominadas 'neurônios de entrada' que, na verdade, simplesmente apresentam na saída um valor igual ao apresentado. Esses neurônios hipotéticos não executam nenhum processamento não passando de um construtor semântico usado para representar uma amostra das energias presentes.

A arquitetura provê a existência de um ou mais neurônios de saída, geralmente muito poucos, que representam a resposta da rede aos dados do mundo. Diferentemente dos de entrada, os neurônios de saída são bastante reais. Cada neurônio de saída aceita entradas que tanto podem advir de outros neurônios como diretamente do mundo exterior, processando essas entradas e produzindo uma saída.

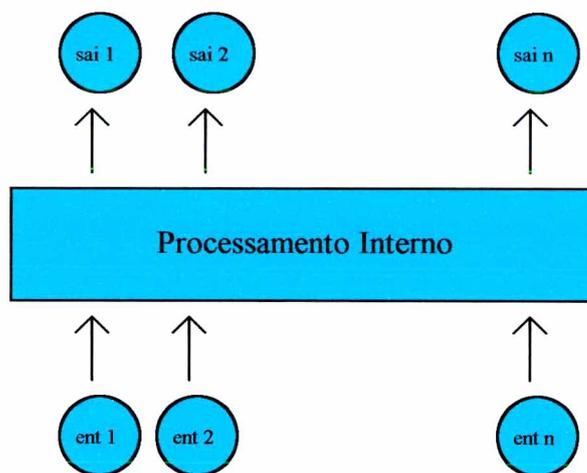


Figura 20 – Exemplo de rede neural

A caixa preta define o tipo de rede. Alguns modelos conectam diretamente as entradas às saídas, omitindo-as. Todo o processamento, nesse caso, é executado pelos neurônios de saída. Outros modelos permitem que os neurônios de saída se conectem uns com os outros da mesma forma que com os pertencentes à camada anterior e, finalmente, em alguns casos, não se distinguem neurônios de entrada e de saída. Através de um processo de relaxamento de

rede, passam de neurônios primários no início do processo quando os dados do mundo lhes são apresentados a neurônios de saída após a estabilização da rede.

Um *neurônio artificial* é, essencialmente, um conjunto de entradas e uma função que determina uma saída, conforme mostrado na Figura 21. O conjunto de entradas pode representar informações vindas do ambiente ou saídas de outros neurônios. Cada entrada tem um peso associado, que corresponde à força de cada estímulo em relação à saída do neurônio.

O conjunto de estímulos de entrada, rotulado como x_1, x_2, \dots, x_n é aplicado ao neurônio. Pode-se denotar este conjunto como um vetor, escrevendo: X . Os pesos associados a cada entrada podem ser denotados como w_1, w_2, \dots, w_n ou como um vetor W .

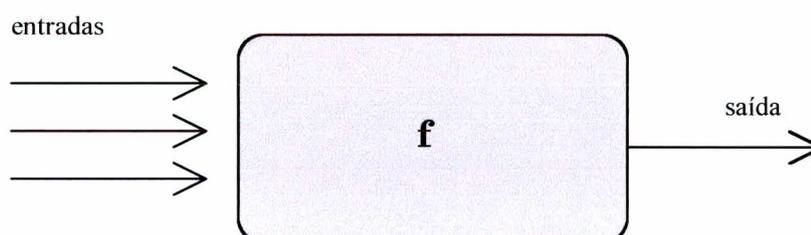


Figura 21- Neurônio artificial

Em muitos modelos, o nível de excitação do neurônio corresponde à soma ponderada dos valores de entrada, ou seja, o somatório dos pesos multiplicados pelas entradas. Se este valor for denotado por Z , pode-se escrever a seguinte fórmula:

$$Z = W \cdot X = w_1x_1 + w_2x_2 + \dots + w_nx_n$$

O valor Z é freqüentemente processado por uma função f , para produzir a saída correspondente do neurônio. Um simples neurônio é capaz de desempenhar reconhecimento de padrões simples. Mas o poder de computação é grandemente incrementado quando ligados em rede.

O conjunto de entradas X tem cada um de seus elementos conectados a um neurônio artificial com um peso específico w_{ji} , onde i denota o elemento de entrada e j o neurônio artificial. É

conveniente considerar os pesos como sendo uma matriz W . A matriz teria n linhas por m colunas, onde n é o número de entradas e m é o número de neurônios. Dessa forma, o vetor de níveis de excitação Z seria formado a partir de uma simples multiplicação de matrizes: $Z = X \times W$, onde Z e X são vetores linha.

O vetor de resultados da rede Y é formado a partir da aplicação da função f a cada um dos elementos de Z .

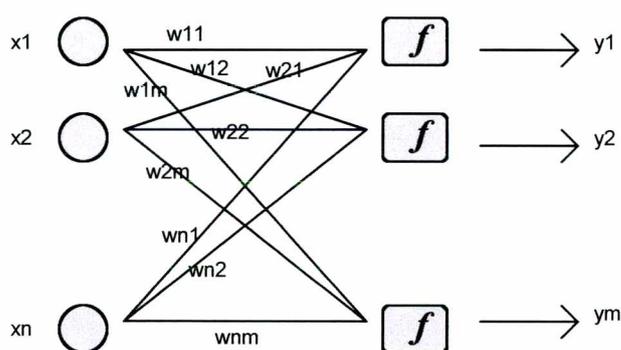


Figura 22 – Exemplo simples de rede neural

Uma rede neural é usualmente treinada de duas maneiras distintas: **supervisionado** e **não supervisionado**. No treinamento supervisionado muitas amostras, que vão servir como exemplares, são coletadas, definindo entradas e saídas que devem ser obtidas quando essas entradas forem apresentadas à rede. Os pesos que conectam cada neurônio na rede são ajustados para minimizar o erro entre a saída desejada e a saída obtida na rede. No treinamento não supervisionado assume-se que cada entrada se origina de uma classe pertencente a um conjunto de classes e que a saída da rede identifica a que classe pertence o exemplar de entrada. Existe um terceiro tipo de treinamento, denominado '**reforço de aprendizagem**' que não é supervisionado na medida em que as saídas desejadas não são especificadas a priori. Ao mesmo tempo é supervisionada, na medida em que a rede é informada se a resposta que forneceu a um determinado padrão de entrada é 'boa' ou 'má' (Fausett 1994; Kartalopoulos 1996)

Depois que a rede neural tiver sido treinada, ela deve ser testada com novos dados não utilizados no treinamento. O procedimento usual é o de se separar casos conhecidos em dois

conjuntos disjuntos. Usa-se o primeiro como conjunto de treinamento e o segundo como conjunto de validação (Shaw e Simões 1999).

3.2 - Aplicações das Redes Neurais

3.2.1 - Redes neurais utilizadas para classificação

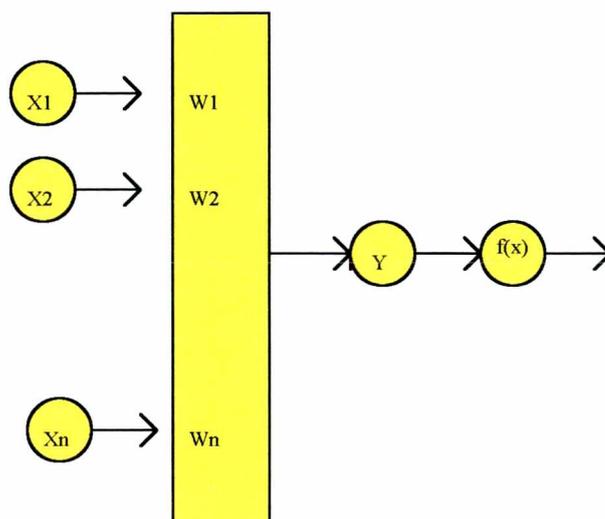
O problema mais simples de classificação é aquele que envolve uma decisão binária, o de se saber, por exemplo, se um copo de cristal, conduzido por uma esteira dentro da fábrica, deve ser aceito ou rejeitado. A escolha entre duas classes não é uma decisão binária posto que o elemento pode pertencer a classe A, a classe B ou a nenhuma das classes.

Um perceptron que decide se uma entrada pertence ou não pertence a uma classe é mostrado na Figura 23. Não é necessário mais do que um neurônio de saída para se representar essa decisão. A rede é treinada para produzir um nível de ativação alto para uma decisão e baixo para a outra.

Quando a rede treinada é testada com uma amostra desconhecida, a decisão é tomada em função da saída obtida que, devido à função logística $f(x)$, nunca é 0.0 ou 1.0. Se o resultado está acima de um limiar diz-se que um nível alto de ativação foi obtido e faz-se a escolha correspondente (Kartalopoulos 1996).

Uma técnica para se analisar o comportamento das redes consiste em traçar as regiões de decisão criadas no espaço multidimensional pelas variáveis de entrada. Estas regiões de decisão especificam quais valores de entrada resultam numa decisão ou em outra.

O perceptron de 1 camada forma duas regiões de decisão separados por um hiperplano. Rosenblatt provou que se as entradas apresentadas são linearmente separáveis, então o procedimento para treinar uma rede de uma única camada converge e coloca um hiperplano entre as duas decisões. Um problema com tal procedimento ocorre quando as entradas não são mais linearmente separáveis, como é o caso do problema lógico associado ao 'ou exclusivo'.



$$Y = \sum X_i W_i \text{ e } f(x) \text{ geralmente tem a forma de } 1/(1+e^{-x})$$

Figura 23- Rede classificadora binária

Um modelo que permite classificação múltipla não permite apenas que se classifique as entradas nas diversas alternativas possíveis, mas possibilita, ainda, uma tomada de decisão que implica na 'rejeição', pois não pertence a nenhuma das classes. A melhor maneira de implementar tais modelos consiste em destinar um neurônio de saída para cada classe. Uma saída alta implica que a classe está presente. O treinamento é tal que, quando isso acontece, os outros neurônios de saída apresentam baixo nível de ativação. Se todos os neurônios estiverem 'off' isso é interpretado como uma 'rejeição'.

Quando casos desconhecidos são apresentados à rede treinada, a regra de decisão implementada computacionalmente é tal que: Se todas as saídas são baixas, a amostra é rejeitada e caso contrário atribui-se a amostra à classe que apresentou maior nível de ativação. Em caso de empate escolhe-se de forma aleatória.

Existem problemas que sugerem uma outra forma de codificação onde, por exemplo, as categorias têm, entre si, alguma forma de relação de ordem (ruim, bom, ótimo) ou quando a classe é definida por mais de um aspecto (tipo de sangue). No caso de classificação sanguínea, por exemplo, é mais lógico se ter três neurônios associados aos três fatores que compõem a classificação, ou seja, um neurônio que diz se o sangue é do tipo A, outro se é do tipo B e um terceiro que determina o fator Rh.

Uma rede que queira 'ouvir um som' e, a partir daí, decidir se a articulação inicial é palatal, se é articulado a partir da língua ou dos lábios, se a continuação é nasal ou não, poderia ser

codificada com quatro neurônios de saída em vez de seis, três para a articulação e um para a continuação.

3.2.2 - Redes neurais utilizadas para auto-associação

Quando uma rede neural tem exatamente tantos neurônios de entrada quanto os de saída e a rede é treinada para reproduzir na saída os dados apresentados na entrada, diz-se que essa rede é auto-associativa. A principal utilização desse tipo de redes é na filtragem de erros e em completar padrões, por exemplo, uma foto ruim ou uma comunicação incompleta (Fausett 1994).

Uma das virtudes desse tipo de redes é que é possível, a partir de uma amostra com exemplos imperfeitos de um determinado padrão, aprender a construir os padrões protótipos.

3.2.3 - Redes neurais utilizadas para predições de séries temporais

A predição de valores futuros de uma série temporal é um problema extremamente comum. Qualquer rede neural que seja capaz de aceitar vetores com valores reais como entrada e produzir valores reais como saída pode ser usado para esse tipo de predições.

A Figura 24 demonstra o uso de n pontos contíguos numa série temporal para predizer o valor associado ao ponto $(n+1)$. Emprega-se uma série de treinamento de m pontos $m \gg n$ e, a partir dessa, geram-se um grande número de amostras de $(n + 1)$ elementos. Cada amostra para o modelo consiste de um valor atual e $(n-1)$ valores históricos. A rede é ajustada para que o valor de saída coincida com o valor $(n+1)$. Um aspecto vital no processamento de séries temporais consiste na necessidade de se eliminar tendências e variações sazonais. Eliminando tendências que são facilmente predizíveis e podem ser adicionadas posteriormente, liberta-se a rede para um processamento em maior profundidade. A sazonalidade pode provocar uma concentração em aspectos lineares quando o que realmente se pretende é explorar a característica não-linear das redes neurais.

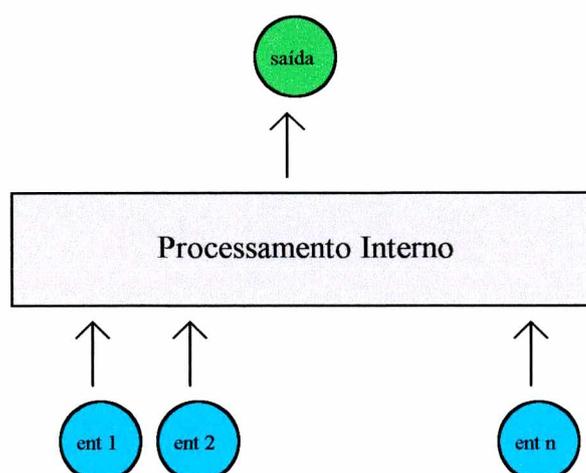


Figura 24 - Exemplo de uma rede simples para previsão

Os sistemas mais simples capazes de previsão se utilizam de dados tirados de uma única série temporal. Isto pode ser estendido utilizando-se dados de diversas séries temporais diferentes. Predições múltiplas são possíveis; mas os resultados obtidos, questionáveis.

Uma empresa de energia elétrica, por exemplo, pode estar interessada em saber qual a potência instalada que será necessária para atender o consumo do próximo ano. Para isso se utiliza de uma série temporal que contém os padrões de consumo das últimas décadas. Outro problema, dentro de uma sala de controle, é saber qual a potência que será necessária na próxima hora. Se for utilizada a mesma série para ambas as previsões obter-se-á, obviamente, melhores resultados para as previsões anuais do que para as previsões horárias. Geralmente preferem-se séries temporais independentes.

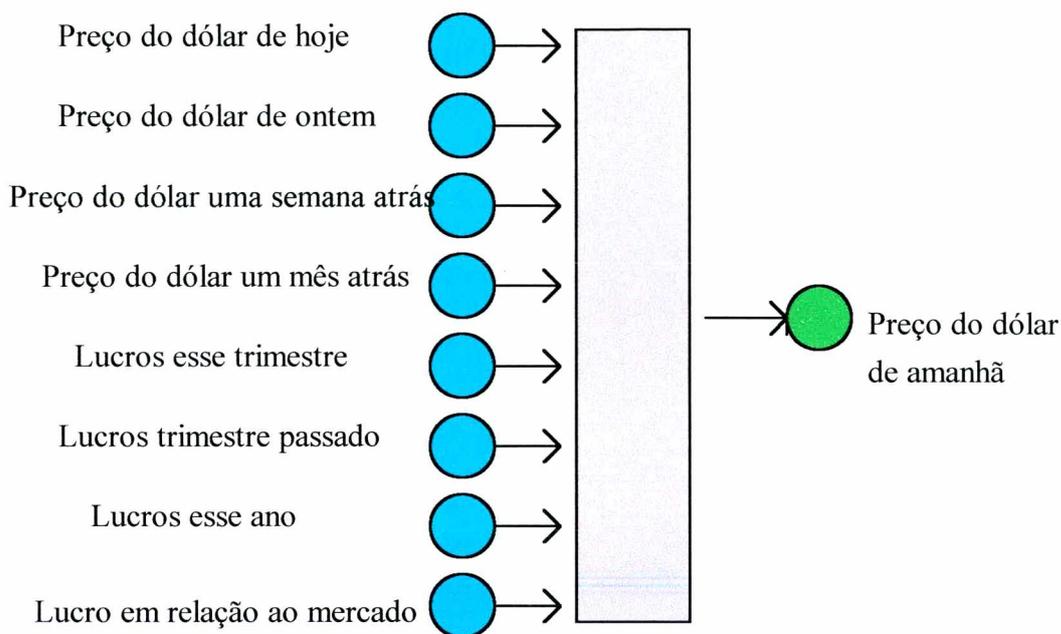


Figura 25 - Uma rede neural para previsão do preço do dólar

3.2.4 - Redes neurais utilizadas para aproximações de funções

Em essência, redes neurais são aproximadoras de funções que mapeiam valores na entrada a outros que aparecem na saída do dispositivo. Classificação, associação e previsões dentro de séries temporais são formas especiais de aproximações de funções. O que se enfoca, no entanto, sob esse tópico, consiste na aproximação de funções que têm como domínio e contra domínio o conjunto dos vetores de valores reais.

O problema a ser resolvido consiste em: dados alguns pontos pertencentes a uma função desconhecida, calcular o resultado dessa função em um ponto qualquer. Testes empregando a função de gauss, um exemplo clássico, mostram que as redes neurais se constituem numa excelente arma para resolução desse tipo de problema.

Suponha-se que haja um vetor real X que descreve uma determinada condição, um tipo de tremor de terra por exemplo, ou uma estrutura atômica ou uma forma sólida. Por causa deste contexto X um resultado Y é produzido. Y pode ser o registro de uma onda de choque, uma difração de raio ou um eco de radar.

Assumindo-se que se tem um modelo m que nos permite prever Y para qualquer vetor X que se apresenta:

$$y = m(x)$$

A técnica consiste na coleta de um conjunto representativo de vetores X e em aplicar o modelo a esse conjunto. Se se está diante de um modelo estocástico tal que Y possa não ser unicamente determinado para um dado X , deve-se replicar cada X até gerar múltiplas amostras de Y . Os vetores X e Y obtidos são então empregados como conjunto de treinamento de uma rede neural.

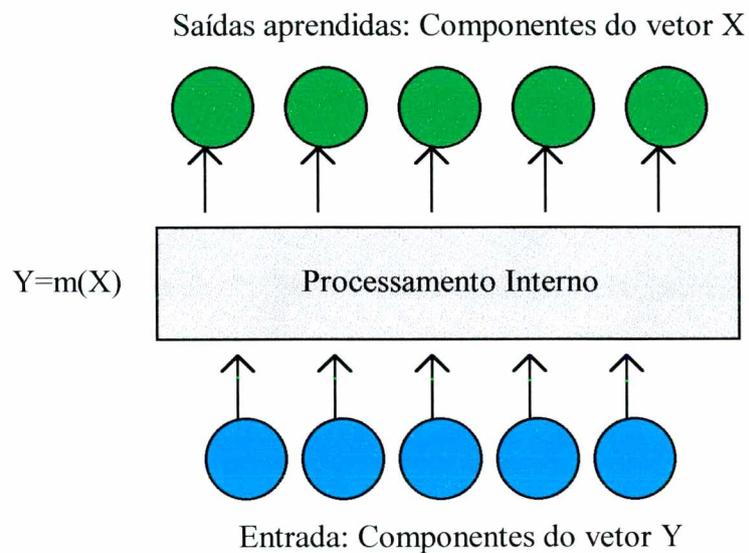


Figura 26 - Rede neural para aproximação de funções

A regressão múltipla é uma técnica estatística padrão pela qual uma ou mais variáveis independentes são usadas para prever uma única variável dependente. Por exemplo, caso tenha-se n variáveis independentes, X_1, X_2, \dots, X_n , e uma variável dependente Y , tem-se que:

$$y = a_0 + \sum_{i=1}^n a_i x_i$$

O problema da técnica de regressão é que quanto mais afastado dos paradigmas de linearidade, maior o erro que se comete. Redes neurais de múltiplas camadas se mostram excelentes como técnicas de caos que são, para lidar com relações complexas não-lineares. Se qualquer das não-linearidades é conhecida, redes neurais com enlace funcional podem ser empregadas para melhorar a aprendizagem.

3.3 - Redes neuronais de múltiplas camadas sem retroalimentação

Uma rede de múltiplas camadas é uma rede com uma ou mais camadas de neurônios entre a entrada e saída de rede. Essas camadas adicionais são denominadas camadas ocultas. O termo sem retroalimentação significa que a informação flui em apenas uma direção.

Um perceptron de quatro camadas com três neurônios nas camadas ocultas é mostrado na Figura 27. Perceptrons de múltiplas camadas resolvem várias limitações dos perceptrons de uma única camada, não sendo usados no passado pela falta de algoritmos eficazes para o treinamento da rede. Ainda não se provou que estes algoritmos convergem como no processo de treinamento de perceptrons de uma camada. Apesar disso, redes desse tipo têm-se mostrado bastante poderosas na solução da maioria dos problemas.

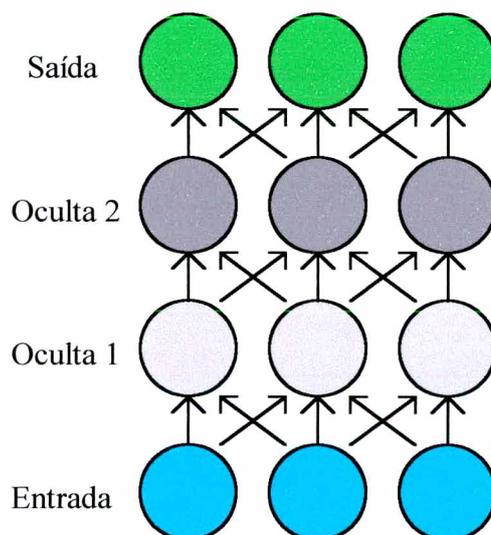


Figura 27 - Rede neural de quatro camadas sem retroalimentação

Os neurônios da primeira camada são 'hipotéticos' na medida em que a entrada é definida pela entrada da rede e que não executam nenhum tipo de processamento. Cada neurônio das demais camadas da rede funciona de acordo com a fórmula $saída = f(\text{total das entradas})$.

Essas entradas vêm das camadas anteriores e suas saídas vão alimentar a entrada da camada seguinte.

O comportamento da rede é determinado pelos pesos que conectam os neurônios de uma camada com os da camada seguinte e que podem ser associados à força ou intensidade das ligações sinápticas, numa metáfora biológica:

$$saida = funcao \left(total \ das \ entradas \right) = f \left(\sum_{i=0}^{n-1} x_i w_i + w_n \right)$$

A função de ativação é não-linear e, quando aplicada ao somatório ponderado das entradas (vide fórmula), determina a saída daquele neurônio. O domínio destas funções é ilimitado e o contradomínio tal que limita a saída dos neurônios a um intervalo $[0 -1]$ ou, às vezes, $[-1 +1]$. A maioria dos modelos usa uma *sigmóide* (função que tem a forma de um s) como função de ativação. A sigmóide mais comumente usada é a *função logística*, $f(x) = 1 / (1 + e^{-x})$. Usualmente a mesma função de ativação é aplicada para todos os neurônios embora isso não seja um requisito. Muitas aplicações usam, para os neurônios de saída, a *função identidade* $f(x) = x$.

O fato de se usar funções de ativação não lineares resulta de que o emprego de funções lineares faz com que a rede de múltiplas camadas possa ser reduzida a outra equivalente de uma camada e, portanto, não resolvendo os problemas típicos de redes de uma camada, como por exemplo os que tentam resolver problemas que não apresentam entradas linearmente separáveis.

Por analogia com sistemas eletrônicos analógicos pode-se pensar na função de ativação como um mecanismo que define um ganho não linear para o neurônio artificial. Esse ganho é calculado encontrando a taxa de mudança que ocorre na saída quando se tem uma pequena variação da soma ponderada das excitações presentes na entrada podendo-se mostrar que o ganho é dado pela inclinação da curva da função para um nível específico de excitação.

O ganho fornece o grau de certeza ou grau de reconhecimento de um padrão. Ele varia de um valor baixo para excitações grandes e negativas até um valor alto para uma excitação nula, caindo novamente quando a excitação se torna muito grande e positiva.

Grosberg (1973) mostra que esta característica não-linear do ganho resolve o dilema apresentado pelo ruído de saturação da rede, isto é, como pode uma mesma rede manusear sinais pequenos e grandes posto que pequenos sinais de entrada exigem um alto ganho através da rede se se pretende que eles produzam saídas utilizáveis, e se esse ganho fosse aplicado também aos sinais maiores um limite de saturação seria ultrapassado.

Pequenos sinais de entrada têm baixa capacidade excitatória, exigindo um fator de ampliação que é dado pelo ganho para produzir saídas úteis, isto é, que contribuam no sentido da convergência da rede. A função logística satisfaz muito bem esta exigência, pois para grandes excitações, positivas ou negativas, apresenta um fator baixo de ampliação. Dessa maneira, um neurônio é capaz de trabalhar sobre um domínio amplo de níveis de entrada.

O treinamento de uma rede é, usualmente, um verdadeiro pesadelo devido a não linearidade do modelo onde um grande número de variáveis correlacionadas torna o problema de otimização extremamente complexo, com falsos mínimos ocorrendo no leito de vales estreitos. Um algoritmo para o treinamento de redes de múltiplas camadas que minimiza o problema é intitulado de *back-propagation* (Rumelhart, 1986). É um algoritmo supervisionado e requer o pareamento de cada vetor de entrada com um vetor de saída desejado.

O processo de treinamento começa fixando-se todos os pesos com valores próximos de zero que, freqüentemente, são gerados aleatoriamente. A partir daí um subconjunto da coleção de amostras de treinamento é apresentada à rede, um de cada vez. Mede-se o erro e atualizam-se os pesos de forma a reduzi-lo, repetindo-se o processo se necessário. Cada passagem por este processo é denominado uma época. O tamanho do subconjunto (número de amostras de treinamento usadas para atualizar os pesos) é chamado de tamanho da época.

A medida de erro mais comum é o erro médio quadrático que é facilmente computável e eficiente, permitindo, ainda, que se calcule, explicitamente, as derivadas parciais com relação aos pesos individuais. O uso do logaritmo do erro médio quadrado como medida melhora, ainda mais, a aprendizagem. Suponha-se que se tem um padrão de treinamento que se está

processando. Quer-se ajustar a rede para uma saída t_{pj} e observa-se uma saída o_{pj} . Se temos n neurônios de saída o erro para uma única apresentação é dado por:

$$E_p = \left(\frac{1}{n}\right) \sum_{j=0}^{n-1} (t_{pj} - o_{pj})$$

Se tem-se o_j como a saída de um neurônio de uma camada anterior; $rede_j$ a soma ponderada vinda através de um neurônio j da camada de saída; o_j a ativação observada e t_j aquela que se desejaria, tem-se:

$$\frac{\partial E}{\partial w_{ij}} = -o_i \delta_j \quad \text{com} \quad \delta_j = f'(rede_j)(t_j - o_j)$$

Se tem-se m apresentações na época, o erro, para a época, é dado por:

$$E_p = \left(\frac{1}{m}\right) \sum_{p=0}^{m-1} E_p$$

O termo 'back propagation' (propagação para trás), se deriva do fato que, para ajustar os pesos, primeiro é necessário calcular para a camada de saída e ir voltando até a camada de entrada. A partir dos valores conhecidos dos δ das camadas mais externas (k), se computam os da camada anterior (j). Na fórmula os w_{kj} são os pesos que conectam um neurônio da camada j com outro da camada seguinte ' k '.

$$\delta_j = f'(rede_j) \sum_k (\delta_k w_{kj})$$

$$\frac{\delta E}{\delta w_{ji}} = -o_i \delta_j$$

Essas fórmulas são para uma única apresentação de um padrão de entrada. Para computar o gradiente de uma época inteira de treinamento, como a derivada da soma é a soma das derivadas, basta somar os gradientes para cada soma.

O algoritmo de retropropagação utilizado para o projeto de redes neurais com múltiplas camadas (sob supervisão), aplica uma técnica recursiva conhecida na estatística como *aproximação estocástica*.

3.4 – Redes de função de base radial

Enfocando o projeto de redes neurais como um *problema de ajuste de curva (aproximação)*, em um espaço de alta dimensionalidade, ter-se-á desse ponto de vista que o ato de aprender é equivalente a encontrar uma superfície, em um espaço multidimensional, que forneça o melhor ajuste para os dados de treinamento, com o critério de “melhor ajuste” sendo medido em um sentido estatístico. A generalização da metodologia equivale na utilização desta superfície multidimensional para interpolar os dados de teste. Esse enfoque em essência é o método das funções de base radial (RBF, *radial-basis function*), considerando que isso o aproxima dos trabalhos de pesquisa em interpolação estrita tradicional em um espaço multidimensional.

No contexto de uma rede neural, as unidades ocultas fornecem um conjunto de “funções” que constituem uma “base” arbitrária para os padrões (vetores) de entrada, quando eles são expandidos sobre o espaço oculto: essas funções são chamadas de *funções de base radial*.

As funções de base radial foram inicialmente introduzidas na solução do problema de interpolação multivariada real. O trabalho inicial neste assunto é detalhado em Powell (1985), e mais recentemente em Light (1992).

A construção de uma *rede de função de base radial* (RBF), em sua forma mais básica, envolve três camadas com papéis totalmente diferentes. A camada de entrada é constituída por nós de fonte (unidades sensoriais) que conectam a rede ao ambiente. A segunda camada, a *única* camada oculta da rede, aplica uma transformação não-linear do espaço de entrada para o espaço oculto; na maioria das aplicações, o espaço oculto é de alta dimensionalidade. A camada de saída é linear, fornecendo a resposta da rede ao padrão (sinal) de ativação aplicado à camada de entrada (Haykin 1999).

Quando uma rede de função de base radial é usada para realizar uma tarefa *complexa* de classificação de padrões, o problema é resolvido basicamente pela sua transformação para um espaço de alta dimensionalidade, de uma forma não-linear. A justificativa desse processo é encontrada no *teorema de Cover* sobre a *separabilidade de padrões*, que, em termos qualitativos, pode ser formulado da seguinte forma (Cover, 1965):

Um problema complexo de classificação de padrões disposto não linearmente em um espaço de alta dimensão tem maior probabilidade de ser linearmente separável do que em um espaço de baixa dimensionalidade.

Um outro ponto importante é o fato de que a dimensão do espaço oculto está relacionado à capacidade da rede de aproximar um mapeamento de entrada-saída suave (Mhaskar,1996; Niyogi e Girosi,1996); quanto mais alta a dimensão do espaço oculto, mais precisa será a aproximação.

A rede de função de base radial, conforme mostrada na Figura 28, apresenta várias estratégias de aprendizagem. O processo de aprendizagem, independente de sua fundamentação teórica, pode ser visualizado como segue. Os pesos lineares associados com as unidades de saída da rede tendem a evoluir em uma “escala de tempo” diferente comparada às funções de ativação não-lineares das unidades ocultas. Assim, como as funções de ativação da camada oculta evoluem lentamente de acordo com alguma estratégia de otimização *não-linear*, os pesos da camada de saída se ajustam rapidamente através de uma estratégia de otimização *linear*. O ponto importante é que as diferentes camadas de uma rede RBF realizam tarefas diferentes, e assim é razoável separar a otimização das camadas ocultas e de saída da rede usando técnicas diferentes e talvez operando em diferentes escalas de tempo (Lowe, 1991).

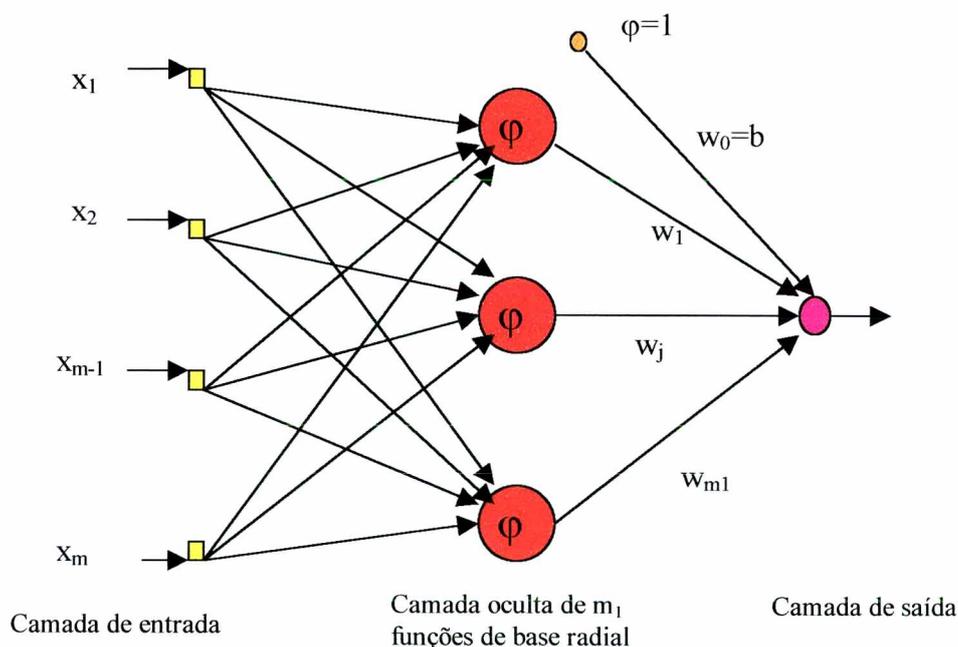


Figura 28 – Rede de função de base radial (Haykin, 1999)

Introduzindo como função de base radial ϕ da camada oculta a *função de Green*, ter-se-á uma rede de regularização que é um *aproximador universal*, já que pode aproximar arbitrariamente bem qualquer função contínua multivariada, dado um número suficientemente grande de unidades ocultas. A rede de regularização tem a *propriedade da melhor aproximação* significando que dada uma função não-linear desconhecida f , sempre existe uma escolha de coeficientes que aproxima f melhor que todas as outras escolhas possíveis. A solução calculada pela rede regularizadora é *ótima*, significando que a rede minimiza um funcional que mede o quanto a solução se desvia de seu valor real como representado pelos dados de treinamento (Haykin,1999).

Existem diferentes estratégias de aprendizagem que podemos seguir para o projeto de uma rede RBF, dependendo de como os centros das funções de base radial da rede são especificados. A estratégia de seleção auto-organizada de centros (Moody e Darken,1989; Lippmann,1989; Chen et al, 1992) utiliza um *processo de aprendizagem híbrido*:

- ◆ Estágio de *aprendizagem auto-organizada*, cujo propósito é estimar localizações adequadas para os centros das funções de base radial na camada oculta.
- ◆ Estágio de *aprendizagem supervisionada*, que completa o projeto da rede estimando os pesos lineares da camada de saída.

3.5 – Redes Neurais e os Métodos Tradicionais

Redes neurais e técnicas tradicionais para a solução de séries temporais têm sido comparadas em diversos estudos. Muitos trabalhos apresentados (Foster et al. 1991, Kang 1991, Sharda e Patil 1990) , utilizaram os dados do trabalho conhecido por “M-competition” (Makidrakis et al. 1982), onde foram efetuados ajustes em 1001 séries temporais reais.

Na competição original, vários grupos de profissionais da área de prognósticos, apresentaram os dados de ajuste mais recentes para cada série temporal.

Cada grupo era constituído de profissionais com experiência e conhecimento reconhecidos em determinada técnica particular. Por exemplo, a técnica de Box-Jenkins foi utilizada por uma autoridade reconhecida no mundo científico.

Cada grupo tinha autonomia para utilizar qualquer uma das técnicas em seu domínio do conhecimento, para efetuar o prognóstico nas várias séries temporais apresentadas.

As séries temporais foram obtidas de várias fontes reais do mundo e os grupos de profissionais não tinham conhecimento da fonte ou da natureza de cada série temporal.

Após cada grupo haver preparado o seu prognóstico, Makridakis comparou os resultados com os valores atuais de cada série não pertencentes aos valores disponibilizados para efetuar o ajuste. Normalmente a modelagem das séries temporais é linear e não apresentam a capacidade de capturar condutas não lineares (Hill et al. 1996)

Apesar de métodos estatísticos tradicionais aplicados a séries temporais apresentarem bons resultados, muitos métodos apresentam limitações, porque é possível cometer enganos na especificação da forma funcional entre as variáveis dependentes e independentes, e conseqüentemente falha na transformação de dados.

Desde 1970, o uso do modelo estacionário linear auto-regressivo de média móvel (ARMA) tem sido muito difundido em muitas análises de áreas científicas e prognóstico de dados de séries temporais (Gooijer e Kumar 1992). Existem, no entanto, várias ocasiões em que o objeto a ser estudado, a teoria ou os dados sugerem que a série temporal seja gerada a partir de processo não-linear.

Se a teoria providenciar algum conhecimento sobre o fenômeno de não-linearidade subjacente aos dados, o processo de modelagem é essencialmente facilitado, com a estimativa de todos os parâmetros requeridos e existentes no modelo. No entanto, esta opção raramente é avaliada na prática. Alternativamente, a seleção de um modelo particular, não-linear, ajustado aos dados da série temporal, e sujeito a uma bateria de diagnósticos teste, para checar a aderência com que o modelo escolhido se adapta aos dados, corresponde à estratégia usual utilizada na análise não-linear de séries temporais, e geralmente envolve mais problemas ao profissional do que no caso linear.

O problema imediato é a seleção de um modelo não-linear. Muitos modelos têm sido propostos, no entanto, na prática, o uso adequado de um modelo não-linear deve satisfazer a capacidade de capturar alguns valores não lineares apresentados pelos dados e, além disso, deve corresponder a um certo conhecimento intuitivo para a escolha. Após a escolha do modelo, o próximo problema é achar evidências bastante fortes nos dados, para abandonar a utilização da modelagem linear convencional.

Esse procedimento é particularmente importante, considerando que na prática a necessidade da modelagem não linear é *a priori* muitas vezes duvidosa.

Muitos testes estatísticos têm sido propostos com este propósito, e suas propriedades de amostras tem sido investigadas. Alguns procedimentos para a aplicação e análise destes testes são necessários, tornando a tarefa de prognosticar séries temporais com modelagem não linear, um problema de difícil solução.

Claramente, a análise e prognóstico de séries temporais não-lineares requerem mais experiência e normas do que nos casos lineares simples. Contudo, ao longo dos últimos anos, a aplicação de modelos não lineares tem sido introduzida no estudo de prognósticos.

A relação geral entre o valor Y_t e uma sucessão finita de valores da série histórica de pontos observados nos tempos $t=0, \pm 1, \pm 2, \dots$, e o valor ε_t , correspondente ao *ruído branco* do processo, pode ser representada pela função de valor real não linear h [3.5-1].

$$Y_t = h\left(Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}\right) + \varepsilon_t \quad [3.5-1]$$

Priestley (1980) introduziu o modelo denominado *State-Dependent Model* (SDM), e considera no equacionamento uma linearização local da equação [3.5-1], utilizando a expansão em série de Taylor em torno de um ponto fixo [3.5-2]

$$Y_t + \sum_{i=1}^p \phi_i(Z_{t-1}) Y_{t-i} = \mu(Z_{t-1}) + \varepsilon_t + \sum_{j=1}^q \theta_j(Z_{t-1}) \varepsilon_{t-j} \quad [3.5-2]$$

Todos os parâmetros desconhecidos do modelo são dependentes do estado do processo no tempo $(t-1)$, e a forma particular de escolha dos mesmos abrange um número importante de modelos de séries temporais em casos especiais.

A adoção de $\mu(x)$, $\phi_i(x) (i=1, 2 \dots p)$, e $\theta_j(x) (j=1, 2 \dots q)$ todos constantes, independentes de x , resulta no modelo linear, denominado *Stationary Linear Autoregressive Moving Average* ARMA(p, q).

O modelo bilinear - BL (Granger e Anderson 1978, Subba Rao e Gabr 1984) considera os parâmetros $\mu(x)$ e $\phi_i(x) (i=1, 2 \dots p)$, constantes, e

$$\theta_j(Z_{t-1}) = \theta_j + \sum_{v=1}^q c_{jv} Y_{t-v} \text{ com } (j=1,2,\dots,q)$$

No entanto, devido a inúmeras dificuldades técnicas, esse resultado é limitado geralmente a um modelo bilinear simples, tendo sido derivada a condição suficiente para a existência da solução estacionária estrita (Liu e Brockwell 1988).

Tomando os valores $\mu(x)$ constante, e

$$\theta_j(x) = 0 \text{ e } \phi_i(\phi_{t-1}) = \phi_i + \pi_i \exp(-\gamma Y_{t-1}^2) \text{ com } (i=1,2,\dots,p)$$

tem-se o modelo *Exponential Autoregressive Model* (EAR) de ordem p , onde o parâmetro γ representa o fator de escala da série temporal. Para valores gerais de p (Jones 1978, Ozaki e Oda 1978), a determinação dos parâmetros envolvidos no equacionamento é essencialmente um problema de otimização não-linear, tendo sido proposto (Haggan e Ozaki 1981) a obtenção dos parâmetros ϕ_i , a partir da fixação de p e γ em um dos valores da grade. O valor $p=2$ (Haggan e Ozaki 1980) incorpora ambos amplitude - dependência da frequência e procedimentos “ciclo limite”, em um modelo de série temporal discreta.

O modelo denominado *Threshold Autoregressive Model* (TAR) foi extensivamente estudado (Tong 1980, 1983), e representa um conjunto finito de possíveis modelos auto-regressivos (AR), cujo processo obedece em cada ponto, com inicializações para a passagem de um membro do conjunto para outro. Se a passagem é determinada através da colocação de valores passados relativamente à inicialização, então o modelo é denominado de “auto excitado”. A essência da idéia subjacente nesta classe de modelos é a linearização de modelos não-lineares, sobre o espaço de estado, pela introdução de inicializações.

O modelo auto-excitado é suficientemente geral (Tong e Lim 1980) para capturar certos aspectos, assim como, ciclos limites, amplitude de frequências e fenômeno de salto, que não é possível capturar com a utilização de modelagem linear.

Tomando $\mu(x) = \phi_0^{(j)}$, $\theta_j(x) = 0$ (para todo x), e $\phi_i(z_{t-1}) = \phi_i^{(j)}$ se $Y_{t-d} \in R^{(j)}$ ($i=1,\dots,p$; $j=1,\dots,l$), onde d é inteiro positivo e $R^{(j)}$ é um subconjunto de números reais (inicializações) da linha de números reais R^1 , pode-se escrever o modelo auto excitado [3.5-3], *Self-Exciting Threshold Autoregressive Model* (SETAR).

$$Y_t + \phi_0^{(j)} + \sum_{i=1}^p \phi_i^{(j)} Y_{t-i} = \varepsilon_t^{(j)} \quad \text{se } Y_{t-d} \in R^{(j)} \quad \text{com } j=1,2,\dots,l \quad [3.5-3]$$

A generalização do modelo auto excitado (Chan e Tong 1986), efetuando mudanças nos parâmetros de um regime para outro suavizado ao longo do tempo, resulta o modelo denominado *Smooth Threshold Autoregressive Model* (STAR) definido por [3.5-4].

$$Y_t + \sum_{i=1}^p \phi_i Y_{t-i} + \left\{ \pi_0 + \sum_{i=1}^p \pi_i Y_{t-i} \right\} F(Z_t) = \mu + \varepsilon_t \quad [3.5-4]$$

Várias formulações para F têm sido propostas, assim como

$$F(Z_t) = \phi\left[K(Y_{t-d} - r)\right]$$

onde $\phi(\dots)$ é uma distribuição cumulativa da curva padrão de distribuição normal.

Aqui d é o parâmetro de atraso, r é a localização ou parâmetro de inicialização que indica a ocorrência de transição e K é o parâmetro de inclinação, e que efetua a suavização do processo de transição.

Outra forma popular de F é a função distribuição cumulativa da distribuição logística

$$F(Z_t) = \frac{\exp(Z_t)}{(1 + \exp(Z_t))} \quad \text{com } Z = K(Y_{t-d} - r)$$

que resulta no modelo denominado *Logistic Smooth Transition Autoregressive Model*.

Esse modelo apresenta uma vantagem sobre o modelo da equação [3.5-4], cujos parâmetros podem ser estimados por mínimos quadrados condicionais usando derivadas analíticas (Luukkonen et al. 1988).

O modelo denominado *Autoregressive with Time-Dependent Coefficients Model* é uma generalização do modelo linear AR(p) da equação [3.5-2], tornando os parâmetros autoregressivos funções determinísticas do tempo.

No contexto da equação [3.5-2], isto nos diz que o vetor de estado é uma função de t ou t/n , onde n é o tamanho da amostra de observações obtidas para $t=1,\dots,n$.

Considere-se a substituição

$$F(K(Y_{t-d} - r)) = F(K(Z_t - r))$$

onde $0 < r < 1$ quando $z_t = 1/n$. Quando K crescer, F se aproxima da função de Heaviside $H(z_t - r)$ e a equação [3.5-2] torna-se a de um processo auto-regressivo AR(p) com coeficientes mutáveis com o ponto.

Uma grande flexibilidade foi observada, assumindo que o mecanismo descrito como um modelo AR linear, muda para um não determinístico e então para a forma determinística ao longo do tempo.

Como exemplo, considere [3.5-2] com $\phi_l(z_{t-l}) = \phi_{lt}$, onde $\{\phi_{lt}\}$ é um processo estocástico, ou seja, um processo AR(2), independentemente de $\{\varepsilon_{jt}\}$. Esta classe de modelo é simplesmente denominada *Doubly Stochastic Time Series Model* (Tjostheim 1986) e foi utilizada para a análise de preços diários de estoques (Tjostheim e Tyssedal 1988).

Uma subclasse do modelo estocástico duplo AR é o denominado *Random Additive Perturbations of AR model* (RCA) (Nicholls e Quinn 1982), que é definido a partir da introdução de perturbações aleatórias nos coeficientes do modelo AR.

Tomando na equação [3.4-2] $\mu(x)$ constante, $\theta_j(x) = 0$ (para todo x) e $\phi_l(z_{t-l}) = \beta_l + b_l(t)$, tem-se o modelo RCA(p), conforme equação [3.5-5].

$$Y_t + \sum_{i=1}^p [\beta_i + b_i(t)] Y_{t-i} = \mu + \varepsilon_t \quad [3.5-5]$$

onde β_i 's ($i = 1, \dots, p$) são constantes e $(b_1(t), \dots, b_p(t))$ é um processo vetorial identicamente distribuído com média zero, independente do processo de $\{\varepsilon_{jt}\}$.

O modelo denominado *Autoregressive Conditional Heteroscedastic Model* (ARCH), introduzido por Engle (1982), apresenta famílias paramétricas de específicos tipos de heterocedasticidade nos modelos AR. O modelo ARCH de ordem q é definido como um modelo AR(q), onde $\{\varepsilon_{jt}\}$, condicionado no conjunto de informações para tempos superiores a t-1, apresenta uma variância

$$h_t = \alpha_0 + \sum_{i=1}^q \alpha_i (\varepsilon_{t-i}^2) \text{ com } \alpha_0 > 0 \text{ e } \alpha_i \geq 0 \text{ (} i = 1, 2, \dots, q \text{)}$$

A equação do modelo ARCH(q) é dada por [3.5-6].

$$Y_t = \varepsilon_t \left[\alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 \right]^{1/2} \quad [3.5-6]$$

Nesse caso a não-linearidade é suportada pela condição da variância $\{\varepsilon_t\}$, enquanto na maioria dos demais modelos não-lineares a média condicional varia com o tempo.

Os modelos não-lineares discutidos anteriormente podem ser combinados entre si (Granger 1989, Weiss 1986, Zakoian 1990, Gouriou e Monfort 1990), ou mesmo utilizar modelos lineares estendidos com a introdução de elementos não-lineares em sua estrutura.

A motivação está em achar uma parametrização simples e flexível da média condicional e da variância condicional, da série temporal. A idéia é aproximar a forma funcional dessas quantidades por meio de funções lineares e constantes da função.

Apesar de estudos recentes terem sido conduzidos no sentido de desenvolver métodos capazes de responder às necessidades de modelagem, teste e prognóstico de séries temporais não-lineares (Gooijer e Kumar 1992), muitos trabalhos (Connor 1988, Hornik et al. 1989, Wasserman 1989, White 1992) têm indicado que a utilização de redes neurais podem contornar as limitações dos processos tradicionais ou mesmo vir a superá-los.

Muitos processos tradicionais não apresentam a capacidade de aprendizagem com o acréscimo de novos valores da série temporal, o que tem sido reivindicado para a modelagem com utilização de redes neurais (Widrow e Sterns 1985).

Redes neurais têm sido aplicadas matematicamente para a aproximação universal de funções (Cybenko 1989, Funahashi 1989, Hornik et al. 1989), e de suas derivadas (White et al. 1992).

Podem ser utilizadas para a aproximação de regressões ordinárias e não-lineares por mínimos quadrados (White 1992, White e Stinchcombe 1992), regressão não-paramétrica (White 1992) e análise de séries de Fourier (White e Gallant 1992).

Redes neurais podem aproximar sempre da melhor forma a caracterização de uma série temporal. Segundo White (White 1992), “redes neurais são capazes, em princípio, de prover boa aproximação sobre tudo o que quisermos representar”.

Redes neurais são inerentemente não-lineares (Rumelhart e McClelland 1986, Wasserman 1989) e são capazes de estimar bem funções não-lineares (White 1992, White e Gallant 1992),

e de extrair elementos residuais não-lineares. As redes podem ao menos transformar parcialmente os dados de entrada, se necessário (Connor 1988, Donaldson et al. 1993, Gorr 1994).

Conforme apresentado anteriormente (Gooijer e Kumar 1992), apesar de haver benefícios na utilização de modelos não-lineares tradicionais, as redes neurais podem apresentar resultados melhores de prognóstico, além de aparentemente estender o horizonte de prognóstico para períodos futuros mais distantes, considerando que as séries temporais não-lineares apresentam modelos de prognóstico mais divergentes, do que as lineares, com o tempo (Hill et al. 1996). As redes neurais podem particionar o espaço amostral e construir diferentes funções em diferentes porções deste espaço (Wasserman 1989). Igualmente, redes neurais são capazes de construir porções verdadeiras de modelos não-lineares, incluindo a modelagem de descontinuidades, tendo sido estudadas para a utilização em prognóstico de séries temporais (White 1994, Gorr 1994).

Estudos efetuados (Sharda e Patil 1990) utilizando algumas séries do trabalho de Makidrakis (Makidrakis et al 1982) concluíram que redes neurais apresentam melhores resultados, no prognóstico de demandas, que aquele obtido com a aplicação da modelagem Box-Jenkins. Foster e outros (Foster et al. 1992), em estudo de prognóstico, utilizando os dados de Makidrakis, encontraram o ajuste por redes neurais, inferior aos métodos de Holt, Brow e os modelos estatísticos, utilizando séries temporais com dados anuais, comparativamente com as séries com dados trimestrais, sendo que o estudo não efetuou comparações com dados mensais.

Alguns trabalhos concordam (Sharda e Patil 1990, Tang et al. 1990) que para séries históricas com uma base de dados bastante grande, redes neurais e a modelagem Box-Jenkins produzem resultados equivalentes.

Outro estudo efetuado, utilizando os dados de Makidrakis (Kang 1991) comparou resultados das 50 séries mais apropriadas para serem trabalhadas pela técnica Box-Jenkins, conforme indicado por Pack (Pack e Downing 1983).

Sobre todos os casos estudados, Kang encontrou que a metodologia Box-Jenkins apresentou um valor MAPE (Erro Médio Percentual Absoluto) superior ou equivalente ao daquele encontrado para 18 arquiteturas diferentes de redes neurais; o erro de prognóstico para o caso das redes neurais somente foi inferior nos casos em que os dados históricos apresentavam tendência de crescimento ou sazonalidade.

Kang também concluiu que as redes neurais geralmente apresentam melhores resultados nos casos em que o horizonte de prognóstico está situado nos primeiros períodos futuros.

Esse efeito tem sido observado na utilização de redes neurais para prognosticar dados macroeconômicos (White 1994) e flutuações das taxas de câmbio (Zimmerman 1994).

Em publicação levada a efeito pelo IEEE Computer Society Press (Vemuri e Rogers 1993), foram transcritos os principais artigos técnicos relacionados com a teoria de redes neurais e suas aplicações no prognóstico de séries temporais, os quais permitem avaliar a aplicação crescente dessa importante ferramenta nos mais diversos campos do conhecimento.

Uma importante aplicação apresentada foi o prognóstico de demanda de carga requerida para um determinado sistema elétrico (Park et al. 1991) que, a partir da aplicação de redes neurais, utilizando dados de temperatura ambiente e carregamentos elétricos medidos no tempo, estabelece a rede para prognóstico das demandas futuras.

Os autores concluíram que a utilização de redes neurais é apropriada para interpolar os dados padrões de temperatura e carregamento, em conjuntos de dados, obtidos do treinamento da rede para obter os padrões futuros de comportamento do carregamento.

A determinação de prognósticos de demandas a partir de séries temporais multivariáveis, com a utilização de redes neurais, foi estudada pela Universidade de Siracusa (Chakraborty 1992), que efetuou a aplicação de redes neurais a uma série temporal da variação do preço de um produto alimentício de grande consumo, com dados de 100 meses, em três cidades americanas Buffalo, Mineápolis e Kansas City.

Segundo os autores, a aplicação de redes neurais foi coroada de sucesso no aprendizado das variações de preço do produto e, da mesma forma, ao prognosticar com bastante precisão, valores futuros do produto.

Segundo o artigo de Hill e outros (Hill et al. 1996), os resultados encontrados com a aplicação de redes neurais em séries históricas foram melhores do que os apresentados em trabalhos anteriores (Sharda e Patil 1990, Foster et al. 1991, Kang 1991), para as mesmas séries.

Os autores sugerem que a heurística desenvolvida para a aplicação da rede neural responde parte da questão. A outra parte da resposta pode ser avaliada a partir do fato que, nos trabalhos anteriores quando não houve a utilização de um método de prognóstico, foi adotado no trabalho o ajuste exponencial simples. A rede neural utilizada apresentava uma arquitetura de prognóstico testada e que requeria uma grande quantidade de parâmetros a serem determinados. Isso sugere que a seleção da melhor arquitetura de rede neural é crucial para o sucesso da modelagem, utilizando redes neurais.

A aplicação de redes neurais a sistemas caóticos tem motivado a pesquisa em arquitetura de redes neurais que possam responder adequadamente a estes sistemas. Deppisch e outros

(Deppisch et al. 1991) desenvolveram uma técnica para treinamento de redes neurais e prognóstico de séries temporais caóticas. O procedimento consiste em sucessivos treinamentos em uma arquitetura hierarquizada e apresenta como princípio: Um sinal de entrada é introduzido em várias redes que são partes de uma única grande rede. Essas redes parciais são treinadas simultaneamente com um sinal de função objetiva de tal modo que a soma de todas as saídas se aproximam da função objetiva do sinal de entrada. O primeiro passo é treinar a rede 1 até a saturação da função erro de aprendizado, na sequência é treinada a rede 2, com a rede 1 mantida fixa, até corrigir a diferença entre a função objetivo e a saída da rede 1. Sucessivamente outras partes da rede são treinadas da mesma forma, sendo finalmente conectadas entre si de forma a formar uma só rede, conforme mostrado na Figura 29.

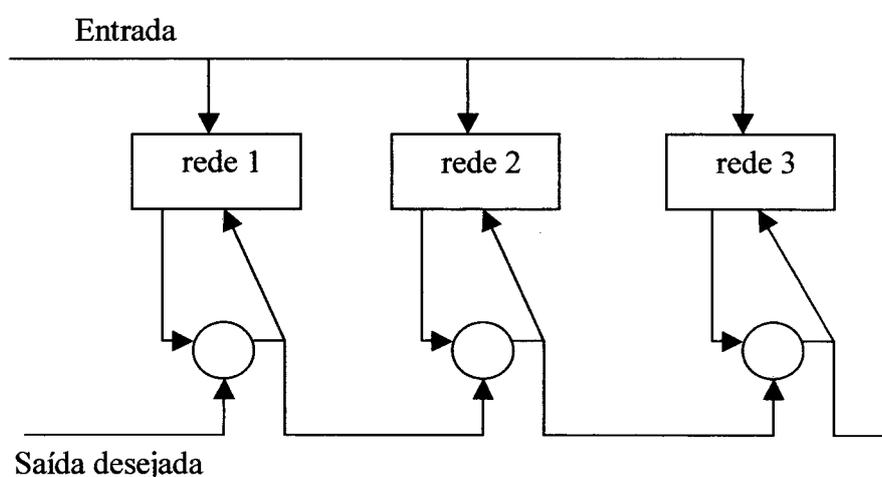


Figura 29-Rede neural para prognóstico de séries caóticas (Deppisch et al. 1991)

Trabalhos recentemente apresentados (Smith e Mason 1997), baseados em aplicações práticas nos mais diversos campos do conhecimento, estabelecem para as redes neurais o caráter de metodologia universal de aproximação de séries temporais.

As redes neurais têm sido utilizadas para a solução, cada vez com maior frequência, de séries temporais (Hoptruff 1993, Marques et al. 1992, Sharda e Patil 1990, Tang e Fishwick 1993), com ênfase na solução de prognóstico de fenômenos financeiros, a exemplo de flutuações de

câmbio (Refenes 1993, Yamaba e Kurashima 1991), avaliação de contratos (Dutta e Sbekkar 1988, Surkan e Singleton 1990), e preço de estoques (Kamijo e Tanigawa 1990, Kimoto e Asakawa 1990, Schoenenburg 1990, White 1988).

Além do campo financeiro, a literatura tem apresentado trabalhos relacionados com a aplicação de redes neurais para a composição de custos estimados nas mais diversas disciplinas, tendo sido publicado (Ahmad e Rahman 1994) trabalho relacionado com a redução de custos na área de Engenharia Civil, custos de softwares (Kumar et al. 1994, Venkatachalam 1993) e a apresentação de uma rede neural com um sistema esperto para a determinação de custos (Rao et al. 1993).

A comparação com os métodos de regressão está sempre presente nas pesquisas e serve para validar a aplicação das redes neurais, como apresentado na modelagem de custos de vasos de pressão (Brass et al. 1994, Gerrard et al. 1994), e na modelagem de custos estimados de materiais para tubulações de aço carbono (Garza e Rouhana 1995).

As pesquisas recentes têm efetuado comparações entre os métodos tradicionais, tais como métodos de regressão, Box-Jenkins e outros, com a metodologia que utiliza redes neurais, com a finalidade de estabelecer padrões de comportamento dos métodos e determinar graus de exatidão.

Com as flutuações experimentadas pelas taxas de câmbio, ocorridas em passado recente, a determinação de prognósticos dessas variações tem consumido os esforços dos pesquisadores e profissionais de finanças internacionais (Hu et al. 1999)..

Uma grande variedade de modelos paramétricos não-lineares foram desenvolvidos (Gooijer e Kumar 1992) e aplicados para o prognóstico financeiro.

Enquanto estes modelos apresentam resultados bastante satisfatórios para séries temporais específicas, eles não apresentam características adequadas para outras aplicações.

Pelo fato de existirem muitas possibilidades de padrões não-lineares, a pré-especificação do modelo a ser utilizado restringe o uso desses modelos paramétricos não-lineares.

A respeito da proliferação de modelos não-lineares Diebold e Nason (Diebold e Nason 1990), anotaram , ” *a enorme quantidade de prováveis candidatos a modelagem não linear torna a tarefa de encontrar uma aproximação que conduza à geração de um processo de dados reais, uma tarefa difícil, e a aparente grande quantidade de modelos paramétricos não lineares é de*

fato uma pequena parte da classe de prováveis processos de geração de dados reais, não lineares”.

Os métodos não-paramétricos e não-lineares investigados (Diebold e Nason 1990, Meese e Rose 1991, Mizrach 1992), foram incapazes de melhorar os resultados da modelagem randômica em prognósticos de flutuações de câmbio.

Recentemente as pesquisas com a utilização de redes neurais têm sido efetuadas e determinaram tratar-se de uma ferramenta importante no prognóstico de séries temporais (Zhang et al. 1998).

As redes neurais apresentam a capacidade de flexibilizar o mapeamento de funções não-lineares, que permitem o seu uso para a aproximação de qualquer função contínua com um valor arbitrário desejado de exatidão (Cybenko 1989, Hornik et al. 1989).

A pesquisa efetuada para validar a performance das redes neurais para o prognóstico de flutuações cambiais (Hu et al. 1999) apresenta algumas implicações administrativas decorrentes dos resultados obtidos.

Os profissionais que efetuam os prognósticos das taxas de câmbio têm maior confiança nos dados obtidos utilizando ferramentas com redes neurais, as quais apresentam uma maior exatidão nos resultados do que os métodos randômicos.

Mesmo nos casos em que as redes neurais produzem prognósticos inferiores, os resultados apresentados ainda são comparáveis aos dos outros métodos.

As decisões de compra e venda na área de câmbio são baseadas em prognósticos de curto prazo, onde as redes neurais apresentam melhor performance que os outros métodos.

Amostras de valores de séries temporais com um grande número de observações podem apresentar resultados com bastante exatidão para prognósticos de longo termo, enquanto para um horizonte de curto termo é necessário um número relativamente pequeno de observações para conduzir a um resultado do prognóstico igualmente com exatidão.

Capítulo IV. Aplicação de redes neurais em séries temporais

4.1 – Introdução

Apesar de existirem inúmeros métodos para o ajuste e prognóstico de valores futuros, a partir de séries temporais, os métodos que melhor se adaptam a prognosticar a curto prazo, mesmo que a partir de uma pequena amostra de valores históricos, estão relacionados com a utilização de algoritmos originados das teorias de IA, mais precisamente neste caso dos métodos com redes neurais, que podem dar respostas mesmo para séries temporais não lineares (Vemuri e Rogers 1993).

Para corroborar a assertiva de que os algoritmos com redes neurais são capazes de efetuar prognósticos de valores futuros a partir de um processo de aprendizagem dos valores históricos da série temporal, este capítulo está dedicado à obtenção e análise de resultados, utilizando esta ferramenta.

4.2 - Metodologia

A metodologia, a ser utilizada para verificar a aplicação de redes neurais em séries temporais, será conduzida de forma a produzir os resultados esperados a partir de uma modelagem que contemple a utilização de valores reais da série temporal escolhida para efetuar o estudo.

O primeiro passo será a obtenção de valores de uma série temporal com dados suficientes, que permitam estabelecer vetores de entrada de dados e respectivas saídas, permitindo a aplicação ao modelo a ser escolhido para efetuar o estudo de ajuste.

O índice BOVESPA –Índice da Bolsa de Valores de São Paulo, será utilizado como série temporal para ajuste do problema proposto.

O passo seguinte será a escolha do modelo de rede neural a ser utilizada para efetuar o estudo de ajuste da série temporal, e que possa dar a resposta adequada aos valores dos vetores de entrada e saída propostos.

Com a obtenção da série temporal a ser trabalhada e escolhido o modelo de rede neural a ser utilizado, o passo seguinte exige a elaboração de programa computacional que possa processar os dados e efetuar os ajustes da rede neural adotada no estudo.

Nesse caso, será utilizado o conjunto de programas existentes MATLAB (Neural Network Toolbox 3), da empresa “ The Math Works Inc.”, que permite o processamento da maioria das aplicações com redes neurais. Com o processamento efetuado, pode-se analisar o comportamento

das redes neurais para o ajuste e o prognóstico de séries temporais, que é o objetivo a que se propõe a pesquisa.

4.3 - Modelo

O modelo de rede neural a ser utilizado para efetuar a pesquisa é do tipo *feedforward* de múltiplas camadas, cuja origem remonta ao trabalho de Rosembat (1958) que apresentou o Perceptron, e o algoritmo denominado *backpropagation* descoberto por Paul Werbos (1974), e redescoberto independentemente por Rumelhart (1986) e Parker(1982), que tem sido amplamente utilizado como algoritmo de aprendizagem.

O modelo utilizado para efetuar o estudo de caso constitui-se de uma sucessão de cem (100) valores do índice BOVESPA de fechamento da Bolsa de Valores de São Paulo, sendo os primeiros noventa (90) valores utilizados para a fase de treinamento e os demais dez (10) valores para a fase de prognóstico.

Os conjuntos de valores para processar a fase de treinamento da rede, utilizando os 90 valores iniciais da série temporal, foram determinados conforme os passos que seguem:

- Para a rede utilizada para efetuar o estudo, adotou-se uma camada de saída com apenas um neurônio, conforme trabalho apresentado por Chakraborty e outros 1992 e mostrada na Figura 30.

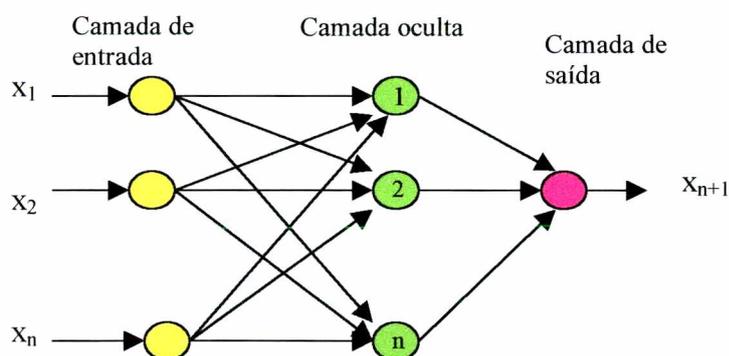


Figura 30 – Rede neural utilizada no caso exemplo da pesquisa

- Para se definirem alguns casos a serem estudados, com o objetivo de verificar o comportamento de arquiteturas diferentes, foram processados casos com 2, 4, 6 e 8 valores de entrada, constituindo-se na camada de entrada.

- A camada oculta igualmente foi processada para 2, 4, 6 e 8 neurônios, tendo-se na combinação das arquiteturas um estudo de caso com 16 situações diferentes.
- Para a determinação do conjunto de vetores de entrada da rede, os 90 valores foram apresentados à rede conforme esquema indicado:

Seja a sucessão de valores $a_1, a_2, a_3, a_4, a_5, a_6, a_7, \dots, a_{n-2}, a_{n-1}, a_n$, onde $n=90$, para uma camada de entrada constituída por 2 valores, por exemplo, ter-se-á para o conjunto de entrada p e o de saída desejada t :

$$p = \begin{Bmatrix} a_1 & a_2 \\ a_2 & a_3 \\ \dots & \dots \\ a_{n-2} & a_{n-1} \end{Bmatrix} \quad t = \begin{Bmatrix} a_3 \\ a_4 \\ \dots \\ a_n \end{Bmatrix}$$

Esse procedimento de aprendizagem é denominado aprendizado supervisionado, pois o processo de treinamento da rede ocorre a partir da existência de pares de valores de entrada p e saída t , com a mensuração do erro cometido a cada novo valor de saída, ou seja, a cada iteração.

O algoritmo de aprendizagem *backpropagation* envolve a fase de propagação para a frente (*forward-propagation*), que inicia no momento em que os valores do vetor de entrada são apresentados aos neurônios de entrada, na primeira camada.

Um nó (neurônio) típico (x_j) recebe diversos sinais de entrada, conforme mostrado na Figura 31, onde cada um dos quais é multiplicado pelos correspondentes pesos (W_i), definido pela equação [4.3-1]

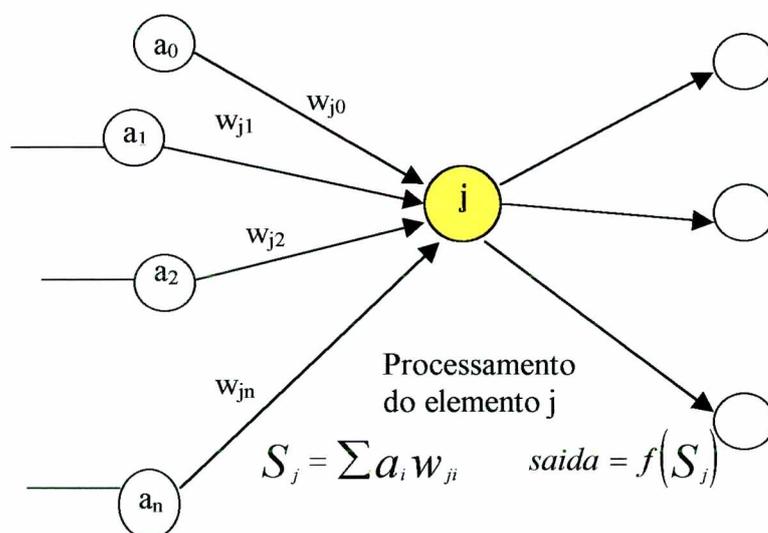


Figura 31 – Processamento do neurônio j da uma rede neural

$$S_j = \sum_{i=1}^n w_{ji} x_i + w_{j0} \quad [4.3-1]$$

Este sinal passa pela função de ativação $f(s_j)$, que é tipicamente não linear, sendo adotada a função sigmoidal, dada pela equação [4.3-2] e mostrada graficamente na Figura 32.

$$f(S_j) = \frac{1}{1 + e^{-S_j}} \quad [4.3-2]$$

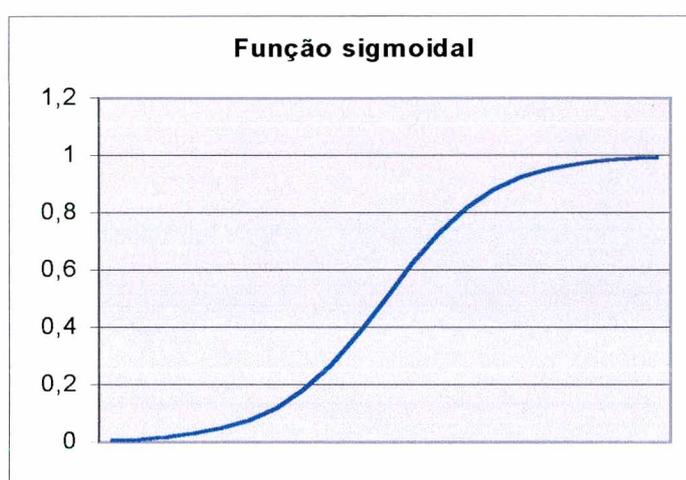


Figura 32- Função sigmoidal

Após computada a função sigmoidal com S_j , o resultado representa o nível da ativação do neurônio j , cujo sinal é enviado para frente, por todas as interconexões existentes, aos demais neurônios.

Cada neurônio recebe também um sinal constante (*bias*), com valor do sinal de entrada unitário ($x_0 = 1$) que, somado à função S_j , apresenta o mesmo efeito que o deslocamento horizontal da curva sigmoidal, providenciando um valor ajustável de entrada na função sigmoidal, para cada unidade alvo.

A terceira camada de neurônios recebe os sinais provenientes da segunda camada, que foi processada com a função sigmoidal e efetua a soma dos produtos dos sinais de ativação pelos correspondentes pesos sinápticos, obtendo-se os valores de ativação desta terceira camada, à qual

é aplicada uma função linear, ou seja, os valores de ativação encontrados correspondem aos valores dos vetores de saída da rede neural.

A segunda fase do algoritmo *backpropagation* inicia logo após o momento em que a primeira aproximação dos vetores de saída é apresentada na última camada de neurônios, e corresponde a um processo de correção dos pesos calculados, do fim para o começo (*backward-propagation*) da rede neural em estudo.

Para a aplicação do algoritmo, é necessário que sejam apresentados os valores de saída desejados, correspondentes às entradas que iniciam o processo de aprendizagem.

Para obter a saída esperada, em função de uma entrada na primeira camada de neurônios, é necessário efetuar a correção dos caminhamentos w (pesos), com base na diferença experimentada entre o vetor de saída calculado e o vetor alvo.

A *regra delta generalizada* efetua as correções necessárias para a obtenção da rede que permita, após o processo de aprendizagem, apresentar um desempenho esperado na determinação de valores futuros da série temporal em estudo.

A equação representativa desta regra é a equação [4.3-3], onde E é uma função de erro do peso que está sendo modificado e β é o coeficiente de aprendizagem.

$$\Delta w_{ij} = -\beta \frac{\partial E}{\partial w_{ij}} \quad [4.3-3]$$

A primeira aproximação do conjunto de vetores de saída é comparado com o vetor de valores estabelecidos $[t]$, permitindo obter o cálculo do erro da unidade neural δ , conforme a equação [4.3-4]

$$\delta_j = (t_j - a_j) f'(S_j) \quad [4.3-4]$$

onde:

t_j - valores atribuídos à saída desejada

a_j - corresponde aos valores obtidos em primeira aproximação do vetor de saída da rede neural.

$f'(S_j)$ - Derivada de primeira ordem da função sigmoidal (Ver Figura 33)

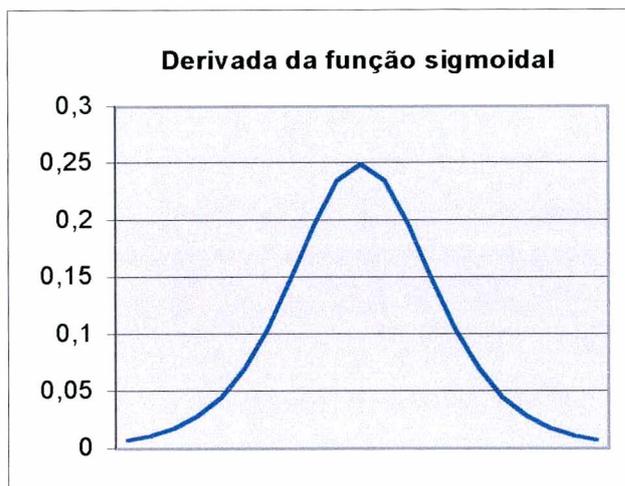


Figura 33-Derivada da função sigmoideal

Para a camada intermediária não há vetores $[t]$ e $[a]$ disponíveis para o processamento do erro. Nesse caso é utilizada a relação [4.3-5]

$$\delta_j = \left[\sum_k \delta_k w_{kj} \right] f'(S_j) \quad [4.3-5]$$

onde os valores δ_j correspondem aos valores dos erros calculados para os neurônios da penúltima camada (camada intermediária), com base nos valores dos erros δ_k , calculados na última camada.

Os valores dos pesos são corrigidos do fim para o começo da rede, utilizando a função erro δ apresentada anteriormente, obtendo-se a equação [4.3-6] para a correção dos pesos.

$$\Delta w_{ji} = \beta \delta_j a_i \quad [4.3-6]$$

Para superfícies de ordem elevada, o gradiente de descida pode ser muito lento se o coeficiente de aprendizagem β for muito pequeno, e produzindo instabilidade se for muito alto; a adição de um termo de momento, α , pode melhorar a situação através da implementação de uma certa inércia da contribuição anterior do peso, modificando a equação conforme [4.3-7] (Shaw e Simões 1999):

$$\Delta w_{ij} = -\beta \frac{\partial E}{\partial w_{ij}} + \alpha \left(\Delta w_{ij} \right)_{anterior} \quad [4.3-7]$$

onde:

Δw_{ij} = Fator de correção dos pesos sinápticos

β = Coeficiente de aprendizagem (normalmente adotado igual a 0,01)

α = Coeficiente de momento (normalmente adotado 0,9)

$(\Delta w_{ij})_{anterior}$ = Fator de correção dos pesos sinápticos do ciclo anterior.

$\partial E / \partial w_{ij}$ = Derivada da função erro relativamente ao peso sináptico.

As redes neurais são sensíveis à escala das variáveis; se a magnitude dos valores de entrada diferirem entre si, a rede pode erroneamente atribuir uma importância maior a valores maiores.

Além dos valores das magnitudes diferenciadas, os valores de entrada devem estar dentro da faixa da função de ativação (de 0 a 1, ou de -1 a +1), para evitar a paralisia do treinamento, pois a função de ativação estará sendo excitada em uma região muito plana, com derivada praticamente nula (Shaw e Simões 1999).

Para evitar os inconvenientes mencionados, os valores dos dados de entrada devem ser normalizados, que no caso em estudo foram efetuados conforme equação [4.3-8]

$$\left(x_i \right)_{normalizado} = \frac{\left(x_i \right)_{atual}}{\sqrt{\sum \left(\left(x_i \right)_{atual} \right)^2}} \quad [4.3-8]$$

Esse procedimento fará com que os valores normalizados estejam situados entre -1 e +1 de forma a atender a relação [4.3-9].

$$\sqrt{\sum \left(\left(x_i \right)_{normalizado} \right)^2} = 1 \quad [4.3-9]$$

As sub-rotinas utilizadas do conjunto de programas do MATLAB e seus respectivos significados são apresentados a seguir.

$net = newff (PR , [S_1 S_2 \dots S_{N_1}] , \{ TF_1 TF_2 \dots TF_{N_1} \} , BTF , BLF , PF)$

onde:

PR – Matriz [R x 2] com valor mínimo e máximo dos R elementos de entrada da rede.

S_i - Número de neurônios em cada camada da rede neural.

TF_i - Função de transferência da camada i . Normalmente adotada sigmoideal para todos os neurônios (*TANSIG*).

BTF – Função de treinamento de retropropagação (*backpropagation*) da rede. Normalmente adotada a função de treinamento de retropropagação Levenberg-Marquardt (*TRAINLM*)

BLF – Função de aprendizagem dos pesos sinápticos e *biases*. Normalmente adotada a função de aprendizagem dos pesos e *biases* do tipo gradiente de descida com momento (*LEARNGDM*).

PF - Função para medição do desempenho da rede. Normalmente adotada a função erro médio quadrático (*MSE*).

A função de treinamento *TRAINLM* – função de retropropagação Levenberg-Marquardt – usada no caso exemplo, apresenta a particularidade de calcular a matriz de sensibilidade da rede neural, determinada pelas derivadas parciais de todos os parâmetros livres W (pesos sinápticos e *biases*), relativamente a todos os conjuntos de treinamento $x(n)$ para $n=1, 2 \dots N$. Obtém-se assim uma matriz de derivadas parciais $W \times N$ que é a matriz Jacobiana J do *perceptron* de múltiplas camadas. Cada linha de jacobiana corresponde a um exemplo particular do conjunto de treinamento (Haykin 1999).

O conjunto de pesos sinápticos e *biases* da rede neural são ajustados de acôrdo com:

$$JJ = JX * JX$$

JX = Matriz Jacobiana da rede neural

E = Matriz de erros

$$JE = JX * E$$

I = Matriz identidade

μ = Valor de ajuste

dX = Matriz de ajuste dos parâmetros livres

$$dX = \frac{- (JJ + I * \mu)}{JE}$$

A função *newff* determina uma rede com N_1 camadas, neste momento ainda não treinada, com algoritmo de aprendizagem com gradiente de descida e momento agregado aos pesos sináptico e *biases*, estruturado com uma rede tipo *feed-forward* com algoritmo de treinamento de retropropagação Levenberg-Marquardt, apresentando uma rede inicializada através do processo de Nguyen-Widrow (Fausset 1994).

$net = train (net , P , T)$

onde:

net = Rede obtida da função anterior *newff*, onde os valores foram inicializados e a rede estruturada.

P = Conjunto de valores do vetor de entrada da rede a ser treinada.

T = Conjunto de valores desejados do vetor de saída, correspondendo aos valores de entrada.

Essa função utiliza a função de treinamento atribuída pela função *newff* – no caso exemplo TRAINLM – retornando após 100 ciclos a rede treinada e o correspondente índice de performance MSE.

$y = sim (net , P)$

net = Rede treinada da função *net* de treinamento.

P = Valores de entrada na forma de vetores, a partir dos quais é desejada a obtenção de prognósticos de valores futuros.

Essa função retorna para cada conjunto de valores de entrada, na mesma dimensão para a qual a rede foi constituída, um valor futuro, correspondendo na sequência ao último valor da entrada.

4.4 - Resultados Obtidos

O caso estudado é uma série temporal das variações da cotação de fechamento da bolsa de valores de São Paulo, índice BOVESPA, ocorrido em um período do ano de 2000.

O gráfico representativo desta variação está apresentado na Figura 34.

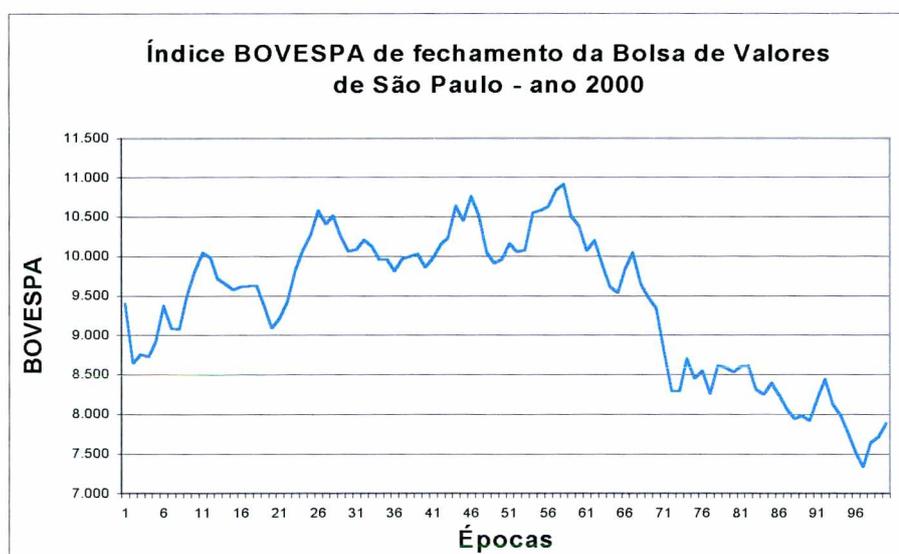


Figura 34 – Índice BOVESPA

Para a aplicação do método é recomendável a normalização dos valores da série em estudo, considerando que na utilização da função de ativação sigmoidal para valores muito elevados da série, a derivada da função não apresenta valores significativos, dificultando a convergência do procedimento de treinamento da rede.

O gráfico representativo dos valores normalizados, que para a mesma série tem a mesma forma que o dos valores atuais, está apresentado na Figura 35 e foi obtido da série original com a utilização da equação [4.3-8] para uma sucessão de 124 valores da série temporal em estudo.

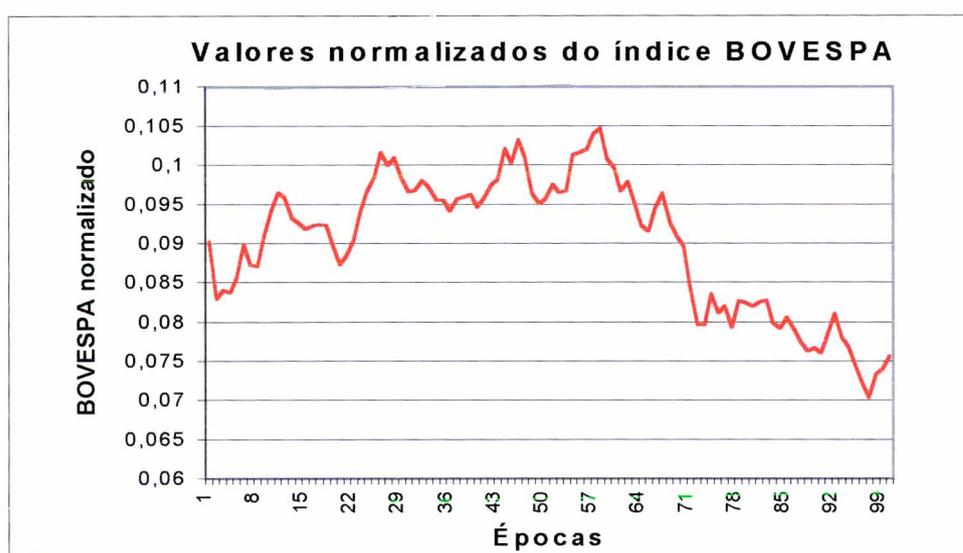


Figura 35 – Valores normalizados do índice BOVESPA

Para possibilitar uma avaliação da aplicação de redes neurais para o ajuste de valores de séries temporais, foram geradas várias arquiteturas diferentes de redes, sempre com a aplicação do algoritmo *backpropagation* e o algoritmo de treinamento Levenberg-Marquardt.

Na tabela a seguir, estão apresentadas as arquiteturas que foram estudadas, e cujos resultados serão apresentados na sequência.

Arquiteturas estudadas no caso exemplo																
Camada de entrada	2				4				6				8			
Camada oculta	2	4	6	8	2	4	6	8	2	4	6	8	2	4	6	8
Camada de saída	1				1				1				1			

A organização da apresentação dos resultados dar-se-á na forma como foi apresentada na tabela anterior, ou seja, o primeiro grupo com um conjunto de dois valores na camada de entrada combinados com a arquitetura da camada oculta, onde sequencialmente analisar-se-á a arquitetura com dois, quatro, seis e oito neurônios, sendo que todas as redes apresentam-se com um neurônio na camada de saída.

4.4.1 – Comportamento das redes neurais com a alteração da camada oculta

Neste item analisar-se-á o comportamento das redes neurais, no que diz respeito à capacidade de treinamento e de prognóstico de valores futuros, mantendo-se constante o conjunto de vetores de entrada.

4.4.1.1 – Rede com conjunto de entrada composto por dois valores

A Figura 36 apresenta a série de valores originais do índice Bovespa com valores normalizados e o resultado da simulação da rede 2 x 8 x 1, após treinada conforme modelagem descrita no item 4.3. Essa configuração apresentou um valor de $MSE = 4,34555E-6$, para os 100 ciclos de treinamento, menor dentre todas as simulações com vetor de entrada com dois valores.

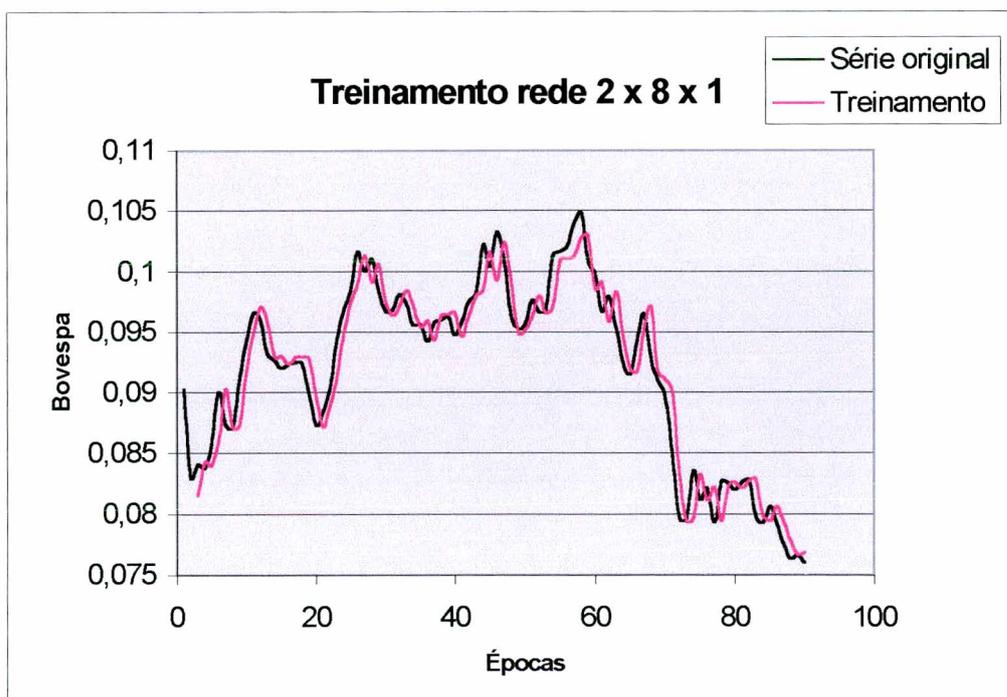


Figura 36 – Resultado do treinamento da rede 2 x 8 x 1

A Figura 37 apresenta um pequeno trecho do ajuste efetuado pela rede treinada $2 \times 8 \times 1$, onde se pode observar que há um deslocamento de uma época entre os valores originais da série e os valores obtidos no treinamento. Na mesma Figura está sendo representada a variação dos valores obtidos em treinamento deslocados de uma época, compensando a defasagem mencionada anteriormente, onde se pode aferir uma maior aderência entre as curvas original e de treinamento.

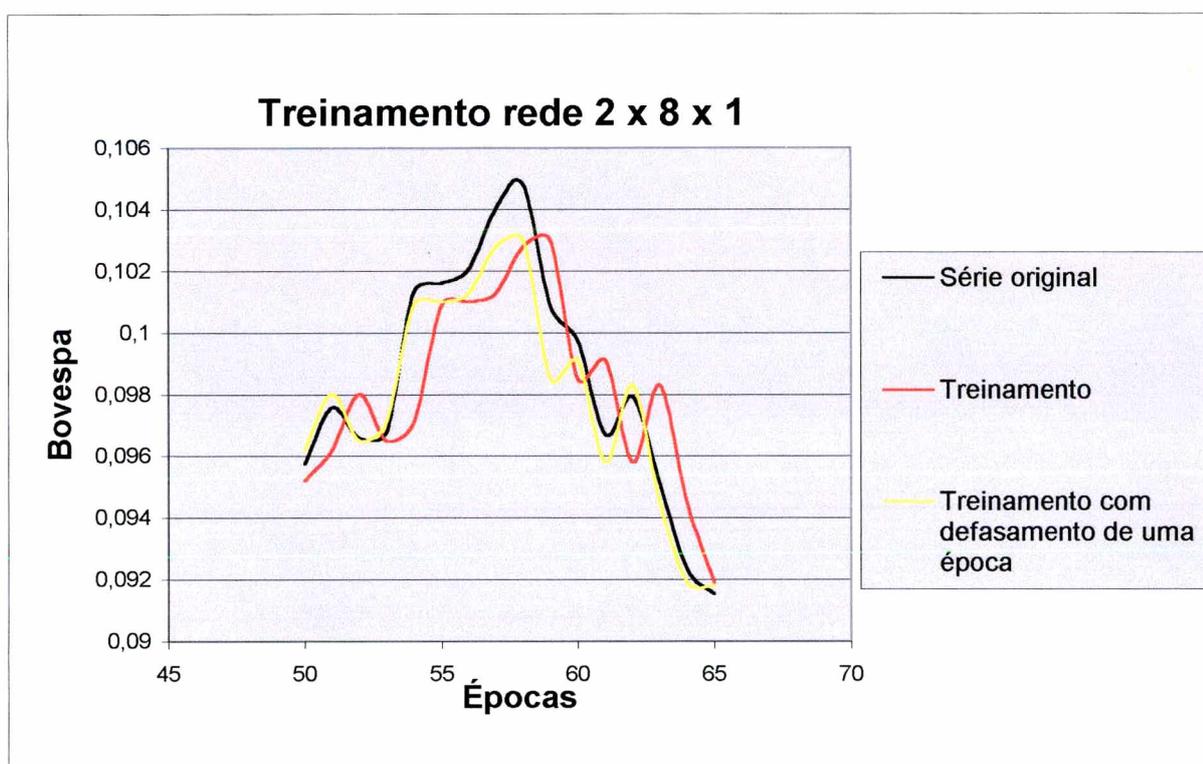


Figura 37 – Resultado de trecho do treinamento da rede $2 \times 8 \times 1$

A relação entre os valores de treinamento e o valor original no ponto de época 58 (maior valor observado no trecho para a série original), é de 0,9809 (linha vermelha da Figura) tomando-se o valor obtido na simulação da rede treinada e 0,9828 (linha amarela da Figura) tomando-se o valor com deslocamento de uma época.

Na Figura 38 está representada graficamente a variação dos valores do erro médio quadrático MSE, para as redes estudadas com vetor de entrada com dois valores. A rede que apresentou melhor desempenho, para o critério do erro médio quadrático, é a configuração $2 \times 8 \times 1$, cujo resultado está apresentado na Figura 36.



Figura 38 – Valores de MSE para as redes com conjunto de dois valores na camada de entrada

A verificação da capacidade apresentada pela rede em representar o conjunto de valores utilizados para o treinamento, é efetuada com o mesmo vetor p de entrada usada naquela fase, obtendo-se os valores treinados que podem ser comparados com os valores originais da série em estudo, permitindo a determinação do valor MSE que indica o índice de performance da rede.

Uma vez verificada a capacidade de generalização da rede através da capacidade de representar os valores utilizados no treinamento, pode-se utilizá-la para determinar valores futuros, ou seja, prognosticar um valor além da época presente.

A metodologia utilizada para efetuar o prognóstico foi a mesma utilizada para verificar a capacidade de treinamento obtido pela rede, utilizando um processo de simulação a partir do último elemento a_n do conjunto de valores utilizados para o treinamento. Assim para uma rede com vetor de entrada formado com dois valores, o vetor pp representado a seguir é o vetor de entrada da rede, o vetor yp a saída determinada pela rede e tp o vetor representativo do conjunto de valores conhecidos da série original, e que serão utilizados para verificar o desempenho da rede para prognóstico de valores futuros.

$$pp = \begin{Bmatrix} a_{n-1} & a_n \\ a_n & a_{n+1} \\ \dots & \dots \\ a_{N-2} & a_{N-1} \end{Bmatrix} \quad yp = \begin{Bmatrix} x_{n+1} \\ x_{n+2} \\ \dots \\ x_N \end{Bmatrix} \quad tp = \begin{Bmatrix} a_{n+1} \\ a_{n+2} \\ \dots \\ a_N \end{Bmatrix}$$

A Figura 39 mostra os resultados obtidos para prognóstico dos valores futuros, determinados conforme procedimento descrito anteriormente. Os valores prognosticados apresentam-se defasados, de uma época, em relação ao conjunto de valores da série original. Apesar da rede $2 \times 8 \times 1$ ter apresentado a melhor performance na fase de treinamento, tal fato não se verifica quando da sua utilização para prognóstico de valores futuros

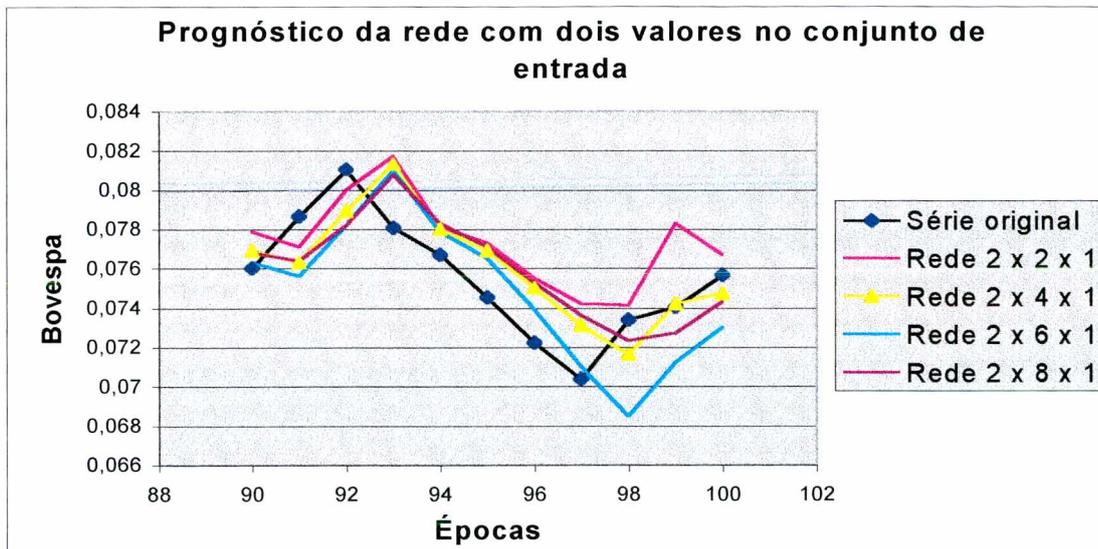


Figura 39 – Valores de prognóstico para redes com dois valores na camada de entrada

A Figura 40 apresenta os valores de MSE calculados com base nos valores prognosticados (a partir da época 91), onde se pode observar claramente o melhor desempenho da rede $2 \times 4 \times 1$.

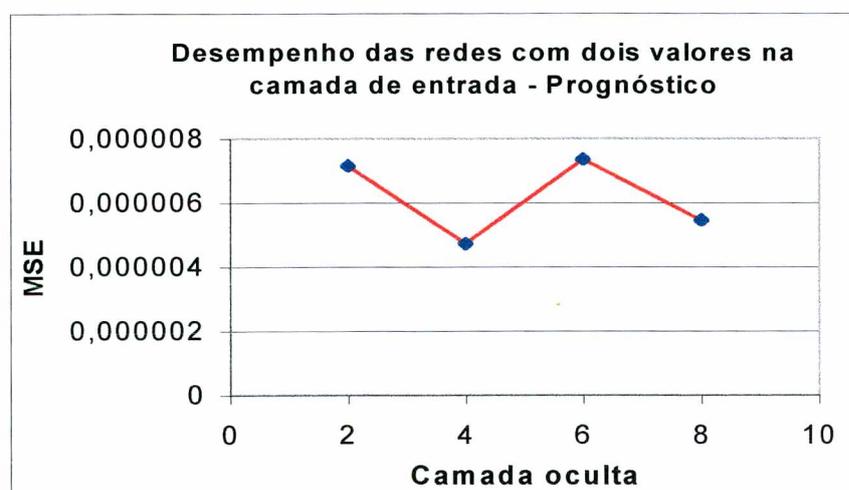


Figura 40 – Desempenho das redes com dois componentes na camada de entrada

4.4.1.2 - Rede com conjunto de entrada composto por quatro valores

Esta rede repete a mesma quantidade de neurônios da camada oculta que do caso anterior, somente que nesse caso a camada de entrada é composta por conjuntos de valores de quatro elementos. A metodologia aplicada segue os mesmos procedimentos já conhecidos.

A Figura 41 apresenta os resultados de treinamento simulado da rede $4 \times 8 \times 1$ treinada utilizando o vetor de entrada usado na fase de formação da rede, que apresentou um $MSE=4.03775E-6$, menor valor entre todos aqueles processados com quatro elementos na camada de entrada.

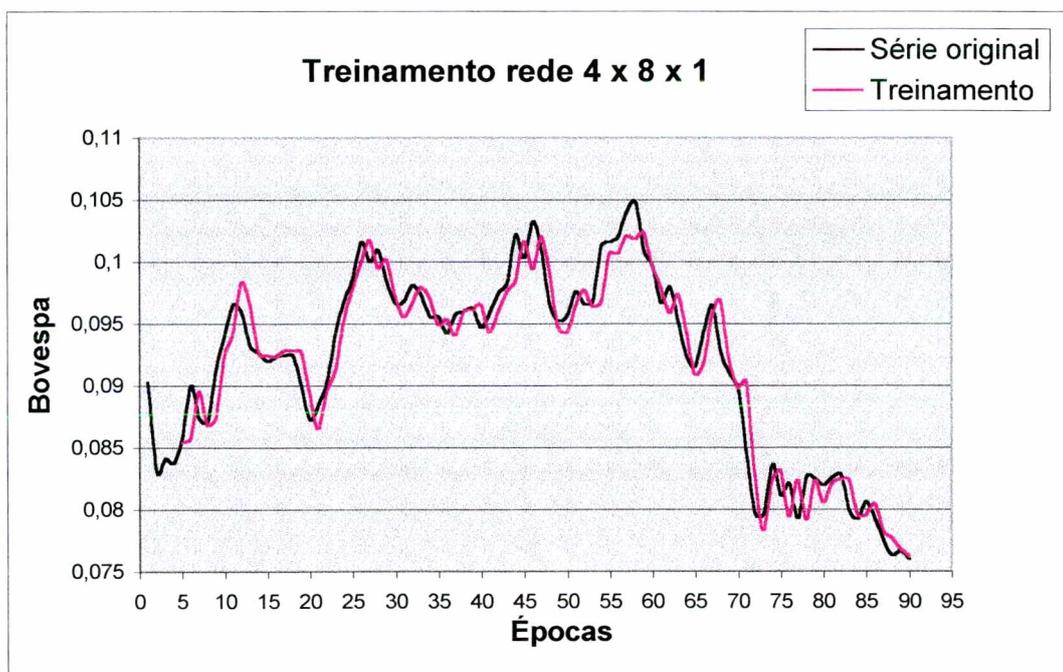


Figura 41 – Resultado do treinamento da rede $4 \times 8 \times 1$

Para melhor visualizar a aderência do treinamento com a série original, far-se-á a análise de um pequeno trecho da rede $4 \times 8 \times 1$, que está sendo apresentado na Figura 42.

A relação calculada entre o valor obtido no treinamento e valor original da série em estudo, para o ponto de época igual a 58, apresenta um resultado igual a 0,971 e utilizando o valor da série de treinamento deslocada de uma época, minimizando o efeito de defasagem, obter-se-á uma relação igual a 0,976. Comparando com o caso anterior foram obtidos valores menos favoráveis neste caso com uma camada de entrada com quatro elementos.

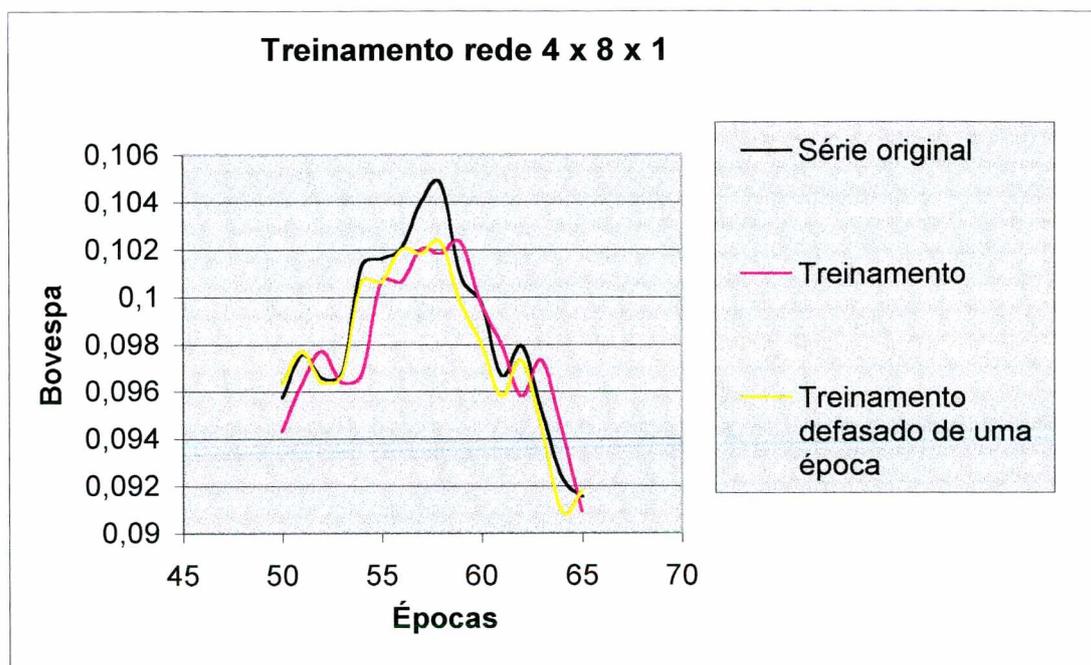


Figura 42 – Resultado de trecho do treinamento da rede 4 x 8 x 1

A Figura 43 apresenta os parâmetros MSE encontrados no treinamento de redes com quatro elementos de entrada, onde o valor para a rede 4 x 8 x 1 se apresenta como menor de todos (MSE = 4,03775E-6)

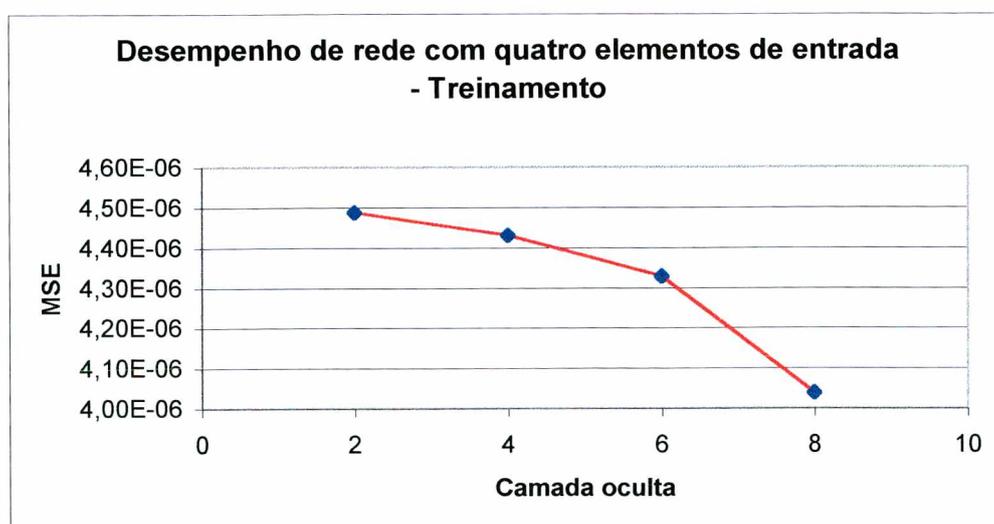


Figura 43 - Valores de MSE para as redes com conjunto de quatro valores na camada de entrada

Efetuada-se o treinamento da rede, o passo seguinte é a verificação do comportamento das arquiteturas processadas na simulação de valores futuros.

Os prognósticos das redes processadas com quatro elementos na camada de entrada estão apresentados na Figura 44, onde se pode verificar que as redes $4 \times 2 \times 1$ e $4 \times 4 \times 1$ apresentam resultados melhores de valores futuros que as outras duas configurações.

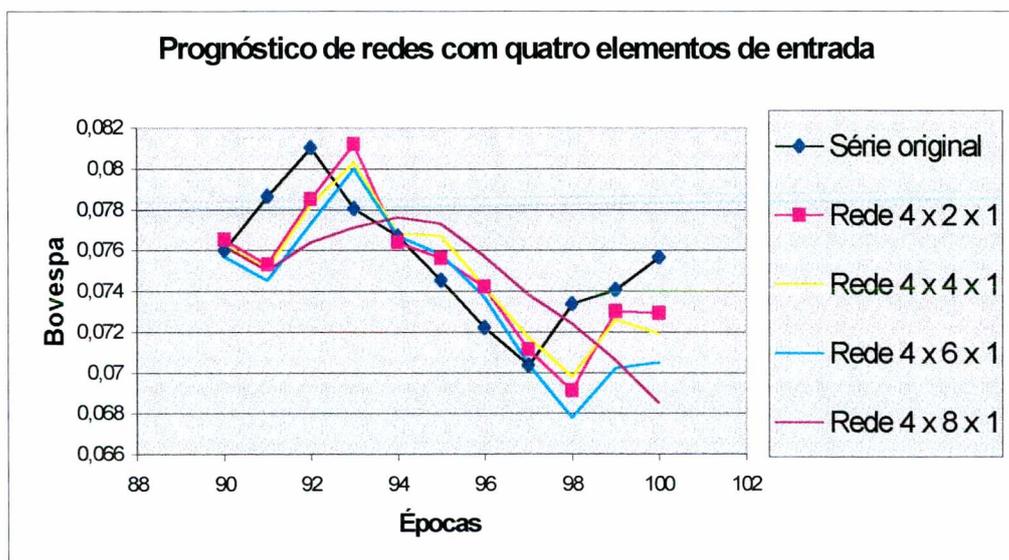


Figura 44 – Valores de prognóstico para redes com quatro valores na camada de entrada

A decisão sobre qual o melhor comportamento para a determinação de valores futuros poderá ser obtida da Figura 45 que compara valores de MSE para o conjunto de prognóstico e para cada arquitetura de camada oculta.

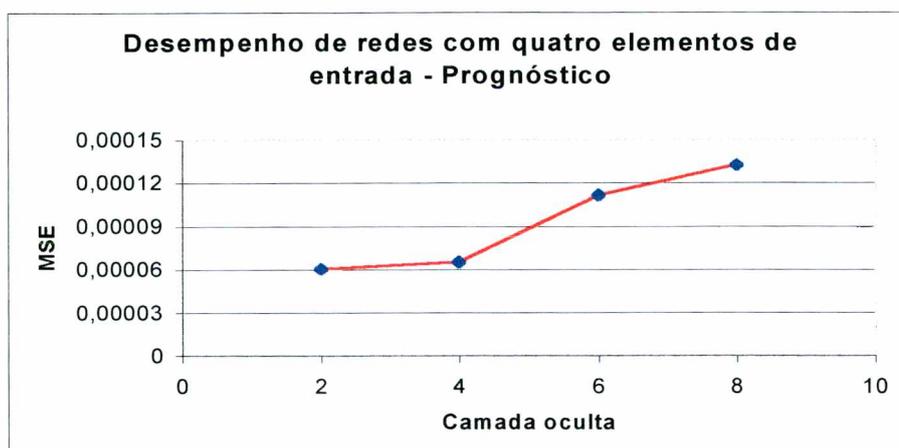


Figura 45 - Desempenho das redes com quatro componentes na camada de entrada

4.4.1.3- Rede com conjunto de entrada composto por seis valores

Neste caso adotou-se um conjunto de entrada composto por vetores com seis elementos, caracterizando a camada de entrada, e processando quatro casos efetuando-se alterações na camada oculta em dois, quatro, seis e oito elementos. A camada de saída permanece imutável com um neurônio.

A rede que apresentou uma melhor performance ($MSE= 4,04035E-6$) na fase de treinamento é constituída pela rede $6 \times 4 \times 1$, cujo resultado da fase de treinamento está apresentada na Figura 46 e, para melhor analisar os aspectos de aderência, a Figura 47 reproduz em escala maior um trecho escolhido conforme já efetuado para os casos anteriores.

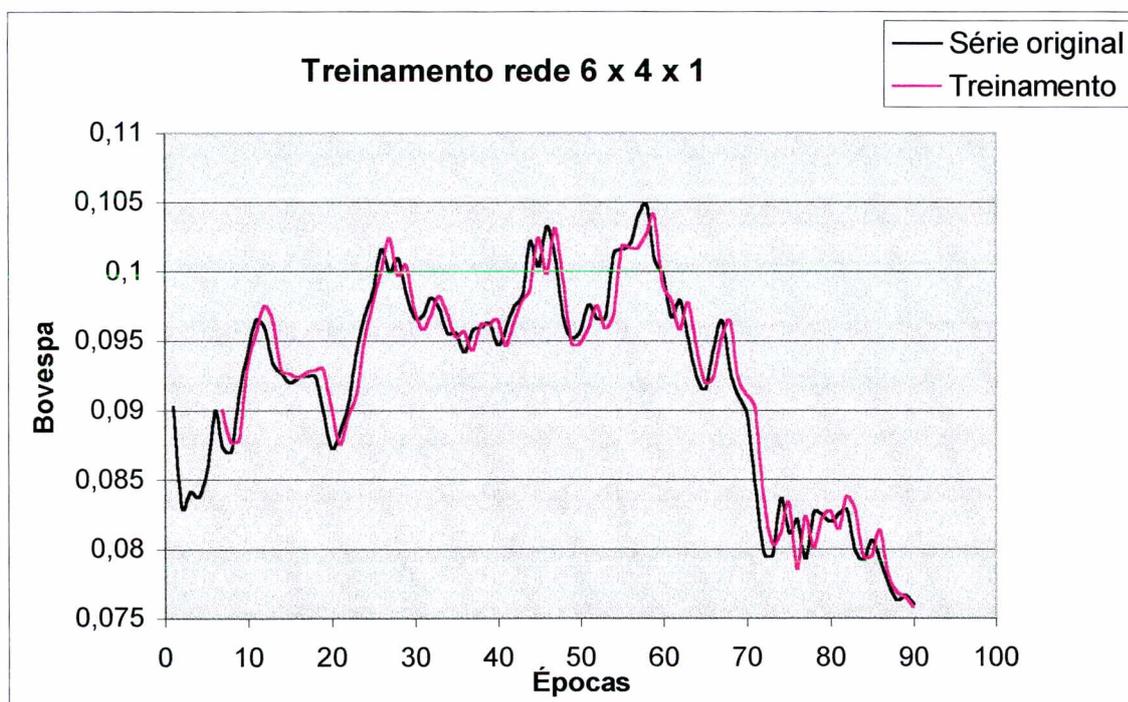


Figura 46 – Resultado do treinamento da rede $6 \times 4 \times 1$

Pela Figura 47 pode-se verificar que a rede continua apresentando um comportamento homogêneo, no que diz respeito ao deslocamento de uma época entre o valor da série original e os valores obtidos da simulação para verificar o grau de generalização da rede.

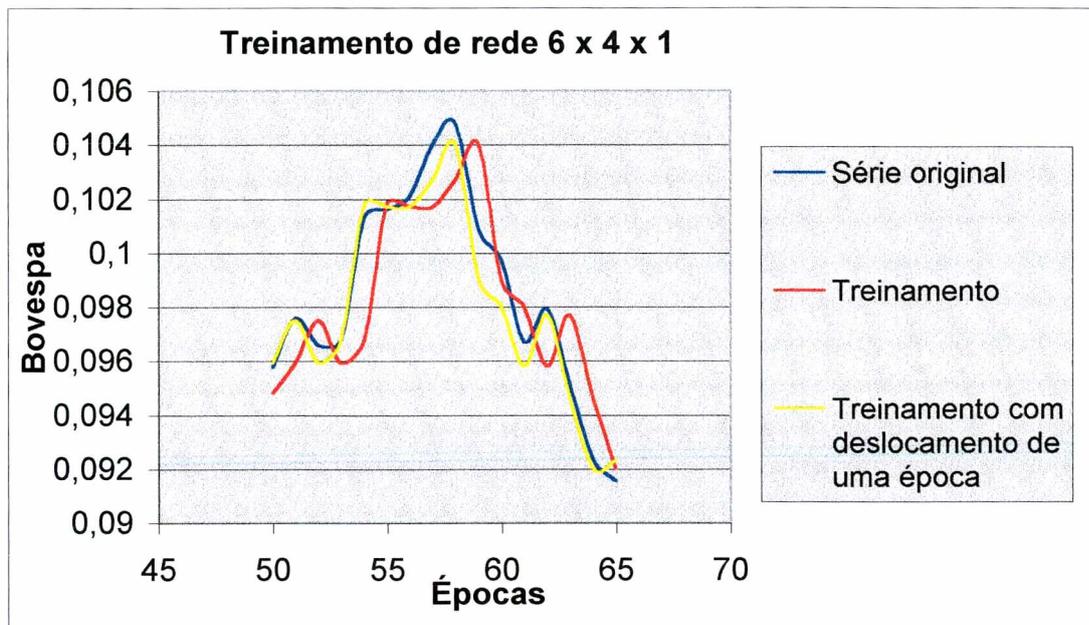


Figura 47 – Resultado de trecho do treinamento da rede 6 x 4 x 1

O valor máximo da série original ocorre no ponto de época 58 escolhido para servir de ponto de análise de valores. Para esta rede a relação entre o valor de treinamento e o correspondente da série histórica para o ponto considerado vale 0,98, sendo que deslocando em uma época a curva de treinamento, essa relação aumenta para 0,992. O fator desempenho das redes em estudo corresponde ao valor MSE (*Mean Square Error*), que para o caso das redes que apresentam arquiteturas com seis elementos de entrada, está mostrado na Figura 48, onde pode-se observar que a rede 6 x 4 x 1 apresentou o melhor desempenho.

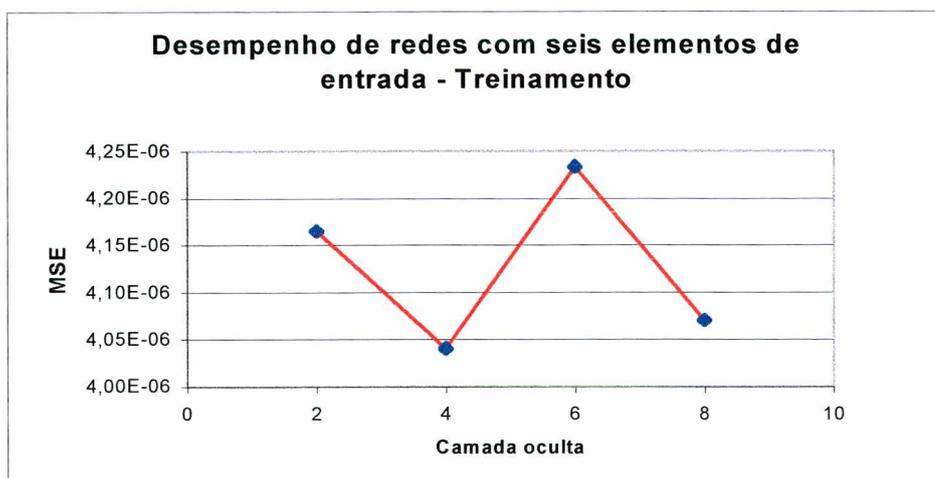


Figura 48 - Valores de MSE para as redes com conjunto de seis valores na camada de entrada

Apesar da rede apresentar um valor baixo de MSE, mostrando tratar-se de uma rede capaz de generalizar o comportamento frente a valores novos, não há garantia de que os resultados de prognósticos venham a apresentar tais características. Na Figura 49 estão apresentados os resultados da simulação de valores futuros.

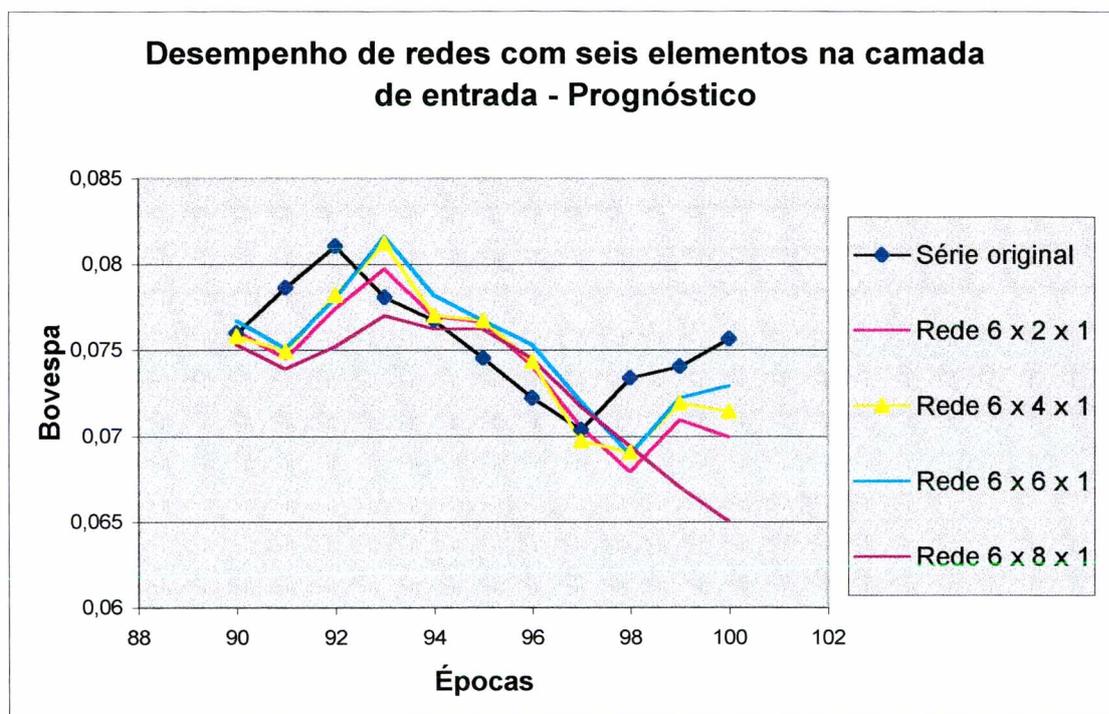


Figura 49 – Valores de prognóstico para redes com seis valores na camada de entrada

Os valores prognosticados com redes, cujas entradas têm seis elementos apresentam um desempenho comparativo pior do que aqueles observados para as redes com dois e quatro elementos na camada de entrada, o que pode ser melhor verificado pela variação do valor de MSE na Figura 50.

Dentre as redes estudadas com seis elementos na camada de entrada a que apresentou o menor valor de erro – $MSE = 8,33092E-5$, foi a arquitetura $6 \times 4 \times 1$.

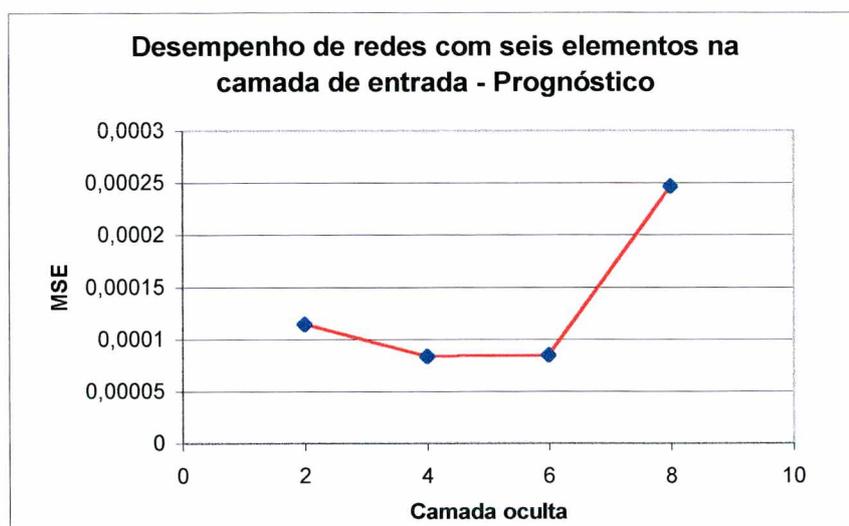


Figura 50 - Desempenho das redes com seis componentes na camada de entrada

4.4.1.4- Rede com conjunto de entrada composto por oito valores

Da mesma forma que foi trabalhada anteriormente apresentar-se-ão os resultados obtidos com redes em que a camada de entrada está constituída por oito elementos, que receberão conjuntos de vetores com o mesmo número de componentes.

A rede desse tipo que melhor desempenhou foi a $8 \times 8 \times 1$, cujos resultados da fase de treinamento estão apresentados nas Figuras 51 e o trecho detalhado para análise na 52.

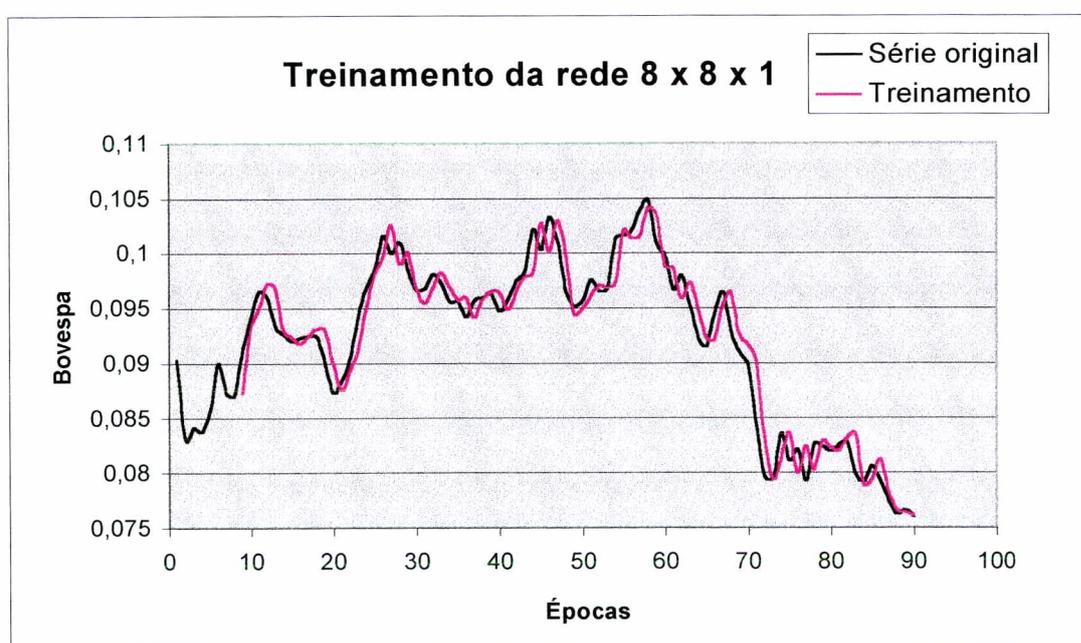


Figura 51 – Resultado do treinamento da rede $8 \times 8 \times 1$

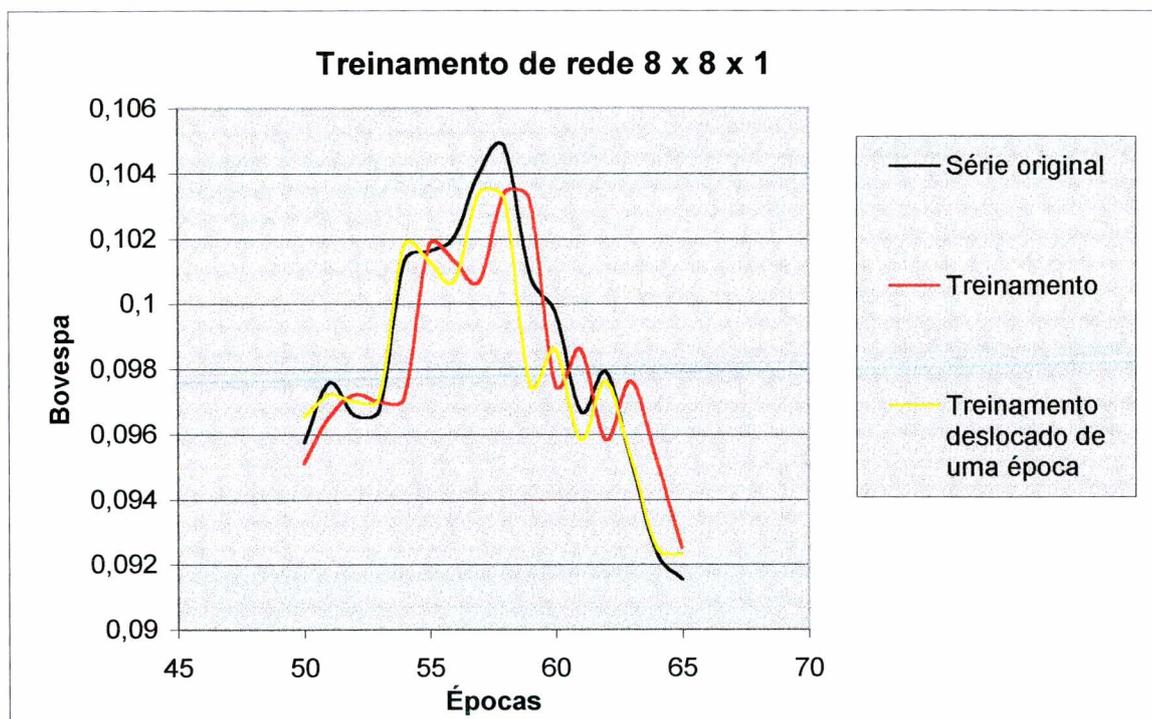


Figura 52 – Resultado de trecho do treinamento da rede 8 x 8 x 1

Na Figura 52 pode-se observar que a curva representativa do treinamento deslocado de uma época apresenta uma aproximação de valores maior em relação à série original do que a sucessão de valores obtidos diretamente da simulação após a rede treinada.

A relação entre o valor máximo da série original e os correspondentes valores obtidos do treinamento e da compensação da defasagem são respectivamente 0,987 e 0,985, estando muito próximos entre si e relativamente ao valor original. Mesmo assim, como nos casos anteriores, o deslocamento da série obtida após o treinamento, de uma época, tem melhor aderência aos valores originais.

A Figura 53 apresenta o desempenho das redes estudadas que possuem oito elementos de entrada, tendo-se para essa situação que o melhor desempenho é apresentado pela arquitetura 8 x 8 x 1 com um valor $MSE = 3,97786E-6$.

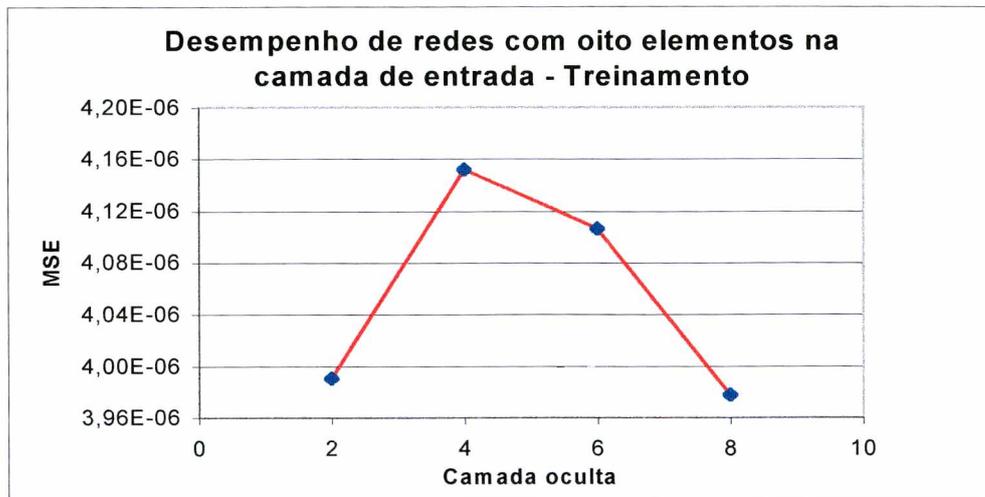


Figura 53 - Valores de MSE para as redes com oito valores na camada de entrada

A Figura 54 apresenta o resultado obtido efetuando-se o prognóstico para valores futuros, tendo-se observado para a rede $8 \times 2 \times 1$ e a rede $8 \times 8 \times 1$ um comportamento melhor do que as redes com 4 e 6 neurônios na camada oculta. É visível um comportamento já observado nos demais casos processados, qual seja, o deslocamento correspondente em uma época das séries obtidas na simulação do prognóstico e da série original.

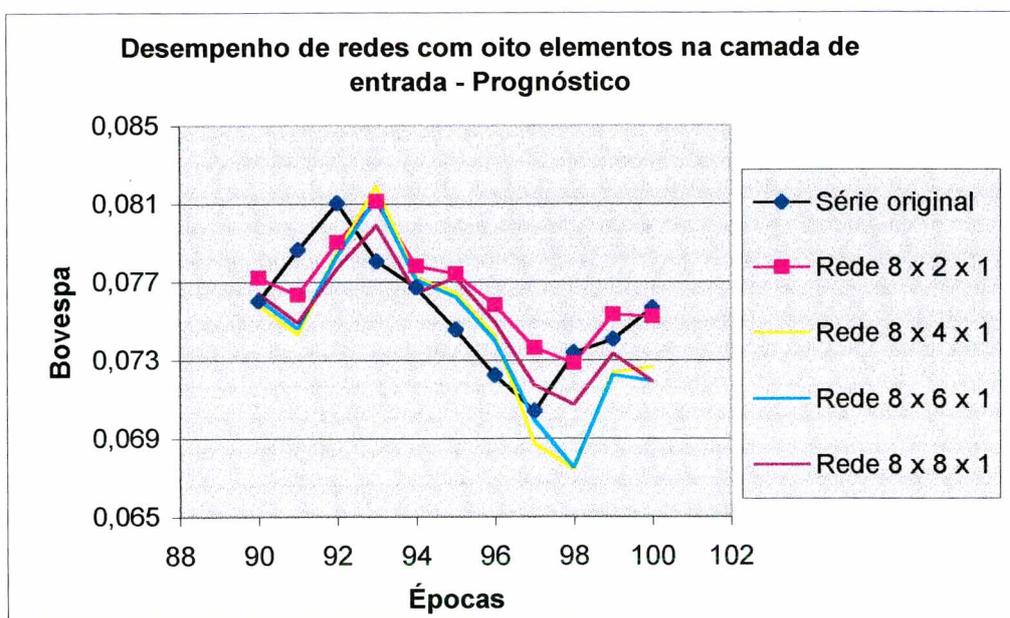


Figura 54 – Valores de prognóstico para redes com oito valores na camada de entrada

O desempenho calculado das séries prognosticadas da Figura 54, está apresentado na Figura 55, onde a rede $8 \times 8 \times 1$ teve o melhor desempenho com $MSE = 5,35998E-5$, valor muito além daqueles apresentados por outras arquiteturas estudadas.

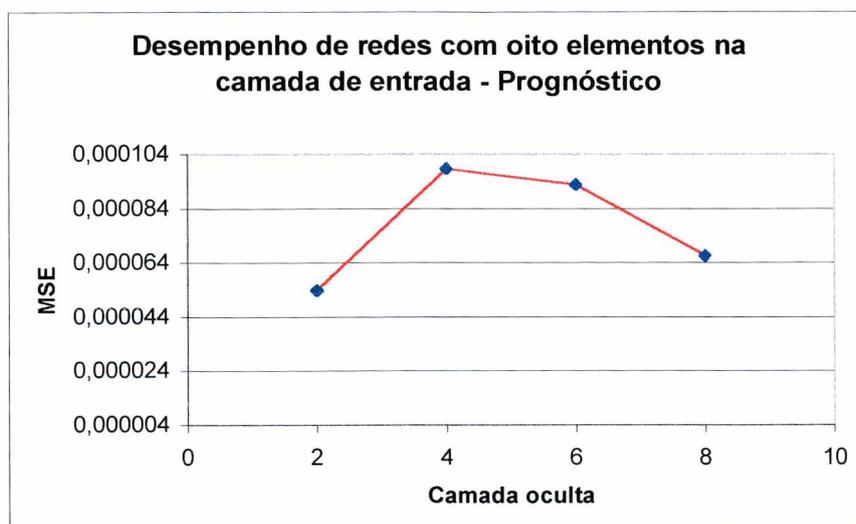


Figura 55 - Desempenho das redes com oito componentes na camada de entrada

4.4.2- Estudo comparativo dos casos processados com base no valor de MSE

Neste item analisar-se-á o comportamento das redes a partir de um estudo comparativo dos valores obtidos para MSE nas diversas configurações processadas, visando determinar a melhor arquitetura dentre as utilizadas para prognosticar os valores futuros da série escolhida no caso exemplo.

No item anterior foram apresentados, nas figuras correspondentes ao desempenho dos valores prognosticados, todos os valores MSE dos casos estudados no exemplo.

Para permitir uma análise visual do comportamento de cada rede relativamente aos valores obtidos na fase de simulação de prognósticos futuros, apresentar-se-á nas figuras que seguem a variação dos índices MSE dos valores obtidos na fase de prognóstico, mantendo-se para cada figura o vetor de entrada constante e variando-se o tamanho da camada oculta. Dessa forma pode-se obter a relação entre a quantidade de elementos da camada oculta e a capacidade de prognosticar da rede treinada.

A Figura 56 apresenta os valores encontrados para MSE na fase de treinamento e na fase de prognóstico de valores futuros, para uma rede com camada de entrada com dois elementos, variando-se a quantidade de elementos da camada oculta (utilizados 2, 4, 6 e 8)

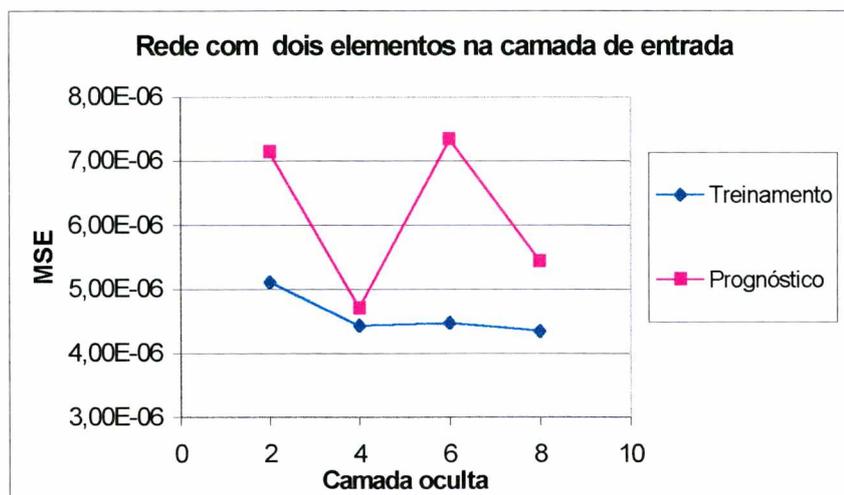


Figura 56 – Desempenho da rede com dois elementos na camada de entrada

Para uma rede com as características utilizadas no estudo, e para a série temporal do índice BOVESPA do caso exemplo, o comportamento se apresenta melhor para uma camada oculta de oito elementos na fase de treinamento e melhor para uma camada oculta com quatro elementos na fase de prognóstico. Globalmente a rede que melhor resultado apresentou está constituída com quatro elementos na camada oculta, onde os valores de MSE para ambas as fases estão bastante próximos.

A Figura 57 apresenta graficamente valores com as características da anterior, porém, para redes com camadas de entrada com quatro elementos.

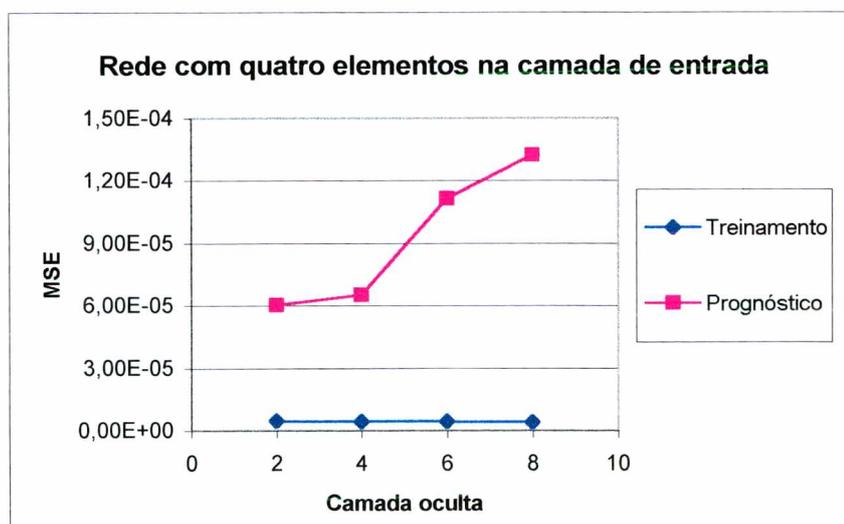


Figura 57 – Desempenho da rede com quatro elementos na camada de entrada

O desempenho da fase de treinamento apresenta-se uniforme não experimentando grandes variações com o número de elementos da camada oculta. Na fase de prognóstico o desempenho diminui com o acréscimo de elementos na camada oculta, apresentando para dois elementos o melhor desempenho global da rede com quatro elementos na camada de entrada .

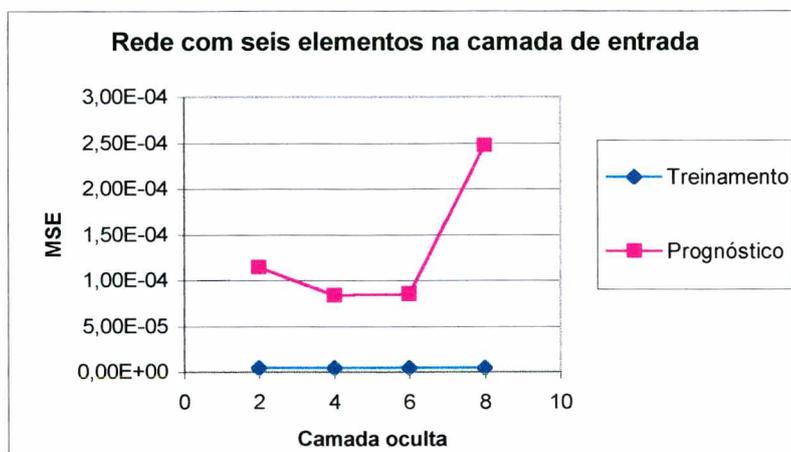


Figura 58 – Desempenho da rede com seis elementos na camada de entrada

O desempenho na fase de treinamento não apresenta alterações sensíveis com o número de elementos da camada oculta. Na fase de prognóstico, as redes com quatro e seis elementos ocultos apresentam aproximadamente o mesmo desempenho, que podem representar globalmente o melhor resultado para esta rede. A Figura 59 apresenta o resultado de estudo de desempenho das redes com oito elementos na camada de entrada.

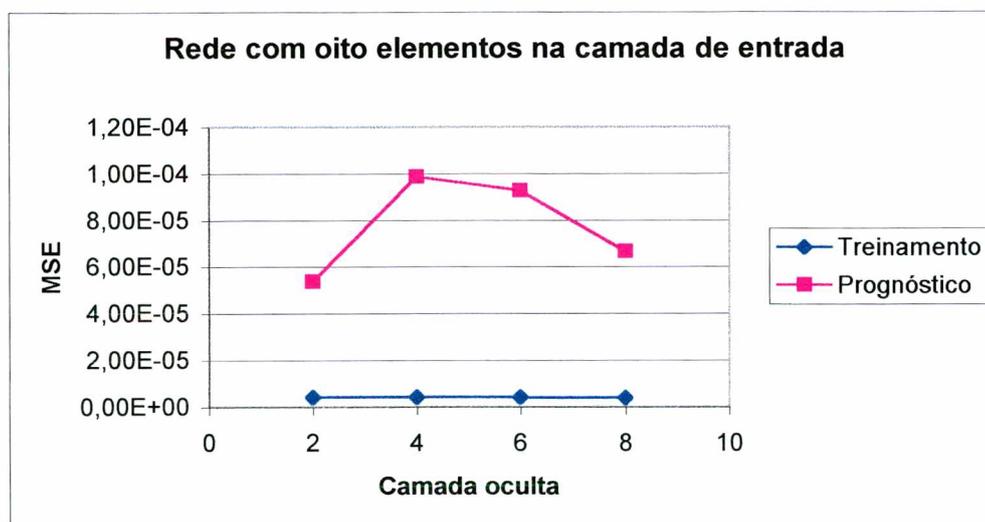


Figura 59 – Desempenho da rede com oito elementos na camada de entrada

O desempenho da fase de treinamento é uniforme, sofrendo variações não observáveis com a alteração da arquitetura da camada oculta da rede. Na fase de prognóstico o problema consiste em dado um conjunto de valores novos na camada de entrada e, com a rede treinada anteriormente com valores históricos obter os resultados para épocas futuras, resultando como é observável nas figuras 56, 57, 58 e 59 valores de MSE bem maiores que os da fase de treinamento, obtendo-se o melhor resultado para dois elementos na camada oculta da rede.

Na Figura 60 apresentam-se graficamente todos os resultados de desempenho observados na fase de treinamento, permitindo avaliar, dentre todas as configurações, aquela que melhor representa a massa de dados ocorridos no período.

O melhor desempenho foi apresentado pela rede $8 \times 8 \times 1$ vindo na sequência a rede $8 \times 2 \times 1$.

As redes compostas com seis e oito elementos de entrada, e para qualquer uma das arquiteturas da camada oculta, apresentaram resultados do desempenho MSE praticamente idênticos, em torno de $4E-6$. A rede com quatro valores de entrada apresentou um desempenho menor e com dois valores de entrada o pior desempenho entre todos os casos processados.

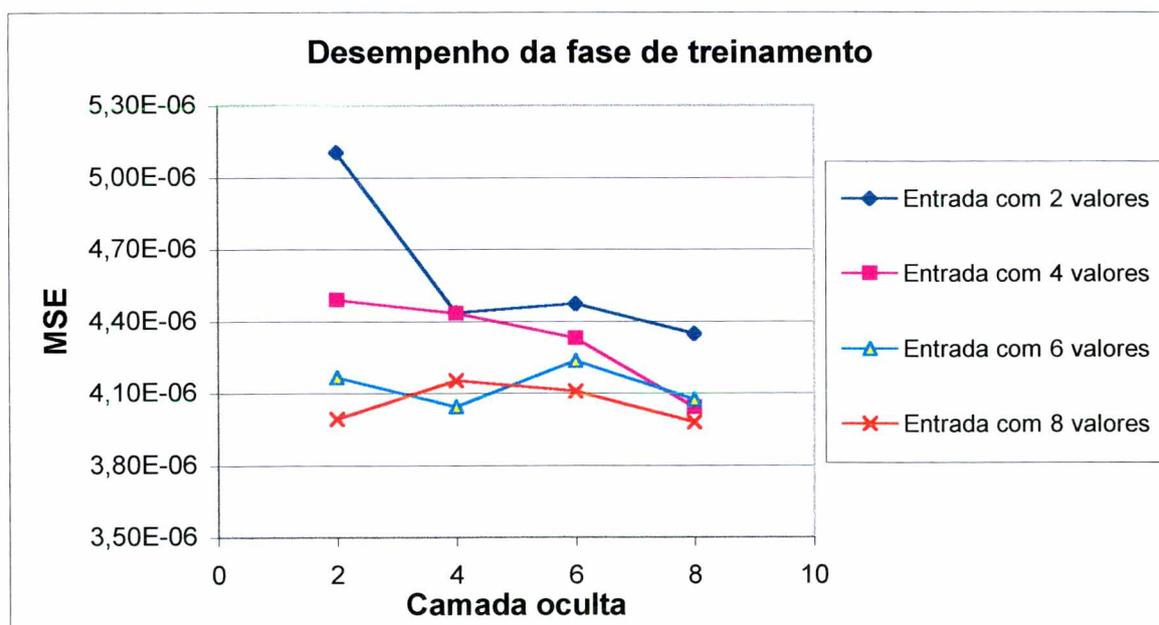


Figura 60 – Desempenho da fase de treinamento das redes do estudo de caso

Outro fator importante a ser comparado é o desempenho determinado a partir dos valores prognosticados, qual seja, o desempenho da fase de prognóstico, que está apresentado na Figura 61.

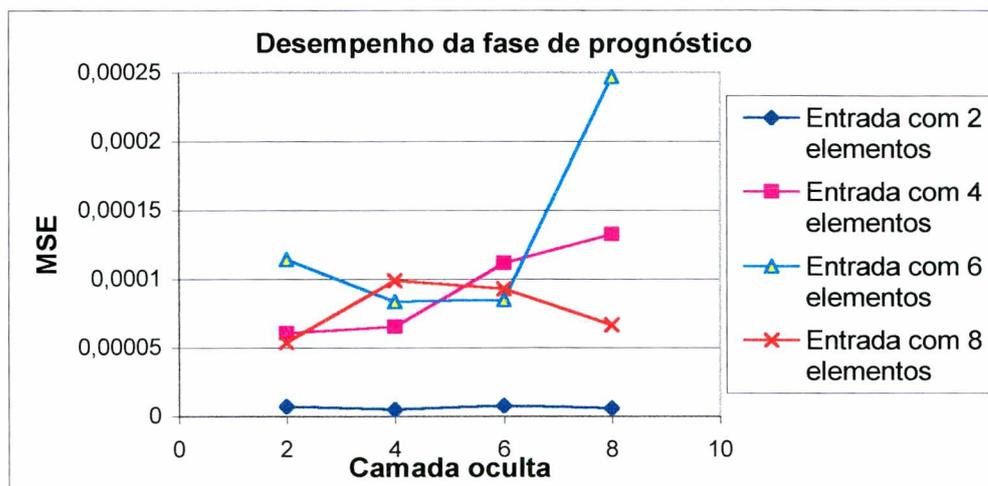


Figura 61 – Desempenho da fase de prognóstico das redes do estudo de caso

Pela Figura 61 pode-se observar que as redes com dois elementos na camada de entrada apresentaram melhor valor de MSE quando comparados com os valores obtidos para camadas de entrada com 4, 6 e 8 elementos de entrada.

Efetuada-se a comparação entre os valores prognosticados para redes com dois elementos na camada de entrada, porém na forma de diferenças percentuais, ter-se-á os gráficos da Figura 62.

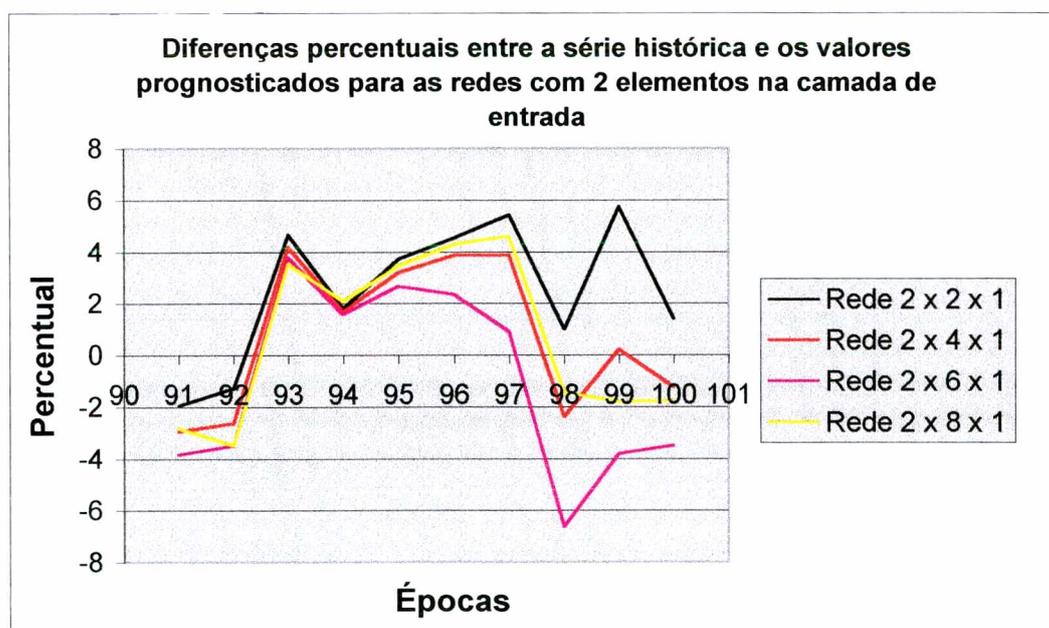


Figura 62 – Diferenças percentuais de valores prognosticados para as redes com dois elementos de entrada

Pela Figura 62 pode-se observar que a rede que apresenta menor valor percentual ponto a ponto é a rede $2 \times 4 \times 1$. Na Figura 63 apresenta-se graficamente a diferença percentual entre valores históricos e simulados, tanto para a fase de treinamento como para a fase de prognóstico, da rede com arquitetura $2 \times 4 \times 1$.

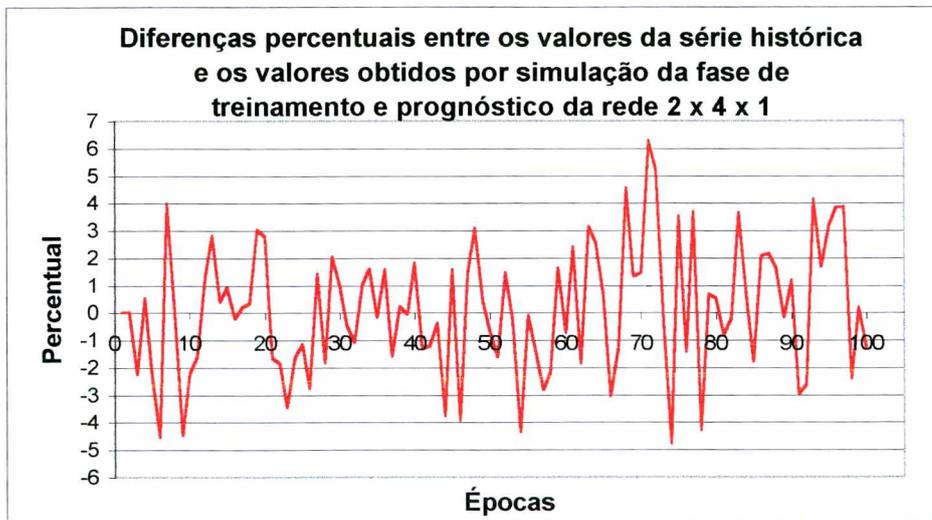


Figura 63 – Diferenças percentuais entre valores históricos e simulados da rede $2 \times 4 \times 1$

A arquitetura de rede que tem dois elementos na camada de entrada, independente da quantidade de neurônios utilizados na camada oculta, apresentou os melhores resultados de desempenho, podendo-se atribuir para a rede $2 \times 4 \times 1$ o melhor desempenho global por apresentar $MSE_{\text{treinamento}} = 4,4344E-6$ e $MSE_{\text{prognóstico}} = 4,70482E-6$, conforme pode ser visto na Figura 64.

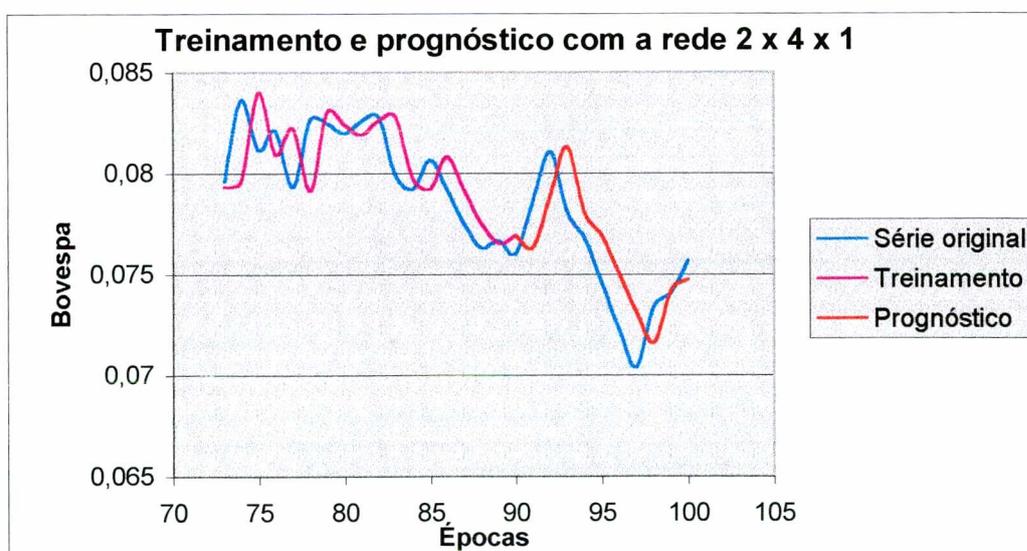


Figura 64 – Desempenho global da rede $2 \times 4 \times 1$

Capítulo V. Conclusões e Sugestões para Futuros Trabalhos

5.1 – Conclusões

De acordo com trabalhos efetuados por vários pesquisadores, as redes neurais, com suas diversas arquiteturas, podem representar os mais diversos fenômenos, apresentando soluções capazes de responder às indagações dos profissionais responsáveis pelo planejamento de demandas dos setores produtivos.

As redes neurais oferecem características inteligentes muito interessantes, tais como: aprendizagem, adaptação, tolerância a falhas e generalização. Assim, o ponto forte de aplicação está no reconhecimento de padrões, reconhecimento de caracteres e formas, estimativa de funções não lineares, previsões financeiras e controle de processos.

O controle de processos também pode ser considerado uma tarefa de reconhecimento de padrões quando a rede neural puder reconstruir modelos de conjuntos de dados, a partir da coleta de sinais de vários sensores e medidores em um processo.

Na ausência de modelos matemáticos de processos, as redes neurais podem utilizar um histórico de dados para construir modelos preditivos.

Depois de treinada, a rede neural pode se tornar um equipamento eletrônico equivalente, ou seja, um modelo que pode então prever as reações do processo a novas condições.

Entretanto, as redes neurais têm certos pontos fracos que devem ser considerados (Shaw e Simões 1999) :

- ◆ Não é possível se traçar a maneira pela qual uma rede neural chegou a uma determinada solução. Nesse sentido, as redes neurais são semelhantes a muitos especialistas humanos que podem expressar uma opinião sem na realidade ter uma explicação plausível para tal.
- ◆ Uma vez que não se pode olhar para o sentido dos pesos e conexões de uma rede neural, torna-se impossível se interpretarem as causas de um comportamento particular; não é possível também se mudar manualmente a estrutura de uma rede neural para se obter um comportamento desejado.
- ◆ A maioria dos métodos de treinamento não são completamente compreendidos, pois além de depender de um enfoque de tentativa e erro, há poucas regras de projeto a serem seguidas.

Por exemplo, a definição do número de neurônios, ou o estabelecimento de uma determinada faixa de convergência depende de experimentação com os dados.

- ◆ O tempo de treinamento não é previsível, podendo em alguns casos ser muito longo. Por essa razão, o retreinamento “on-line” em geral requerido para um sistema de auto-sintonia só pode ser aplicado para sistemas com dinâmica muito lenta.
- ◆ O tempo de execução depende do número de conexões, sendo aproximadamente proporcional ao quadrado do número de neurônios utilizados. Isso significa que apenas alguns nós a mais podem aumentar consideravelmente o tempo de execução. Todavia, *hardware* especializado para redes neurais está se tornando disponível para aplicações específicas.

Mesmo que as redes neurais apresentem pontos fracos quando da sua aplicação em problemas relacionados com o reconhecimento de padrões, trata-se de uma ferramenta que permite inúmeras variações na estruturação da arquitetura da rede, permitindo ao pesquisador buscar resultados que possam representar o desafio particular em estudo.

Assim, lembrando os objetivos iniciais propostos para a pesquisa, que eram a apresentação de uma ferramenta para determinar prognósticos de demandas mercadológicas capazes de representar séries temporais com características irregulares, e revelar resultados da utilização de arquiteturas diferentes de redes neurais, podemos concluir:

- ◆ Para que a rede possa apresentar resultados compatíveis com os valores da série temporal em estudo, é necessário que os dados sejam normalizados, para que as não-linearidades possam ser adequadamente manipuladas, através da função de não-linearidade geralmente sigmoideal uma vez que a função somente apresenta derivada significativa no intervalo $[-1;+1]$;
- ◆ No sistema computacional utilizado (MATLAB), não houve a ocorrência de problemas de convergência, nem durante o processo de aprendizagem, nem na fase de simulação com a rede treinada;

- ◆ Na fase de treinamento das diversas arquiteturas utilizadas, todas efetuadas com 100 ciclos do algoritmo *backpropagation*, o comportamento do erro foi uniforme, mantendo-se na ordem de $4.0E-6$;
- ◆ Conforme é possível observar nas Figuras 36, 41 e 46, que apresentam graficamente a série original e a série obtida por simulação com o conjunto de valores utilizados como vetor de entrada na fase de treinamento, existe uma aderência bastante grande entre os dois conjuntos de valores para todas as configurações, tendo-se observado somente um deslocamento de aproximadamente uma unidade de tempo (com a série calculada em atraso) entre os valores das duas séries.
- ◆ A relação entre um valor da série simulada e o correspondente da série original, para um ponto particular, mostrados nas Figuras 37, 42 e 47, manteve-se entre 0,97 e 0,98, mostrando que o erro, para os valores da rede treinada, situa-se na ordem de 2% a 3%. Portanto, as redes neurais, podem representar com bastante precisão séries temporais com comportamento não uniforme.
- ◆ O efeito de deslocamento de aproximadamente uma época, entre a série original e a série correspondente obtida pela simulação da rede treinada, mostrada nas Figuras 37, 42 e 47, não apresentou uma discrepância significativa nos valores, tendo-se encontrado para o mesmo ponto particular, relações variando entre 0,98 e 0,99, o que representa uma melhora na representação da série original de aproximadamente 1% do valor.
- ◆ O desempenho das redes treinadas obtido a partir dos valores de erro MSE mostrados nas Figuras 38, 43, 48, 53 e 60, não apresentam uma uniformidade, tendo-se observado que o pior desempenho nessa fase foi apresentado pela rede $2 \times 2 \times 1$ com um valor $MSE_{\text{treinamento}}=5.1E-6$. A rede com melhor desempenho nessa fase foi a $8 \times 8 \times 1$ com um valor $MSE_{\text{treinamento}}=3.98E-6$ representando 78% do maior erro observado.
Para as redes com a camada oculta com 2 neurônios observou-se que o acréscimo do número de elementos da camada de entrada resultou em um decréscimo no erro MSE calculado. Assim os erros calculados para a entrada com 2 elementos foi $MSE_{\text{treinamento}}=5.1E-6$, para 4 elementos $MSE_{\text{treinamento}}=4.48E-6$, para 6 elementos $MSE_{\text{treinamento}}=4.17E-6$ e para 8 elementos $MSE_{\text{treinamento}}= 3.99E-6$.

- ◆ O desempenho observado na fase de diagnóstico não acompanha o desempenho da fase de treinamento, lembrando que a rede que apresentou o menor valor de MSE naquela fase foi a rede 8 x 8 x 1.

Dentre as redes com dois elementos de entrada, aquela que apresentou o melhor resultado na fase de prognóstico foi a rede 2 x 4 x 1 com um valor $MSE_{\text{prognóstico}}=4.7E-6$, que pode ser observado visualmente nas Figuras 39 e 40. Dentre as redes com quatro elementos de entrada o melhor valor do erro foi $MSE_{\text{prognóstico}}=6.0E-5$ para a rede 4 x 2 x 1 conforme desempenho nas Figuras 44 e 45. Dentre as redes com seis elementos de entrada, a arquitetura 6 x 4 x 1 apresentou o melhor resultado para um erro $MSE_{\text{prognóstico}}=8.33E-5$ conforme mostrado nas Figuras 49 e 50. Para as redes com oito valores de entrada, desempenho das Figuras 54 e 55, a rede 8 x 2 x 1 apresentou o melhor prognóstico com um erro $MSE_{\text{prognóstico}}=5.36E-6$.

Comparando os valores da série original com os valores simulados do prognóstico, observa-se que nas redes com 4, 6, e 8 elementos de entrada há coincidência em dois valores entre as duas séries (Figuras 44, 49 e 54), sendo que na série com dois valores de entrada há somente um valor coincidente (Figura 39), no entanto, essa apresentou resultado global melhor, pois os demais pontos mantêm uma maior proximidade dos correspondentes pontos da série original, reduzindo o valor de MSE.

- ◆ Para o desempenho comparativo entre as fases de treinamento e de prognóstico pode-se recorrer às Figuras 56, 57, 58 e 59 onde estão representadas graficamente as variações do erro MSE para arquiteturas com 2, 4, 6 e 8 elementos da camada de entrada, respectivamente, em combinação com as variações da camada oculta estudadas nesse caso. Pelas figuras pode-se concluir que a rede que apresentou o melhor resultado, tanto na fase de treinamento como na fase de prognóstico, apresenta na sua camada de entrada um conjunto de vetores com dois elementos de cada vez, estando os valores de MSE situados na ordem de 10^{-6} , ao passo que nas demais situações estudadas o valor de MSE situa-se na ordem de 10^{-5} . Observando as Figuras 60 e 61, que mostram as comparações entre valores de MSE da fase de treinamento e da fase de prognóstico, respectivamente, podemos assumir a rede 2 x 4 x 1 como sendo aquela que melhor resultado apresentou globalmente nas duas fases. A rede que apresentou o pior desempenho global foi a arquitetura 6 x 8 x 1.
- ◆ Um elemento importante para efeito de verificação de desempenho é a obtenção de gráficos representativos das diferenças percentuais, que calculam ponto a ponto o percentual que

representa a diferença entre o valor obtido do processo de simulação e o valor histórico em relação ao próprio valor histórico (série original).

A Figura 62 apresenta o gráfico das diferenças percentuais relativamente ao prognóstico para as redes com dois elementos na camada de entrada, onde se pode observar que a rede que apresenta valores percentuais menores em relação aos valores da série original é a rede $2 \times 4 \times 1$ (lembrar que para um prognóstico exato os pontos do gráfico apresentam valor nulo). Mesmo apresentando o melhor desempenho, os pontos mais distantes apresentam uma diferença percentual na ordem de 4%, o que representa o erro no prognóstico para aquele ponto particular e para a rede treinada com os 90 pontos iniciais da série.

A Figura 63 apresenta o gráfico das diferenças percentuais relativamente à totalidade da rede $2 \times 4 \times 1$ incluindo valores do treinamento e da fase de prognóstico. Pela figura podemos observar que não há uma representação exata da sucessão de valores que constituem a série original, tendo-se obtido um valor de 6,3% no valor da diferença percentual para um determinado ponto específico da série.

- ◆ A representação total da série original em estudo, juntamente com os valores obtidos na simulação para obter a série da fase de treinamento e a simulação para os valores de prognóstico, permite visualmente constatar a capacidade de representação e generalização das redes neurais a partir de valores conhecidos do passado, bem como constatar que a partir desses valores é possível estender o processo para além do tempo presente, mostrando que com aquela arquitetura de rede e com aquele conjunto de algoritmos, é possível prognosticar para obter valores futuros da variável em estudo. Isso está mostrado na Figura 64, onde em uma mesma representação estão a série original, os valores simulados para a rede na fase de treinamento e os valores simulados com aquela rede para obter os valores futuros.

5.2 – Sugestões para futuros trabalhos

Com base no histórico levantado, e nas simulações efetuadas, sugerimos alguns estudos que podem ser efetuados, e que não foram explorados nesta pesquisa:

- ◆ Estudo sobre o comportamento das várias metodologias e arquiteturas na aplicação a uma determinada série temporal, pertencente a uma família, com parâmetros similares, visando estabelecer regras de utilização em conformidade com a série temporal em estudo, podendo

estabelecer uma correlação entre arquiteturas de redes e séries temporais pertencentes a uma mesma classe.

- ◆ A combinação de lógica Fuzzy com redes neurais pode prover um sistema com características de resposta que integram as vantagens de ambos os algoritmos. A rede neural, que é uma caixa preta que necessita ser definida a partir de um processo de aprendizagem com os dados medidos no ambiente, e o sistema Fuzzy que exige o conhecimento do comportamento prévio das variáveis fuzzy e das funções dos membros, para definir as regras fuzzy que mais contribuem para a solução da aplicação. São os sistemas *fuzzy artificial neural network*.
- ◆ A utilização de redes neurais para prognóstico de valores futuros da curva de carga de sistemas elétricos, associando as variáveis que interferem na determinação do problema, pode definir a operação ótima do sistema elétrico, as necessidades de troca de energia com outras concessionárias e consequentemente conferindo segurança para o sistema.
- ◆ A necessidade de processar redes com uma grande quantidade de dados, ou mesmo para aquelas redes em que é requerido um elevado grau de precisão na saída, o procedimento de aprendizagem para séries temporais caóticas pode tornar-se extremamente dispendioso. A redução do tempo de treinamento e principalmente do erro no processo de aprendizagem podem ser dramaticamente reduzidos, efetuando sucessivos treinamentos em uma arquitetura hierarquizada.
- ◆ A comparação de métodos que utilizam redes neurais com aqueles que usam metodologias convencionais, a exemplo da metodologia Box-Jenkins, continua sendo investigada, considerando que ambos podem apresentar resultados comparáveis, principalmente para séries com base histórica elevada.

REFERÊNCIAS BIBLIOGRÁFICAS

- [AHM e RAH 94] -Ahmad U., Rahman S.,-Refinement of cost estimated with artificial neural nets Congress on Computing in Civil Engineering 1994
- [ARM 92] -Armstrong J.S.,- Error measures for generalizing about forecasting methods: Empirical comparisons – International Journal of Forecasting – 1992
- [ASA e TAK 94] - Asakawa K., Takagi H. – Neural networks in Japan – Communcations of the ACM - 1994
- [BAL 99] -Ballou R. H – Business Logistics Management – Prentice Hall Fourth Edition.- 1999
- [BED e BAI 86] -Bedworth D.D., Bailey J.E., - Integrated Production Control Systems – John Wiley and Sons – 1986
- [BOW e COM 87] -Bowerman B.L., O’Connell R.T., – Times Series Forecasting - Boston: Duxbury Press 1987
- [BRA et al. 94] -Brass J., Gerrard A M., Peel D.,-Estimating vessel costs via neural networks International Cost Engineering Congress 1994
- [BOX e JEN 70] - Box G. E. e Jenkins G. M. – Time-series analysis: Forecasting and control. - San Francisco: Holden-Day 1970
- [BRO 63] -Brown R.G., - Smoothing, Forecasting and Prediction of Discrete Time Series - Prentice-Hall - 1963
- [CHA et al. 92] -Chakraborty K, Mehrotra K, Mohan C K, Ranka S,-Forecasting the Behavior of Multivariate Time Series Using Neural Network -IEEE -1992
- [CHA et al. 71] -John C. Chambers,J.C., Satinder K. Mulick S.K., Smith D.D., – How to Choose the Right Forecasting Technique - Harvard Business Review 1971
- [CHA e TON 86] -Chan K S, Tong H.,-On estimating threshold in autoregressive models Journal of Time Series Analysis. 1986
- [CHE et al. 92] - Chen S., Mulgrew B., McLaughlin S. – Adaptive bayesian feedback equalizer based on a radial function network – IEEE 1992
- [CON 88] -Connor D., - Data transformation explains the basics of Neural Network – EDN –1988
- [COU et al. 97] -Courtney H., Kirkland J., Viguerie P.,-Strategy Under Uncertainty – Harvard Business Review – 1997

- [COV 65] - Cover T.M. – Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition – IEEE - 1965
- [CYB 89] -Cybenko G.-Approximtion by Superpositions of a Sigmoidal Function Mathematical Control, Signals and Systems -1989.
- [CUR e PEE 98] -Curry B., Peel M.J., -Neural networks and business forecasting: An application to cross-sectional audit fee data – International Journal of Commerce and Management – 1998
- [DAL 87] -Dalrymple D. J., - Sales Forecasting Practices – Internationa Journal of Forecasting – 1987
- [DAV e DRU 96] -Davis H. W., Drumm W. H., - Logistics Costs and Customer Service Levels Council of Logistics Management – 1996
- [DAY 90] -Dayhoff J., - Neural Network Architectures – An Introduction – Van Nostrand Reinhold Book - 1990
- [DEP et al. 91] -Deppisch J, Bauer H U, Geisel T.,- Hierarchical training of neural networks and prediction of chaotic time series -Physics Letters –1991
- [DIE e NAS 90] -Diebold F.X., Nason J.A, -Nonparametric exchange rate prediction? – Journal of International Economica – 1990
- [DON 93] -Donaldson R G, Kamstra M, Kim H Y –Evaluating Alternative Models for Conditional Stock Volatility: Evidence from International Data – University of British Columbia –1993
- [DUT e SBE 88] -Dutta S., Sbekkar S.,-Bond rating: a non-conservative application International Joint Conference on Neural Networks 1988
- [ENG 82] -Engle R F –Autoregressive conditional heteroskedasticity whit estimates of the variance of UK inflation Econometrica 1982.
- [FAU 94] - Fausset L. – Fundamentals of neural networks architectures, algorithms and applications – Prentice Hall - 1994
- [FIS et al 94] -Fisher M. L., Hammond J.H.,Obermeyer W.R., Rama A. – Making Supply Meet Demand in na Uncertain World – Harvard Business Review – 1994
- [FLO 80] -Flowers A D - A Simulation Study of Smoothing Constant Limits for an Adaptive Forecasting System -Journal of Operations Management 1980
- [FOS et al. 92] -Foster B., Collopy F.,Ungar L. – Neural Network Forecasting of Short Noisy Time Series – Computers and Chemical Engineering – 1992
- [FUN 89] -Funahashi K., -On the approximate relization of continuous mappings by Neural Networks- Neural Networks 1989.

- [GAR e ROU 95] -Garza J. M., Rouhana K. G.,- Neural networks versus parameter-based applications in cost estimating – Cost Engineering - 1995
- [GEO e MUR 86] -Georgoff D.M., Murdick R.G.,- Manager's Guide to Forecasting – Harvard Business Review – 1986
- [GOO e KUM 92] -Gooijer J.G., Kumar K.,-Some recent developments in non-linear time series modelling, testing, and forecasting- International Journal of Forecasting -1992.
- [GER et al. 94] -Gerrard A M., Brass J., Peel D.,-Using neural nets to cost chemical plants – European Symposium on Computer-Aided - 1994
- [GRA e AND 78] -Granger C.W.J. e Andersen A P –An introduction to bilinear time series models – Vandenhoeck e Ruprecht-1978
- [GRA 89] -Granger C.W.J.- Stochastic or deterministic non-linear models? A discussion of the recent literature in economics. SERC 1989
- [GOU e MON 90] -Gourieroux C., Monfort A,-Qualitative threshold ARCH models ENSAE/INSEE 1990
- [GOR 94] -Goor W. –Research Perspective on Neural Networks -International Journal of Forecasting- 1994
- [GRO 73] -Grossberg S., - Contour enhancement, short-term memory, and constancies in reverberating neural networks – Studies Applied Mathematics – 1973.
- [HAG e OZA 80] -Haggan V., Ozaki T,-Amplitude dependent exponential AR model fitting for non-linear random vibrations -Time-Series - 1980
- [HAG e OZA 81] -Haggan V., Ozaki T.,-Modelling non-linear random vibrations using an amplitude-dependent autoregressive time series model -Biometrika -1981.
- [HAY 99] - Haykin S. - Neural networks: a comprehensive foundation – Prentice Hall - 1999
- [HIL et al. 96] -Hill.T, O'Connor M., Remus W.- Neural Network Models for Time Series Forecasts- Management Science - 1996.
- [HOP 93] -Hoptroff A R –The principles and practice of time series forecasting and business modelling using neural networks Neural Computing and Applications 1993
- [HOR et al.89] -Hornik K. M.,Stinchcombe, White H.-Multilayer Feedforward Networks are Universal Approximators - Neural Networks – 1989
- [HU et al. 99] -Hu M. Y., Zhang G., Jiang C. X., Patuwa B.E.,- A cross-validation analysis of neural network out-of-sample performance in exchange rate forecasting – Decision Sciences – 1999

- [JON 78] -Jones D A – Nonlinear Autoregressive Processes- Proceedings of the Royal Society of London - 1978.
- [KAM e TAN 90] -Kamijo K., Tanigawa T.,-Stock price pattern recognition-a recurrent neural network approach International Joint Conference of Neural Network 1990
- [KAN 91] -Kang S. – An Investigation of the Use of Feedforward Neural Networks for Forecasting, Ph.D. Dissertation, Kent State University, Ohio 1991.
- [KAR 96] -Kartalopoulos S.V. – Understanding Neural Networks and Fuzzy Logic – IEEE Press - 1996
- [KIM e ASA] -Kimoto T.,Asakawa K.,-Stock Market prediction system with modular neural networks- International Joint Conference on Neural Networks 1990
- [KUM et al. 94] -Kumar S., Krishna A, Satsangi P.,-Fuzzy systems and neural networks in software engineering Applied Intelligence 1994
- [LAL e ZIN 76] -LaLonde B. J., Zinszer P.H.- Customer Service: Meaning and Measurement National Council of Physical Distribution Management – 1976
- [LAP 99] -Lapide L.,-New developments in business forecasting – The Journal of Business Forecasting Methods and Systems – 1999
- [LIG 92] -Ligth W. – Some aspects of radial basis function approximation - Approximation Theory Spline Functions na Applications - 1992
- [LIN 93] -Lin F.C.,Lin M.,-Neural Networks in the Financial Industry – AI Expert 1993
- [LIP 89] - Lippmann R. P. – Pattern classification using neural networks – IEEE 1989
- [LIU e BRO 88] -Liu J., Brockwell P.J.-On the general bilinear time series model- Journal of Applied Probability 1988.
- [LOW 91] - Lowe D. – What have neural networks to offer statistical pattern processing? – Conference on adaptive signal processing - 1991
- [LUU et al. 88] -Luukkonen R., Saikkonen P., Terasvirta T.,-Testing Linearity Against Smooth transition autoregressive models- Biometrika; Testing linearity in univariate time series model Scandinavian Journal of Statistics, 1988.
- [MAK et al. 82] -Makridakis S., Anderson A, Carbone R., Fildes R., Hibon M., Lewandowski R., Newton J., Parzen E., Winkler R., - The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition – International Journal of Forecasting - 1982
- [MAK e HOG 81] -Makridakis S., Hogarth R.M., - Forecasting and Panning: An Evaluation – Management Science – 1981
- [MAR et al. 92] -Marquez L., Hill T., O'Connor M., Remus W.-Neural Networks Models for

- Forecasting: A Review International Conference on System Sciences 1992
- [MCM 97] -McMenamin J.S., - A primer on neural networks for forecasting – The Journal of Business Forecasting Methods and Systems – 1997
- [MEN e KEN 99] -Mentzer J.T., Kent J.L., - Forecasting demand in the Logabarger Company Marketing Management – 1999
- [MEE e ROS 91] -Meese R. A, Rose A,- An empirical assessment of non-linearities in models of exchange rate determination – Review of Economic Studies – 1991
- [MHA 96] - Mhaskar H. N. – Neural networks for optimal approximation of smooth and analytical functions – Neural computation - 1996
- [MIZ 92] -Mizrach B., - Multivariate nearest-neighbour forecasts of EMS exchange rates -Journal of Applied Econometrics – 1992
- [MOO e DAR 89] - Moody J. e Darken C. J. – Fast learning in networks of locally-tuned processing units – Neural computation - 1989
- [NIC e QUI 82] -Nicholls D F, Quinn B G,-Random Coefficient Autoregressive Models:An Introduction- Springer-Verlag 1982.
- [NIY e GIR 96] - Niyogi P. e Girosi F. – On the relationship between generalization error, hypothesis complexity, and sample complexity for radial basis function – Neural computation 1996
- [OZA e ODA 78] -Ozaki T., Oda H,-Non-linear time series model identification by Akaike's information criterion-Information and Systems - 1978.
- [PAC e DOW 83] -Pack D J,Downing D J.,-Why Didn't Box-Jenkins Win (Again)? – International Symposium of Forecasting 1983
- [PAR 82] -Parker D.,-Learning Logic -Invetion Report-Stanford University – 1982
- [PAR et al. 91] -Park D C, El-Sharkawi M A, MarksII R J, Atlas L E , Damborg M J- Electric Load Forecasting Using na Artificial Neural Network. -IEEE –1991
- [POW 95] - Powell M. J. D. – Radial basis function for multivariate interpolation-A review – Conference on algorithms for the application of function and data - 1985
- [PRI 80] -Priestley M. B., - State-dependent models : a general approach to non-linear time series analysis – Journal of Time Series Analysis – 1980
- [RAO et al. 93] -Rao G N., Grobler F., Kim S.,-Conceptual cost estimating International Conference on Computing in Civil and Building Engineering 1993
- [REF 93] -Refenes A N.,-Currency exchange rate prediction and neural network design strategies Neural Computing and Applications 1993

- [ROS 58] -Rosenblatt F.,-The Perceptron: A probabilistic model for information storage and organization in the brain – Psychological Review 1958
- [RUM et al. 86] -Rumelhart D., Hinton G., Williams R., -Learning Representations by Backpropagation Errors – Nature 1986
- [RUM e MCC 86] -Rumelhart D., McClelland J.,-Parallel Distributed Processing -MIT -1986
- [SCH 90] -Schoenenburg E., -Stock price prediction using neural network strategies – Neural Computing and Applications – 1990
- [SHA e PAT 90] -Sharda R., Patil R. – Neural Networks as Forecasting Experts:An Empirical Test- Proc. IJCNN Meeting - 1990.
- [SHA e SIM 99] -Shaw I.S., Simões M.G.,- Controle e modelagem Fuzzy – Editora Edgard Blucher 1999
- [SJO 82] -Sjoberg L., - Aided and Unaided Decision Making: Improved Intuitive Judgment – Journal of Forecasting – 1982
- [SMI e MAS 97] -Smith A, Mason A K,- Cost estimation predictive modeling: Regression versus neural network The Engineering Economist 1997
- [SUB e GAB 84] -Subba Rao T., Gabr M. M.-An introduction to bispectral analysis and bilinear time series models , Springer-Verlag, 1984
- [SUR e SIN 90] -Surkan A, Singleton J.,-Neural Networks for bond rating improved by multiple hidden layers-International Joint Conference on Neural Networks 1990
- [TAN et al. 90] -Tang Z., Almeida C., Fishwick P.,-Time Series Forecasting Using Neural Networks vs. Box-Jenkins Methodology - Simulation -1990
- [TAN e BET 99] -Tanwari A U., Betts J – Impact of Forecasting on Demand Planning – Production and Inventory Journal – 1999
- [TAN e FIS 93] -Tang Z., Fishwick P A,-Feedforward neural nets as models for time series forecasting Journal on Computing 1993
- [TJφ 86] -Tjøstheim D., - Estimation in non-linear time series model – Stochastic Processes and their Applications – 1986.
- [TON 80] -Tong H.,- A view on non-linear time series building Time Series 1980
- [TON 83] -Tong H.,-Threshold Models in Non-Linear Time Series Analysis Springer-Verlag 1983
- [TON e LIM 80] -Tong H., Lim K S,-Threshold autoregression, limit cycles and cyclical data Journal of the Royal Statistical Society 1980.

- [TYS e TJφ 88] -Tyssedal J.S. , Tjøstheim D J,-An autoregressive model with suddenly changing parameters and na application to stock market prices Applied Statistics 1988.
- [VEM 93] -Venkatachalam A R., -Software coxt estimation using artificial neural networks – International Joint conference on Neural Networks – 1993
- [VEM e ROG 93]-Vemuri V R., Rogers R D.,- Artificial Neural Networks- Forecsting Time Series IEEE 1993
- [WAS 89] -Wasserman P. D. -Neural Computing: Theory and Practice Van Nostrand-Reinhold 1989
- [WEI 86] -Weiss A A,-ARCH and bilinear time series model:Comparison and combination Journal of Business and Economic Statistics 1986
- [WER 74] -Werbos P., -Beyond Regression : New Tools for Prediction and Analysis in the Behavioral Sciences – Ph.D dissertation, Harvard University - 1974
- [WID e STE 85] -Widrow B., Sterns S.D,-Adaptive Signal Processing - Prentice Hall - 1985
- [WHI 92] -White H.-Connectionist Nonparametric Regression: Multilayer Feedforward Networks can Learn Arbitrary Mappings; Consequences and Detection of Nonlinear Regression Models; Learning and Statistics- Artificial Neural Networks. 1992.
- [WHY 72] -Whybark D.C., -A Comparison of Adaptive Forecasting Techniques - Logistics and Transportation Review 8 1972
- [WHI 88] -White H.,-Economic prediction using neural network: the case of IBM daily stock returns-International Joint Conference on Neural Networks 1988
- [WHI 94] -White H. -Can Neural Networks Forecast in the Big Leagues? Comparing Network Forecasts to the Pros? – International Symposium of Forecasting Proceedings -1994.
- [WHI et al. 92] -White H., Hornik K., Stinchcombe M., - Universal Approximation of an Unknown Mapping and its Derivatives – Artificial Neural Network: Approximations and Learning Theory - 1992
- [WHI e GAL 92] -White H., Gallant A R, -There Exists a Neural Network That Does Not Make Avoidable Mistakes -Artificial Neural Network -1992.
- [WHI e STI 92] -White H., Stinchcombe M.-Approximating and Learning Unknown Mappings Using Multilayer Feedforward Networks whit Bounded Weights Artificial Neural Networks 1992

- [YAM e KUR 91]-Yamaba S., Kurashima H.,-Decision Support System for position optimization on currency option dealing International Conference on Artificial Intelligence Applications 1991
- [ZAK 90] -Zakoian J.M.-Modèle autoregressif à un seuil ENSAE/INSEE; Modèles hétéroscédastiques à seuil ENSAE/INSEE Paris 1990
- [ZHA et al. 98] -Zhang G., Patuwo B.E., Hu M.Y.,-Forecasting with artificial neural networks: The state of the art – International Journal of Forecasting 1998
- [ZIM 94] -Zimmerman H G.,-Neural Networks-The Future of Forecasting in Finance? International Symposium of Forecasting -1994
- [] - (.....) – Engineering Statistics Handbook – Chapter 6.Process or Product Monitoring and Control – 6.4 Introduction to Time Series Analysis <http://www.itl.nist.gov/div898/handbook/pmc/section4/....htm>