

HEITOR ANDRÉ KIRSTEN

**COMPARAÇÃO ENTRE OS MODELOS HOLT-
WINTERS E REDES NEURAIIS PARA PREVISÃO
DE SÉRIES TEMPORAIS FINANCEIRAS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Engenharia de Produção e Sistemas.

CURITIBA

2009

HEITOR ANDRÉ KIRSTEN

**COMPARAÇÃO ENTRE OS MODELOS HOLT-
WINTERS E REDES NEURAIS PARA PREVISÃO
DE SÉRIES TEMPORAIS FINANCEIRAS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Engenharia de Produção e Sistemas.

Área de Concentração: *Automação e Controle de Processos.*

Orientador: Prof. Dr. Leandro dos Santos Coelho

CURITIBA

2009

Kirsten, Heitor André

Comparação entre os modelos Holt-Winters e redes neurais para previsão de séries temporais financeiras. Curitiba, 2009. 87p.

Dissertação – Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Engenharia de Produção e Sistemas.

1. Previsão 2. Séries temporais financeiras 3. Holt-Winters 4. Redes neurais. I. Pontifícia Universidade Católica do Paraná. Centro de Ciências Exatas e de Tecnologia. Programa de Pós-Graduação em Engenharia de Produção e Sistemas

Agradecimentos

À Deus, por me dar o talento necessário para todos os meus objetivos.

Aos meus pais, Ivo Kirsten e Liseta Teresinha Holdefer, que de alguma forma ou de outra me proporcionaram os meios necessários para chegar até aqui.

À minha esposa, Fabiane Genovai Kirsten, por ter tido uma enorme paciência e carinho comigo em todos os momentos de privação devido este longo trabalho.

Ao meu orientador, Leandro dos Santos Coelho, que me deu a chance de fazer este trabalho e que me ajudou de forma incalculável a finalizá-lo.

Aos demais parentes, professores e amigos que, de alguma forma, ajudaram na realização deste.

Sumário

Agradecimentos	iv
Lista de Figuras	vii
Lista de Tabelas	ix
Lista de Abreviaturas	x
Resumo	xi
Abstract.....	xii
Capítulo 1 - Introdução.....	13
1.1. Justificativa da pesquisa	15
1.2. Definição do problema de pesquisa.....	18
1.3. Objetivo	18
1.4. Descrição da proposta.....	18
1.5. Contribuição da pesquisa.....	19
1.6. Organização da dissertação	19
Capítulo 2 - Modelos de previsão.....	20
2.1. Modelo Holt-Winters.....	22
2.2. Média móvel centrada com sazonalidade.....	25
2.3. Redes neurais artificiais.....	27
2.3.1 <i>Perceptron</i> de múltiplas camadas.....	33
2.3.2. Função de base radial	36
2.4. Índices de desempenho	40
2.4.1. Medidas estatísticas padrões (ME, MAE, MSE e MAPE)	41
Capítulo 3 - Metodologia.....	43
3.1. Caracterização da pesquisa	43
3.2. Séries Temporais	44
3.2.1. Mackey-Glass	44
3.2.2. Índice Bovespa	46
3.2.3. Índice Dow Jones	47
3.2.4. Índice Nasdaq	49
3.3. Modelos de previsão.....	50

3.4. Índices de desempenho	50
3.5. Aplicação da previsão.....	50
Capítulo 4 - Apresentação e Análise dos Resultados	52
4.1. Simulações para a saída do sistema Mackey-Glass.....	52
4.1.1. Previsão com MMC com índice de sazonalidade.....	52
4.1.2. Previsão usando o método Holt-Winters	53
4.1.3. Previsão usando RNs-MLP	54
4.1.4. Previsão usando RNs-RBF	56
4.1.5. Comentários sobre os resultados da saída do sistema Mackey-Glass	57
4.2. Simulações para a série Ibovespa	58
4.2.1. Previsão com média móvel centrada com índice de sazonalidade	58
4.2.2. Previsão usando o método Holt-Winters	59
4.2.3. Previsão usando RNs-MLP	61
4.2.4. Previsão usando RNs-RBF	62
4.2.5. Comentários sobre os resultados da série Ibovespa	63
4.3. Simulações para a série Dow Jones	64
4.3.1. Previsão com média móvel centrada com índice de sazonalidade	64
4.3.2. Previsão usando o método Holt-Winters	65
4.3.3. Previsão usando RNs-MLP	66
4.3.4. Previsão usando RNs-RBF	67
4.3.5. Comentários sobre os resultados da série Dow Jones	69
4.4. Simulações para a série Nasdaq.....	70
4.4.1. Previsão com média móvel centrada com índice de sazonalidade	70
4.4.2. Previsão usando o método Holt-Winters	71
4.4.3. Previsão usando RNs-MLP	72
4.4.4. Previsão usando RNs-RBF	73
4.4.5. Comentários sobre os resultados da série Nasdaq.....	74
4.5. Síntese conclusiva	75
Capítulo 5 - Conclusão	77
Referências Bibliográficas.....	80

Lista de Figuras

Figura 2.1 – Estrutura básica de um neurônio artificial. Adaptado de Santos (2005).....	29
Figura 2.2 – Padrões linearmente separáveis	33
Figura 2.3 – Exemplo de estrutura de uma rede neural MLP.....	34
Figura 2.4 – Estrutura básica de uma rede neural RBF. Adaptado de Santos (2005).	38
Figura 3.1 – Saída do sistema Mackey-Glass.....	45
Figura 3.2 – Série temporal do índice Bovespa utilizada neste trabalho.....	47
Figura 3.3 – Série temporal do índice Dow Jones.....	48
Figura 3.4 – Série temporal do índice Nasdaq utilizada neste trabalho	49
Figura 4.1 – Previsão da saída do sistema Mackey-Glass usando o modelo MMC com índice de sazonalidade e o erro quadrático dos segmentos de validação e teste	53
Figura 4.2 – Previsão da saída do sistema Mackey-Glass usando o modelo Holt-Winters e o erro quadrático dos segmentos de validação e teste	54
Figura 4.3 – Previsão de saída do sistema Mackey-Glass usando RN-MLP e o erro quadrático dos segmentos de validação e teste.....	55
Figura 4.4 – Previsão da saída do sistema Mackey-Glass usando RN-RBF e o erro quadrático dos segmentos de validação e teste.....	56
Figura 4.5 – Previsão da série Ibovespa usando o modelo MMC com índice de sazonalidade e o erro quadrático dos segmentos de validação e teste.....	59
Figura 4.6 – Previsão da série Ibovespa usando o modelo Holt-Winters e o erro quadrático dos segmentos de validação e teste.....	60
Figura 4.7 – Previsão da série Ibovespa usando RN-MLP e o erro quadrático dos segmentos de validação e teste	61
Figura 4.8 – Previsão da série Ibovespa usando RN-RBF e o erro quadrático dos segmentos de validação e teste	62
Figura 4.9 – Previsão da série Dow Jones usando o modelo MMC com índice de sazonalidade e o erro quadrático dos segmentos de validação e teste.....	64

Figura 4.10 – Previsão da série Dow Jones usando o modelo Holt-Winters e o erro quadrático dos segmentos de validação e teste.....	66
Figura 4.11 – Previsão da série Dow Jones usando RN-MLP e o erro quadrático dos segmentos de validação e teste	67
Figura 4.12 – Previsão da série Dow Jones usando RN-RBF e o erro quadrático dos segmentos de validação e teste	68
Figura 4.13 – Previsão da série Nasdaq usando o modelo MMC com índice de sazonalidade e o erro quadrático dos segmentos de validação e teste.....	70
Figura 4.14 – Previsão da série Nasdaq usando o modelo Holt-Winters e o erro quadrático dos segmentos de validação e teste.....	71
Figura 4.15 – Previsão da série Nasdaq usando RN-MLP e o erro quadrático dos segmentos de validação e teste	72
Figura 4.16 – Previsão da série Nasdaq usando RN-RBF e o erro quadrático dos segmentos de validação e teste	73

Lista de Tabelas

Tabela 2.1 – Dados para exemplo do método Holt-Winters. Adaptado de Makridakis <i>et al.</i> (1998).	24
Tabela 2.2 – Dados para exemplo da média móvel centrada com índice de sazonalidade. Fonte: Tubino (2000).....	26
Tabela 2.3 – Funções de ativação. Adaptado de Cortez (1997) e Santos (2005).	30
Tabela 2.4 – Principais funções de base radial. Fonte: Santos (2005).	37
Tabela 2.5 – Algoritmos de aprendizagem para ajuste de parâmetros livres. Adaptado de Castro (2001).	39
Tabela 3.1. – Distribuição das amostras para cada segmento	51
Tabela 4.1 – Índices de desempenho para a previsão da saída do sistema Mackey-Glass usando RN-MLP.....	55
Tabela 4.2 – Índices de desempenho para a saída do sistema da saída do sistema Mackey-Glass usando RN-RBF	57
Tabela 4.3 – Comparativo dos resultados para a saída do sistema Mackey-Glass.....	57
Tabela 4.4 – Índices de desempenho para a previsão da série Ibovespa usando RN-MLP.....	61
Tabela 4.5 – Índices de desempenho para a previsão da série Ibovespa usando RN-RBF	63
Tabela 4.6 – Comparativo dos resultados para a série Ibovespa	63
Tabela 4.7 – Índices de desempenho para a previsão da série Dow Jones usando RN-MLP ..	67
Tabela 4.8 – Índices de desempenho para a previsão da série Dow Jones usando RN-RBF ...	68
Tabela 4.9 – Comparativo dos resultados para a série Dow Jones.....	69
Tabela 4.10 – Índices de desempenho para a previsão da série Nasdaq usando RN-MLP.....	73
Tabela 4.11 – Índices de desempenho para a previsão da série Nasdaq usando RN-RBF.....	74
Tabela 4.12 – Comparativo dos resultados para a série Nasdaq	74

Lista de Abreviaturas

ARIMA	<i>AutoRegressive Integrated Moving Average</i>
ARMA	<i>AutoRegressive Moving Average</i>
DAX	<i>Deutscher Aktien Index</i>
GARCH	<i>Generalized AutoRegressive Conditional Heteroskedasticity</i>
MAE	<i>Mean Absolute Error</i>
MAPE	<i>Mean Absolute Percentage Error</i>
ME	<i>Mean Error</i>
MLP	<i>MultiLayer Perceptron</i>
MMC	<i>Média Móvel Centrada</i>
MPE	<i>Mean Percentage Error</i>
MSE	<i>Mean Squared Error</i>
PE	<i>Percentage Error</i>
RBF	<i>Radial Basis Function</i>
RN	<i>Rede Neural</i>
SARIMA	<i>Seasonal AutoRegressive Integrated Moving Average</i>

Resumo

A previsão de séries temporais é um problema que tem recebido especial atenção dos pesquisadores nos últimos anos. Prever o futuro, e em especial o comportamento de séries temporais, é fundamental em análises e apoio à tomada de decisões, e continua sendo um desafio para a estatística e a computação. O Holt-Winters é um modelo de previsão exponencial, conhecido por lidar com elementos de tendência e sazonalidade de uma série temporal. As redes neurais têm se mostrado grandes ferramentas na aplicação de previsões de séries temporais, bem como sua capacidade de paralelismo e não-linearidade chamam a atenção para sua utilização. O *perceptron* é a forma mais simples de uma rede neural, porém, com capacidade apenas para trabalhar com tarefas lineares. A rede neural *MultiLayer Perceptron* surge para tratar os problemas não lineares. As redes neurais de funções de base radial são de uma classe de redes neurais com uma camada oculta, onde cada neurônio é ativado por uma função de base radial, de onde é derivado seu nome. Os neurônios na camada oculta fornecem um conjunto de funções não-lineares, de base radial, que constituem uma base arbitrária para os padrões de entrada. Neste contexto este trabalho se propõe a realizar um comparativo de alguns modelos lineares e não-lineares para a previsão um passo a frente de três séries temporais distintas, tentando encontrar qual modelo se adapta melhor a qual série temporal. Os modelos comparados são o de média móvel centrada (MMC) com índice de sazonalidade, Holt-Winters, rede neural *perceptron* de múltiplas camadas e rede neural função de base radial. Quanto às séries temporais, o objetivo é comparar séries diferentes, porém dentro da área de finanças: índices Bovespa, Dow Jones e Nasdaq, além da saída do sistema Mackey-Glass, para ser utilizada como *benchmark*.

Palavras-Chave: Previsão, séries temporais, Holt-Winters, redes neurais, mercado financeiro.

Abstract

The time series forecasting is a problem that has received special attention from researchers in recent years. Predict the future, and in particular the behavior of time series, is essential to analyze and support decision making, and remains a challenge for statistics and computing. The Holt-Winters model is a kind of exponential estimator, known for dealing with time series elements of trend and seasonality. The neural networks have been major tools in the application of time series forecasts, and its capacity for parallelism and nonlinearity draw attention to their use. The perceptron is the simplest form of a neural network, however, it is able to work only with linear tasks. The MultiLayer Perceptron neural network is to treat the nonlinear problems. The neural network of radial basis functions are of a class of neural networks with one hidden layer, where each neuron is activated by a function of radial basis from which it derived its name. The neurons in the hidden layer provide a set of nonlinear functions of radial base, which constitute a basis for arbitrary patterns of entry. In this context this study is to present a comparison of some linear and nonlinear forecasting models for one step ahead of three different time series, trying to find which model fits best to which time series. The models are compared to the centered moving average with index of seasonality, Holt-Winters, the multi-layered perceptron neural network and radial basis function neural network. As for time series, the goal is to compare different series, but within the finance area: Bovespa, Dow Jones and Nasdaq index. Besides that, the output chaotic system Mackey-Glass is used as a benchmark.

Keywords: Forecasting, time series, Holt-Winters, neural networks, financial market.

Capítulo 1

Introdução

Uma série temporal é uma coleção de observações feitas seqüencialmente ao longo do tempo. Uma característica importante deste tipo de dado é que as observações vizinhas são dependentes umas das outras. Enquanto para modelos de regressão a ordem não importa, para as séries temporais essa ordem é crucial. Ehlers (2007) menciona algumas áreas do conhecimento onde surgem séries temporais, tais como economia (preços diários de ações, taxa mensal de desemprego, produção industrial), medicina (eletrocardiograma, eletroencefalograma), epidemiologia (número mensal de novos casos de meningite), meteorologia (precipitação pluviométrica, temperatura diária, velocidade do vento), entre outros.

Outras características inerentes das séries temporais são que a ordem temporal das observações deve ser levada em consideração, pelas observações serem correlacionadas torna-se difícil a análise e técnicas mais específicas são necessárias, tendências e sazonalidade podem se tornar difíceis de estimar ou remover, a seleção de modelos pode ser complicada e torna-se mais difícil de lidar com *outliers* e observações discrepantes, dada sua natureza seqüencial.

Freqüentemente existe uma diferença de tempo entre um evento ou necessidade iminente e a ocorrência desse evento ou necessidade. Esta diferença de tempo é a principal razão para prever e planejar. A previsão é uma importante ajuda para um planejamento eficiente e eficaz (MAKRIDAKIS *et al.*, 1998).

A previsão de séries temporais é um problema que tem recebido especial atenção dos pesquisadores nos últimos anos. Prever o futuro, e em especial o comportamento de séries temporais, é fundamental em análises e tomada de decisões, e continua sendo um desafio para a estatística e a computação (ABELÉM, 1994).

Esta idéia que a previsão serve como meio de determinar base para o apoio à tomada de decisão já vinha de Morettin e Toloí (1987), que qualificavam as previsões como sendo simples e intuitivas ou complexas e quantitativas. O primeiro grupo necessita de pouca ou nenhuma análise, enquanto o segundo grupo já necessita de uma quantidade considerável de análise dos dados.

As previsões também podem ser divididas em dois enfoques, segundo Morettin e Toloí (1987), onde o primeiro é baseado em alguma teoria para construir o modelo, enquanto no segundo o modelo é obtido diretamente dos dados disponíveis, sendo este segundo enfoque chamado de estatístico ou de séries temporais. É este segundo enfoque o utilizado neste trabalho.

Estes dois enfoques são conhecidos como previsões qualitativas e quantitativas, respectivamente.

As previsões quantitativas são possíveis quando, segundo Makridakis, *et al.* (1998), a informação sobre o passado está disponível, essa informação pode ser quantificada de forma numérica e pode-se considerar que alguns padrões que aconteceram no passado irão se repetir no futuro.

Quando se trata de previsão qualitativa, os dados do passado não são tratados da mesma forma como na previsão quantitativa. As entradas dependem do método específico selecionado e são em sua maioria produto de julgamento e conhecimento acumulado. Esse tipo de previsão requer freqüentemente entrada de pessoas especialmente treinadas.

Outra preocupação na previsão de séries temporais é determinar quão bem um modelo se adaptou à série histórica. Neste contexto, a acuracidade é tida como um dos principais critérios para a seleção de um modelo de previsão. A acuracidade se refere à quão bem um modelo de previsão é capaz de reproduzir os dados de uma série temporal já conhecidos. A acuracidade pode ser determinada através de várias medidas, tais como *Mean Error* (ME, Erro Médio), *Mean Absolute Error* (MAE, Erro Médio Absoluto), *Mean Squared Error* (MSE, Erro Quadrático Médio), *Mean Absolute Percentage Error* (MAPE, Erro Percentual Médio Absoluto) e R^2 (MAKRIDAKIS *et al.*, 1998).

1.1. Justificativa da pesquisa

O fato de que a previsão de séries temporais é um problema que tem recebido especial atenção, dado que é fundamental na análise e tomada de decisões, como já constatado por Abelém (1994) e Morettin e Toloí (1987), chamou a atenção por suas dificuldades e complexidades.

A verdade é que vários trabalhos já estudaram a comparação entre os modelos de previsão para uma série temporal, como Souza (2005a), Falco (2005) e Lima (2004); ou várias séries temporais, como Teixeira (2003), Cortez (1997) e Cortez (2002). No entanto, a vasta quantidade de possibilidades para fazer a previsão de uma série temporal e as várias conclusões encontradas pelos autores faz com que se justifique a pesquisa comparativa entre alguns dos modelos mais populares encontrados, como o Holt-Winters e as redes neurais. Uma parte dos trabalhos encontrados que pesquisaram esses dois modelos para a previsão de séries temporais são relacionados a seguir.

Ehlers (2007), em seu trabalho, fez uma análise detalhada sobre as séries temporais.

A previsão de séries temporais, em todas as suas etapas, teve uma revisão de literatura realizada por Gooijer e Hyndman (2006).

Sobre o modelo Holt-Winters (WINTERS, 1960), os autores a seguir fizeram os estudos que se seguem.

Lopes (2002) aplicou diversos métodos exponenciais em comparação com métodos de médias móveis para previsão de estoque de peças para automóveis e concluiu que os métodos exponenciais têm vantagem sobre métodos mais simples, como o de médias móveis.

Barbiero (2003) aplicou os modelos SARIMA (*seasonal autoregressive integrated moving average*), Regressão com erros ARMA (*autoregressive moving average*) e Holt-Winters em dados sobre a demanda de diversos produtos da empresa Correios. Encontrou que os modelos SARIMA e regressão com erros ARMA eram os que mais se aproximavam da realidade.

Barbosa (2005) estudou a aplicação de médias móveis e métodos exponenciais quando aplicados em dados da demanda do álcool. Encontrou que as médias móveis geravam um menor erro quadrático do que os métodos exponenciais (incluindo Holt-Winters) e que, usando o MSE como indicador de desempenho, a previsão usando média móvel apresentou o melhor resultado.

Souza (2005a) encontrou que a previsão da demanda por energia elétrica no estado de Santa Catarina através de redes neurais pode ser melhorada, ou até mesmo substituída, por um modelo usando Holt-Winters e regressão dinâmica, através da média ponderada das previsões encontradas. Souza *et al.* (2004) também concluiu apropriado o uso do modelo Holt-Winters e regressão dinâmica para a previsão da demanda de energia elétrica consumida por indústrias no estado de Santa Catarina.

Souza (2005b) comparou a previsão de preços das *commodities* cana-de-açúcar, açúcar e álcool, quando utilizados modelos exponenciais, médias móveis, Holt-Winters e outros. Descobriu que o método Holt-Winters era o que melhor se adaptava para aquelas séries em específico, e que a previsão gerada era acurada.

Souza (2007) identificou que a previsão de curto prazo da demanda de energia elétrica de um determinado distribuidor no sudeste do Brasil, quando utilizada uma variação do modelo Holt-Winters, é possível ter um ganho sobre a utilização do modelo Holt-Winters tradicional. A variação Holt-Winters com múltiplos ciclos, proposta por Taylor (2003), pode manipular séries temporais com duas sazonalidades (uma diária e outra semanal, por exemplo).

Miranda (2007) propôs a utilização de uma variação do modelo Holt-Winters para a previsão de vários passos à frente em uma série temporal de alta frequência de energia elétrica, utilizando entradas exógenas de feriados e temperatura. Comparando o modelo proposto com um modelo ingênuo, o modelo proposto apresentou resultado melhor.

Sobre previsão utilizando redes neurais, pode-se destacar as literaturas a seguir.

Zhang *et al.* (1998) fizeram uma revisão do estado da arte das redes neurais aplicadas à previsão.

Zhang (2001) fez uma investigação da capacidade de utilização de redes neurais na previsão de séries temporais lineares, tendo como resultado que as redes neurais se mostraram competentes para a realização desse tipo de previsão.

Zhang e Qi (2005) estudaram a aplicação de redes neurais em séries temporais com sazonalidade e tendência. Fizeram uma comparação entre um modelo Box-Jenkins e as redes neurais, o resultado encontrado foi que as redes neurais não conseguem capturar tendência e sazonalidade de forma efetiva e tratando a sazonalidade e a tendência antes da utilização das redes neurais mostrou-se ser uma forma efetiva de resolver o problema.

Braga (2006) verificou a correlação de outros índices com o Índice Bovespa e o impacto disso na previsão do preço de fechamento utilizando redes neurais e o modelo GARCH (*Generalized Autoregressive Conditional Heteroskedasticity*). Concluiu que os resultados utilizando redes neurais são motivadores e indicam a possibilidade de se utilizar a solução. A previsão gerada pelo modelo GARCH serviu como entrada para a previsão da rede neural, porém não houve ganho de performance com isso.

Freitas (2007) analisou os modelos ARMA-GARCH e redes neurais, quando usados para fazer previsão das séries temporais de taxa de cambio brasileira e da umidade em microclima. Encontrou que os modelos de redes neurais *radial basis function* (RBF, função de base radial) eram satisfatórios para ambas as séries temporais, já o modelo ARMA-GARCH se adaptou melhor a série de taxa de câmbio e não obteve previsão satisfatória para a série de microclima.

Zou *et al.* (2007) comparou a performance do modelo ARIMA (*autoregressive integrated moving average*), redes neurais e uma combinação de modelos lineares para a previsão de preços do alimento em grão no mercado chinês, encontrando como o melhor modelo as redes neurais.

Santos *et al.* (2007) analisaram a diferença entre os modelos de redes neurais *MultiLayer Perceptron* (MLP, *perceptron* multicamadas), redes neurais RBF, sistemas nebulosos, ARMA e ARMA-GARCH no que tange a acuracidade da previsão para uma série temporal de taxa de cambio real/dólar. A conclusão foi que os modelos não lineares (redes neurais e sistemas nebulosos) conseguiram um melhor desempenho do que os modelos lineares (ARMA e ARMA-GARCH).

Selim (2009) comparou a regressão *hedonic* com redes neurais para a previsão dos preços de casas na Turquia. Dada a característica não-linear da série o autor observou que o modelo de redes neurais conseguiu um desempenho melhor do que a regressão *hedonic*.

Amjady e Farshid (2008) propuseram um modelo híbrido de redes neurais e algoritmos evolucionários para a previsão de carga de energia elétrica para o médio prazo. O modelo foi testado com dados de energia elétrica do Irã e dados de teste da *European Network on Intelligent Technologies* obtendo resultados satisfatórios.

Hamzaçebi (2008) propôs uma estrutura de redes neurais para melhorar o desempenho de previsões de séries temporais com sazonalidade. Naquele trabalho comparou a estrutura sugerida com outras estruturas de redes neurais e encontrou que quando a série temporal tem

uma sazonalidade acentuada a estrutura proposta obtém desempenho melhor do que estruturas tradicionais, porém se a sazonalidade é fraca o desempenho é o mesmo das outras estruturas comparadas.

Considerando todos os estudos apresentados, pode-se notar que diferentes resultados foram encontrados e que o tema da previsão de séries temporais, apesar de não ser vago, ainda desperta dúvidas e hipóteses, justificando assim o objetivo deste trabalho.

1.2. Definição do problema de pesquisa

Separando-se os resultados por cada série temporal, qual modelo de previsão, dentre os escolhidos nesta pesquisa, consegue gerar uma previsão um passo a frente melhor que os demais?

1.3. Objetivo

O objetivo geral deste trabalho é a comparação de alguns modelos lineares e não-lineares para a previsão um passo a frente de três séries temporais distintas, tentando encontrar qual modelo se adapta melhor a qual série temporal.

Os modelos comparados são o de média móvel centrada (MMC) com índice de sazonalidade, Holt-Winters, rede neural *perceptron* de múltiplas camadas e rede neural função de base radial.

Quanto às séries temporais, o objetivo é comparar séries diferentes, porém dentro da área de finanças: índices Bovespa, Dow Jones e Nasdaq, além da série temporal Mackey-Glass, para ser utilizada como *benchmark*.

Os objetivos específicos são:

- i) Entender o conceito e implementar cada modelo de previsão comparado;
- ii) Conhecer as características de cada série temporal estudada; e
- iii) Identificar e analisar o melhor modelo de previsão, linear ou não-linear, para cada série temporal.

1.4. Descrição da proposta

Propõe-se para esta dissertação um estudo comparativo entre os modelos de previsão MMC com índice de sazonalidade, Holt-Winters, RN-MLP e RN-RBF; aplicados as séries temporais Mackey-Glass e índices Bovespa, Dow Jones e Nasdaq.

1.5. Contribuição da pesquisa

Este trabalho contribui no aspecto de ajudar a elucidar qual modelo de previsão pode se adequar melhor para as séries temporais financeiras. Acaba contribuindo também para as comparações dos modelos lineares com os modelos não-lineares de previsão. É um diferencial deste trabalho a comparação do modelo Holt-Winters com as RNs-MLP e RNs-RBF para a previsão de séries financeiras.

1.6. Organização da dissertação

A organização do restante deste trabalho deu-se da seguinte forma: no capítulo 2 está apresentada a fundamentação teórica dos modelos de previsão de MMC com índice de sazonalidade, Holt-Winters, RNs-MLP e RNs-RBF, bem como uma breve introdução sobre séries temporais. No capítulo 3 encontra-se a caracterização da pesquisa, as séries temporais escolhidas, os modelos escolhidos, os índices de desempenho escolhidos e também a metodologia aplicada para gerar previsões para as séries temporais usando os modelos escolhidos. No capítulo 4 são apresentados os resultados das aplicações dos modelos de previsões, separados por série temporal, bem como comentários sobre o desempenho de cada método para cada série temporal. Por fim, o capítulo 5 traz as conclusões encontradas pela pesquisa, as limitações do trabalho e as sugestões para trabalhos futuros.

Capítulo 2

Modelos de previsão

Segundo Gaither e Frazier (2002), “uma série temporal é um conjunto de valores observados medidos ao longo de períodos sucessivos de tempo”.

Para Corrar e Theóphilo (2004), “uma série temporal é um conjunto de observações seqüenciais de determinada variável, expressas numericamente, obtidas em períodos regulares de tempo. A disponibilidade dos dados é que condiciona a formatação das séries. O analista deve cuidar para que as observações sejam dispostas em pontos equidistantes no tempo, obtendo, dessa forma, séries temporais discretas”.

Os autores supracitados também explicam que o gráfico de uma série temporal deve ser traçado de maneira que o tempo seja representado no eixo das abscissas (eixo x), por meio de valores discretos seqüenciados. E no eixo das ordenadas (eixo y), dispõem-se os valores da variável de interesse. Mesmo que seja possível representar a escala de tempo em campos alfanuméricos (por exemplo: ano1, bim2) em ferramentas de planilha eletrônica, é necessário representar essa escala de tempo com campos numéricos, cuja distância entre eles seja homogênea.

Corrar e Theóphilo (2004) complementam mencionando que “a principal característica de uma série temporal, que a torna distinta de uma amostra aleatória simples, é a vinculação das observações com o fator tempo”.

Quanto aos componentes de uma série temporal, pode-se elencar:

- i) Tendência: “É, frequentemente, o componente mais importante de uma série temporal. Consiste no movimento de direção geral que cobre um longo período de tempo e reflete as mudanças nos dados ocorridas de modo bastante consistente e gradual. A tendência é observada, por exemplo, nos indicadores

econômicos, preço de ações, etc.” (CORRAR; THEÓPHILO, 2004). Seguindo a mesma linha, Tubino (2000) entende que “tendência consiste num movimento gradual de longo prazo, direcionando os dados”.

- ii) Variações cíclicas: “Consistem em variações que apresentam movimentos regulares em torno da reta ou da curva de tendência. Referem-se às oscilações de longo prazo – períodos de um ano ou mais. Os ciclos podem ser ou não periódicos, isto é, podem ou não seguir padrões análogos, depois de intervalos de tempos iguais.” (CORRAR; THEÓPHILO, 2004). Davis *et al* (2001) complementam: “os fatores cíclicos são mais difíceis de se determinar, já que tanto o período considerado, quanto a causa do ciclo, podem não ser conhecidos. Por exemplo: eleições políticas, guerra, condições econômicas ou pressão psicológica”.
- iii) Variações sazonais: “São variações cíclicas de curto prazo (menos de um ano). Referem-se a movimentos de padrões regulares ou repetidos em torno da linha de tendência, em que os dados de uma série temporal parecem obedecer a períodos como meses, semanas, dias, etc.” (CORRAR, THEÓPHILO, 2004).
- iv) Variações irregulares: “Referem-se aos deslocamentos esporádicos das séries temporais. Decorre, normalmente, de causas naturais ou sociais provocadas por eventos imprevisíveis e não periódicos, chamados de ruídos, como, por exemplo: enchentes, greves, eleições, etc.” (CORRAR, THEÓPHILO, 2004). Segundo Tubino (2000) “as variações irregulares são alterações na demanda passada resultantes de fatores excepcionais, como greves ou catástrofes climáticas, que não podem ser previstos e, por tanto, nem incluídos no modelo”.

Além dos quatro itens propostos por Corrar e Theóphilo, Tubino (2000) coloca mais um componente (que é comum a outros autores): a variação aleatória.

Davis *et al* (2001) ensinam que as variações aleatórias são causadas por eventos casuais. E que quando todos os componentes da série temporal conhecidos são subtraídos da demanda total, “o que sobra é uma parte remanescente inexplicável da demanda. Se não for possível identificar a causa da mesma, esta é presumida como puramente aleatória”.

Na seqüência serão descritos os modelos de previsão Holt-Winters, MMC com índice de sazonalidade, RNs-MLP e RNs-RBF.

2.1. Modelo Holt-Winters

Corrar e Theóphilo (2004) explicam que Holt-Winters “é um método destinado ao tratamento de séries temporais que apresentam tendência e sazonalidade. Operando de forma semelhante ao método de Holt, o método de Winters incorpora parâmetros adicionais para o tratamento da sazonalidade da série”.

Resumindo, o modelo Holt-Winters, segundo Lewis (1997), assume que uma demanda baseada em tendência e sazonalidade é composta dos seguintes elementos:

- i) um elemento de nível, estacionário, que não exhibe tendência ou sazonalidade, mas que pode incluir alguma variação aleatória;
- ii) um elemento de tendência, em adição ao elemento de nível;
- iii) um elemento de sazonalidade em adição ao elemento de tendência e nível.

Makridakis *et al.* (1998) apresenta as seguintes fórmulas do modelo Holt-Winters:

$$\text{Nível: } L_t = \alpha \frac{Y_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + b_{t-1}) \quad (2.1)$$

$$\text{Tendência: } b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1} \quad (2.2)$$

$$\text{Sazonalidade: } S_t = \gamma \frac{Y_t}{L_t} + (1 - \gamma)S_{t-s} \quad (2.3)$$

$$\text{Previsão: } F_{t+m} = (L_t + b_t m)S_{t-s+m}, \quad (2.4)$$

onde

- L_t : elemento nível para o período t ;
- b_t : elemento tendência para o período t ;
- S_t : elemento sazonalidade para o período t ;
- F_{t+m} : valor da previsão para o período m ;
- Y_t : valor observado no período t ;
- s : intervalo de tempo da sazonalidade;
- m : número do período previsto;
- α : constante de ajuste do nível;
- β : constante de ajuste da tendência;
- γ : constante de ajuste da sazonalidade.

Sendo que $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$ e $0 \leq \gamma \leq 1$.

No modelo Holt-Winters existe a necessidade de iniciar os valores de nível, tendência e sazonalidade para aplicar as equações (2.1), (2.2), (2.3) e (2.4).

Sobre a geração dos valores de nível, tendência e sazonalidade, Makridakis *et al.* (1998) menciona que, para determinar os valores da sazonalidade é necessário pelo menos um período sazonal de dados históricos completo. Enquanto isso, os valores iniciais do nível e da tendência são iniciados no período s . O nível pode ser iniciado com a média da primeira época sazonal, conforme a equação apresentada a seguir:

$$L_s = \frac{1}{s}(Y_1 + Y_2 + \dots + Y_s). \quad (2.5)$$

Nota-se que o nível é uma média móvel de ordem s , com o intuito de retirar o efeito da sazonalidade. Para iniciar o elemento da tendência, Makridakis *et al.* (1998) usa duas épocas sazonais, e expõe a seguinte equação:

$$b_s = \frac{1}{s} \left[\frac{Y_{s+1} - Y_1}{s} + \frac{Y_{s+2} - Y_2}{s} + \dots + \frac{Y_{s+s} - Y_s}{s} \right]. \quad (2.6)$$

Cada termo é uma estimativa da tendência sobre uma sazonalidade completa. Para o valor inicial da tendência faz-se uma média de todos os termos.

A seguir, encontra-se o valor da sazonalidade para os primeiros períodos da época sazonal. Esse índice é encontrado dividindo-se os primeiros valores observados pelo valor inicializado do elemento nível, tal que:

$$S_1 = \frac{Y_1}{L_s}, S_2 = \frac{Y_2}{L_s}, \dots, S_s = \frac{Y_s}{L_s}. \quad (2.7)$$

Makridakis *et al.* (1998) ainda coloca que existem vários outros meios de iniciar esses valores, como é o caso de Corrar e Theóphilo (2004) que explicam que primeiro deve-se iniciar os índices sazonais da primeira época sazonal tal que:

$$S_t = \frac{Y_t}{\frac{1}{s}(Y_1 + Y_2 + \dots + Y_s)}. \quad (2.8)$$

Depois, para iniciar o primeiro elemento do nível deve-se dividir o valor observado no período s pelo índice sazonal do mesmo período s . A tendência é iniciada com valor zero.

Martins e Laugeni (2005) sugerem, diferentemente dos outros autores citados acima, iniciar o valor do nível igual ao valor observado do mesmo período, a tendência iniciada com valor zero e o índice da sazonalidade com valor unitário.

A forma de iniciar os valores de nível, tendência e sazonalidade, proposta por Makridakis *et al.* (1998), foi a utilizada neste exemplo e também neste trabalho.

Após iniciar os valores de nível, tendência e sazonalidade, a ordem a ser seguida para encontrar os valores dos elementos para cada um dos períodos é a seguinte:

- i) calcular o elemento nível;
- ii) calcular o valor da tendência; e
- iii) calcular o índice da sazonalidade.

Quando todos os períodos estiverem com os valores preenchidos usa-se a equação (2.4) para determinar os valores projetados.

O exemplo da aplicação do método Holt-Winters pode ser observado na tabela 2.1.

Tabela 2.1 – Dados para exemplo do método Holt-Winters. Adaptado de Makridakis *et al.* (1998).

Período t	Valor observado Y_t	Nível L_t	Tendência b_t	Sazonalidade S_t	Valor previsto F_t	Período previsto m
1	362			0,953		
2	385			1,013		
3	432			1,137		
4	341	380,00	9,75	0,897		
5	382	398,99	10,26	0,953	371,29	
6	409	404,68	10,01	1,013	414,64	
7	498	433,90	11,07	1,137	471,43	
8	387	433,70	10,45	0,897	399,30	
9	473	487,20	12,83	0,953	423,11	
10	513	505,21	13,11	1,013	506,60	
11	582	513,08	12,82	1,137	589,26	
12	474	527,80	12,93	0,897	471,93	
13	544	565,65	14,31	0,953	515,12	
14	582	575,42	14,06	1,013	587,59	
15	681	597,33	14,49	1,137	670,14	
16	557	619,12	14,89	0,897	549,03	
17	628	654,74	16,04	0,953	603,98	
18	707	693,01	17,27	1,013	679,60	
19	773	685,35	15,89	1,137	807,47	
20	592	667,10	14,00	0,897	629,27	
21	627	662,26	12,96	0,953	648,84	
22	725	708,40	14,80	1,013	684,10	
23	854	746,22	16,07	1,137	822,16	
24	661	741,17	14,90	0,897	684,05	
25				0,953	720,26	1
26				1,013	781,12	2
27					893,41	3
28					718,59	4
29					777,04	5
30					841,50	6

No exemplo apresentado na tabela 2.1, os valores de α , β e γ que apresentaram o menor MSE foram, respectivamente 0,822, 0,055 e 0. O exemplo de como iniciar os valores de nível, tendência e sazonalidade, utilizando as equações (2.5), (2.6) e (2.7), respectivamente, é conforme segue:

$$L_4 = \frac{1}{4}(362 + 385 + 432 + 341) \Rightarrow L_4 = \frac{1520}{4} \Rightarrow L_4 = 380,$$

$$b_4 = \frac{1}{4} \left[\frac{382 - 362}{4} + \frac{409 - 385}{4} + \frac{498 - 432}{4} + \frac{387 - 341}{4} \right] \Rightarrow b_4 = \frac{1}{4}(39) \Rightarrow b_4 = 9,75,$$

$$S_1 = \frac{362}{380} \Rightarrow S_1 = 0,953, S_2 = \frac{385}{380} \Rightarrow S_2 = 1,013, S_3 = \frac{432}{380} \Rightarrow S_3 = 1,137 \text{ e}$$

$$S_4 = \frac{341}{380} \Rightarrow S_4 = 0,897.$$

É importante ressaltar que no exemplo demonstrado na tabela 2.1 o tamanho da sazonalidade (s) é de 4.

Pode-se observar o cálculo para encontrar a previsão do período 27:

$$F_{24+3} = (741,17 + 14,90 \cdot 3)1,137 \Rightarrow F_{24+3} = (785,87)1,137 \Rightarrow F_{24+3} = 893,41.$$

Observa-se que a complexidade aumenta quando uma seqüência de dados históricos apresenta tendência e sazonalidade, isso acaba exigindo melhores e complexos modelos para tratar essa junção de elementos. O modelo Holt-Winters é uma boa escolha quando se trata do exposto.

2.2. Média móvel centrada com sazonalidade

De acordo com Martins e Laugeni (2005), “existem diversos métodos para a realização de previsões quando o consumo é sazonal. Um dos mais utilizados é o método do coeficiente sazonal”.

Tubino (2000) lembra que “a sazonalidade caracteriza-se pela ocorrência de variações, para cima e para baixo, a intervalos regulares nas séries temporais da demanda. Deve existir uma razão plausível para a ocorrência, e posterior repetição, destas variações”.

De acordo com Corrar e Theóphilo (2004), “para a captação dos efeitos das variáveis sazonais, podemos calcular índices sazonais, que fornecem o valor percentual médio das observações nas estações sazonais observadas. Esse procedimento gera um ajuste na projeção

da tendência, aumentando sua acurácia, porque considera os impactos causados por valores sazonais ao longo de toda a série”.

Tubino (2000) menciona que “a forma mais simples de considerar a sazonalidade nas previsões da demanda consiste em empregar o último dado da demanda, no período sazonal em questão, e assumí-lo como previsão”.

Na tabela 2.2 pode ser encontrado um exemplo, onde a demanda é o número de refeições servidas por dia de semana.

Tabela 2.2 – Dados para exemplo da média móvel centrada com índice de sazonalidade. Fonte: Tubino (2000)

Dia	Demanda	Média móvel centrada	Índice de sazonalidade
Segunda-feira	50		
Terça-feira	55		
Quarta-feira	52		
Quinta-feira	56	$443/7 = 63,28$	$56/63,28 = 0,88$
Sexta-feira	65	$448/7 = 64$	$65/64 = 1,01$
Sábado	80	$443/7 = 63,28$	$80/63,28 = 1,26$
Domingo	85	$449/7 = 64,14$	$85/64,14 = 1,32$
Segunda-feira	55	$443/7 = 63,28$	$55/63,28 = 0,86$
Terça-feira	50	$448/7 = 64$	$50/64 = 0,78$
Quarta-feira	58	$443/7 = 63,28$	$58/63,28 = 0,91$
Quinta-feira	50	$438/7 = 62,57$	$50/62,57 = 0,79$
Sexta-feira	70	$435/7 = 62,14$	$70/62,14 = 1,12$
Sábado	75	$435/7 = 62,14$	$75/62,14 = 1,20$
Domingo	80	$431/7 = 61,57$	$80/61,57 = 1,29$
Segunda-feira	52	$441/7 = 63$	$52/63 = 0,82$
Terça-feira	50	$436/7 = 62,28$	$50/62,28 = 0,80$
Quarta-feira	54	$446/7 = 63,71$	$54/63,71 = 0,84$
Quinta-feira	60	$456/7 = 65,14$	$60/65,14 = 0,92$
Sexta-feira	65	$454/7 = 64,85$	$65/64,85 = 1,00$
Sábado	85	$457/7 = 65,28$	$85/65,28 = 1,30$
Domingo	90	$458/7 = 65,42$	$90/65,42 = 1,37$
Segunda-feira	50		
Terça-feira	53		
Quarta-feira	55		

Para calcular o índice de sazonalidade Tubino (2000) explica que é “dividindo-se o valor da demanda no período pela média móvel centrada neste período. O período empregado para o cálculo da média móvel é o ciclo da sazonalidade. Quando se dispõe de dados suficientes, calculam-se vários índices para cada período e tira-se uma média”.

De acordo com Tubino (2000), “quando o ciclo da sazonalidade for um número par, por exemplo, 12 meses, com o centro dos dados caindo no meio de um período, primeiro

calcula-se a média móvel centrada. A seguir, corrigem-se estes valores para que coincidam com os períodos analisados, fazendo a média de dois valores descentrados corresponderem a um valor centrado”.

Com os valores obtidos na tabela 2.2 pode-se calcular o índice de sazonalidade para cada dia da semana como a média dos índices encontrados:

$$I_{\text{seg}} = (0,86 + 0,82)/2 = 0,84$$

$$I_{\text{ter}} = (0,78 + 0,80)/2 = 0,79$$

$$I_{\text{qua}} = (0,91 + 0,84)/2 = 0,87$$

$$I_{\text{qui}} = (0,88 + 0,79 + 0,92)/3 = 0,86$$

$$I_{\text{sex}} = (1,01 + 1,12 + 1,00)/3 = 1,04$$

$$I_{\text{sáb}} = (1,26 + 1,20 + 1,30)/3 = 1,25$$

$$I_{\text{dom}} = (1,32 + 1,29 + 1,37)/3 = 1,32.$$

Este é o índice que deverá ser usado sobre a última média móvel para encontrar a previsão da demanda. Seguindo o método, o cálculo da previsão para o próximo dia da semana pode ser obtido conforme demonstrado:

$$F_{\text{qui}} = \left(\frac{60 + 65 + 85 + 90 + 50 + 53 + 55}{7} \right) \times I_{\text{qui}} \Rightarrow F_{\text{qui}} = 65,43 \times 0,86 \Rightarrow F_{\text{qui}} = 56,27.$$

Caso a previsão seja de vários passos à frente o valor previsto entra no cálculo da média móvel, enquanto a observação mais antiga deixa a equação.

2.3. Redes neurais artificiais

As redes neurais artificiais, originalmente desenvolvidas para mimetizar sistemas neurais biológicos básicos (o cérebro humano, basicamente), são compostas por um número de elementos simples de processamento interconectados denominados neurônios. (ZHANG *et al.*, 1998). Estes elementos simples, os neurônios, são, segundo Haykin (2001), implementados utilizando-se componentes eletrônicos ou através de simulação por programação em um computador digital.

A motivação para a utilização de redes neurais artificiais vem do conhecimento que o cérebro processa informações de uma maneira diferente de um computador digital tradicional, sendo esse um computador altamente complexo, não-linear e paralelo.

Uma rede neural é inspirada no cérebro humano em dois aspectos:

- i) o conhecimento é adquirido através de um processo de aprendizagem, com informações do próprio ambiente;
- ii) o peso de cada sinapse do neurônio representa o conhecimento adquirido (HAYKIN, 2001).

Abelém (1994) ressalta que embora o neurônio artificial não tenha algumas características do neurônio biológico, tais como atraso de tempo, efeito de sincronismo e modulação de frequência, os neurônios artificiais possuem atributos semelhantes aos neurônios biológicos e que em essência o funcionamento do neurônio biológico foi absorvido.

Cortez (2002) menciona que as redes neurais artificiais têm características únicas, não encontradas em outros métodos:

- i) aprendizagem e generalização, que diz respeito a conseguir descrever o todo a partir de algumas partes;
- ii) processamento maciçamente paralelo, fazendo com que tarefas complexas sejam realizadas em um curto espaço de tempo.

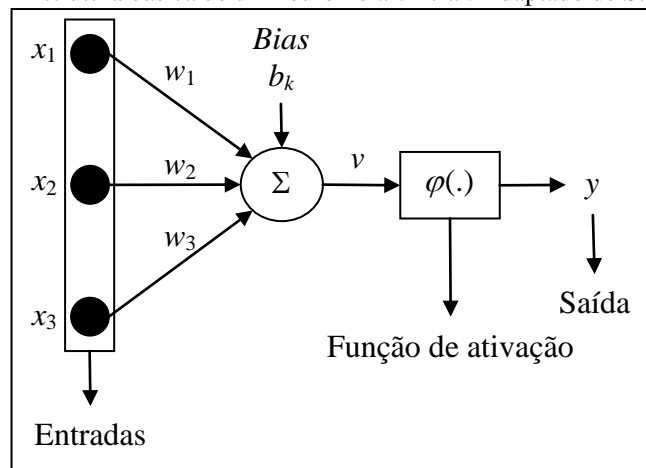
O neurônio é a unidade principal para uma rede neural artificial e sua estrutura básica é apresentada na figura 2.1. Assim como o neurônio biológico, o neurônio artificial apresenta divisões e similaridades com o neurônio biológico:

- i) Sinapses: representa a conexão de um neurônio com outro ou a conexão entre um valor de entrada e o neurônio. Para cada sinapse é atribuído um peso (ou força), sendo que toda informação que passa pela sinapse é multiplicada por esse peso, fazendo com que o conjunto de pesos represente a inteligência ou a memória de uma determinada rede neural. A influência de um neurônio sobre outro é o produto da saída do neurônio antecessor com o peso da sinapse que os conecta (SOTO, 1999). Haykin (2001) coloca que diferente do cérebro humano, o neurônio artificial aceita pesos sinápticos tanto negativos como positivos.
- ii) Somador: faz a somatória da multiplicação de cada valor de entrada pelo respectivo peso sináptico. Após a somatória ainda existe a opção de aplicar o *bias* sobre o resultado obtido.
- iii) Função de ativação: o resultado da somatória do passo anterior é passado como entrada em uma função de ativação, que tem o objetivo de limitar a amplitude da saída. Segundo Haykin (2001), normalmente o intervalo de saída, depois da

função de ativação, pode ser escrito como intervalo unitário fechado $[0, 1]$ ou $[-1, 1]$. Existem várias funções de ativação para uma rede neural, entre elas a mais usada, segundo Santos (2005), é a função sigmóide, definida por uma função estritamente crescente que exibe um balanceamento adequado entre comportamento linear e não-linear. A tabela 2.3 contém algumas funções de ativação.

A figura 2.1 mostra a estrutura básica de um neurônio artificial, que foi proposto inicialmente por McCulloch e Pitts (1943), onde x_1 , x_2 e x_3 representam as entradas do neurônio artificial, w_1 , w_2 e w_3 os pesos das sinapses, b_k o nível de *bias*. O valor da variável v , que é a entrada da função de ativação, é o somatório das entradas após a multiplicação pelos respectivos pesos (u) mais a aplicação do *bias*, antes de ser passado para a função de ativação. O resultado da função de ativação é a variável y , a saída do neurônio.

Figura 2.1 – Estrutura básica de um neurônio artificial. Adaptado de Santos (2005).



Matematicamente podem-se formular as seguintes equações:

$$u = \sum_{i=1}^n w_i \cdot x_i \quad (2.9)$$

$$v = u + b_k \quad (2.10)$$

$$y = \varphi(v) \Rightarrow y = \varphi\left(\left(\sum_{i=1}^n w_i \cdot x_i\right) + b_k\right). \quad (2.11)$$

A função do *bias* é aumentar ou diminuir o valor de entrada na função de ativação. Se o valor passado para a função de ativação é aumentado ou diminuído então isso acaba influenciando a função de ativação no seu resultado. Este é um recurso a mais para tentar ajustar a saída obtida com a saída esperada.

Tabela 2.3 – Funções de ativação. Adaptado de Cortez (1997) e Santos (2005).

Nome	Função	Domínio
Limiar	$\varphi(v) = \begin{cases} v \geq 0 \Rightarrow 1 \\ v < 0 \Rightarrow 0 \end{cases}$	[0,1]
Linear	$\varphi(v) = v$	$] -\infty, +\infty [$
Sigmóide	$\varphi(v) = \frac{1}{1 + \exp(-av)}$	[0,1]
Sigmóide 1	$\varphi(v) = \frac{2}{1 + \exp(-av)} - 1$	[-1,1]
Sigmóide 2	$\varphi(v) = \frac{v}{1 + v }$	[-1,1]
Seno	$\varphi(v) = \sin(v \bmod 2\pi)$	[-1,1]
Cosseno	$\varphi(v) = \cos(v \bmod 2\pi)$	[-1,1]
Gaussiana	$\varphi(v) = e^{\left(\frac{-x^2}{2}\right)}$	[-1,1]
Saturação	$\varphi(v) = \begin{cases} v > 0 \Rightarrow +1 \\ v \in [-1,1] \Rightarrow v \\ v < 0 \Rightarrow -1 \end{cases}$	[-1,1]
Tangente hiperbólica	$\varphi(v) = \frac{e^v - e^{-v}}{e^v + e^{-v}}$	[-1,1]

A arquitetura de uma rede neural artificial é composta de pelo menos três camadas, os neurônios das camadas intermediárias e da camada de saída, os nós de entrada, o tipo de conexão de cada camada e a topologia da rede. O neurônio, que seria a menor unidade da rede neural, compõe as camadas intermediárias e de saída. Cada camada intermediária e de saída tem pelo menos um neurônio, sendo que pelo menos duas camadas estão sempre presentes em uma rede neural: a camada de entrada e a camada de saída. Existe a possibilidade de camadas intermediárias sejam adicionadas para permitir que a geração da saída seja através de uma função contínua e não-linear das entradas.

As redes neurais são ferramentas computacionais flexíveis em um ambiente dinâmico. Elas têm capacidade de aprender rapidamente padrões complexos e tendências presentes nos dados e de se adaptar rapidamente às mudanças, características essas que são desejáveis em se tratando de previsão de séries temporais (CASTRO, 2001).

Essa habilidade de realizar previsões para séries temporais, reconhecimento de padrões, entre outras aplicações das redes neurais só é possível por causa do processo de aprendizagem sobre o ambiente que ela está inserida. Haykin (2001) entende que esse

processo de aprender a partir do ambiente em que a rede neural está inserida e melhorar seu desempenho através da aprendizagem tem importância primordial.

Aprender, neste caso, está relacionado a encontrar os melhores valores para os pesos sinápticos de cada neurônio, bem como o melhor valor para o nível de *bias*. Já o processo de aprendizagem diz respeito à busca desses melhores valores. Cada nova iteração nesse processo de aprendizagem altera os valores dos pesos sinápticos e do nível de *bias*, tornando a rede mais inteligente.

Santos (2005) resume o processo de aprendizagem: “deve-se encontrar o ajuste de peso sináptico e do nível de *bias* que produzam o menor nível de erro entre a resposta desejada e a resposta estimada pela rede (no caso de uma série temporal) ou a melhor fronteira de separação (decisão) entre dois padrões presentes em um conjunto de dados.”

Basicamente existem dois paradigmas de aprendizagem:

- i) aprendizagem supervisionada: necessita dos valores de entrada e suas respectivas saídas esperadas. A cada iteração desse conjunto de entradas e saídas é encontrado um valor de erro, utilizado para corrigir os pesos sinápticos e níveis de *bias*;
- ii) aprendizagem não-supervisionada: somente os valores de entrada são esperados. Para este caso a rede tenta descobrir padrões nos valores de entrada através de identificação de regularidades estatísticas. Dessa forma a rede neural automaticamente adapta os pesos sinápticos para representar e agrupar os padrões de entrada com características semelhantes.

Haykin (2001) lista algumas propriedades das redes neurais:

- i) não linearidade: um neurônio pode ser linear ou não linear. Quando a RN é constituída de conexões entre neurônios não lineares, ela mesma é não linear. Esta propriedade torna-se mais importante se o mecanismo gerador das entradas da rede for inerentemente não linear (como é o caso de um sinal de voz).
- ii) Mapeamento de entrada-saída: diz respeito a forma de aprendizagem com um professor ou supervisionada. É quando a RN é treinada com um conjunto de dados de entrada e os respectivos dados de saída, fazendo com que os pesos sinápticos sejam acertados de forma a criar um mapeamento entre os dados de entrada e os dados esperados de saída.

- iii) Adaptabilidade: as RNs têm uma capacidade inata de adaptar os pesos sinápticos de acordo com modificações no meio ambiente. Desta forma, uma rede pode ser facilmente retreinada caso haja alguma alteração no ambiente operacional. Também é possível que os pesos sejam reajustados em tempo real, para o caso de ambientes não-estacionários.
- iv) Resposta a evidências: no contexto de classificação de padrões, uma RN pode ser projetada para fornecer informação não somente sobre qual melhor padrão particular selecionar, mas também sobre a confiança ou crença na decisão tomada.
- v) Informação contextual: o conhecimento é representado pela própria estrutura e estado de ativação de uma RN. Cada neurônio da rede é potencialmente afetado pela atividade de todos os outros neurônios na rede. Conseqüentemente, a informação contextual é tratada naturalmente pela RN.
- vi) Tolerância a falhas: uma RN, quando implementada na forma física (em *hardware*), tem o potencial de ser inerentemente tolerante a falhas, ou capaz de realizar computação robusta, no sentido que seu desempenho de degrada suavemente sob condições de operação adversas.
- vii) Implementação em *Very Large Scale Integration*: a natureza maciçamente paralela de uma RN a faz ser potencialmente rápida na computação de certas tarefas. Esta mesma característica torna uma RN adequada para implementação utilizando tecnologia de integração em escala muito ampla.
- viii) Uniformidade de análise e projeto: a mesma noção é utilizada em todos os domínios envolvendo a aplicação de RNs.
- ix) analogia neurobiológica: o projeto de uma RN é motivado pela analogia com o cérebro: uma prova que o processamento paralelo tolerante a falhas é não somente possível mas também rápido e poderoso.

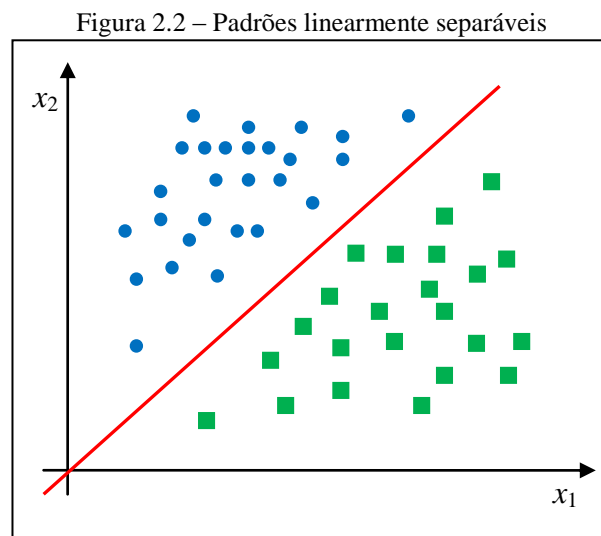
A construção de um modelo de redes neurais artificiais envolve desde a modelagem adequada até as transformações utilizadas para transmitir os dados a ela e os métodos utilizados para interpretar os resultados obtidos. Esses três aspectos (modelagem, transformações e interpretação) são fundamentais na utilização das redes neurais artificiais para, não somente, mas também, previsão de preços (BRESSAN, 2004).

Nas próximas seções são detalhadas as RNs *perceptron* multicamadas e função de base radial.

2.3.1 *Perceptron* de múltiplas camadas

Rosenblatt (1958) propôs o *perceptron* como primeiro modelo para aprendizagem com um professor. Segundo Haykin (2001), o *perceptron* é a forma mais simples de uma rede neural usada para a classificação de padrões denominados linearmente separáveis. O *perceptron* de uma única camada consiste de um único neurônio com pesos sinápticos e *bias*. O procedimento de aprendizagem pode ser o desenvolvido pelo próprio Rosenblatt (1958, 1962), no entanto esse modelo de *perceptron* de um único neurônio é limitado a realizar classificação de padrões com apenas duas classes. A figura 2.2 mostra um exemplo de classes linearmente separáveis.

Para resolver o problema de padrões não linearmente separáveis é necessário adicionar mais camadas na estrutura do *perceptron*.



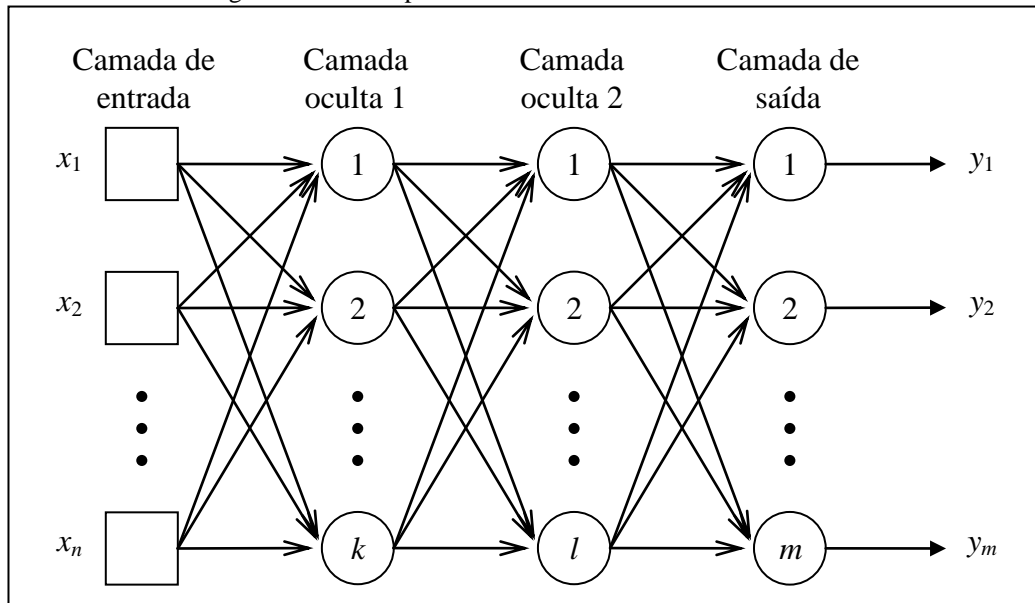
A figura 2.3 mostra a estrutura básica de uma rede neural MLP, com uma camada de entrada, duas camadas ocultas e uma camada de saída. Esta mesma estrutura pode ser representada na forma de equação, segundo Santos (2005), conforme apresentada na equação (2.12),

$$y = \varphi_s \left\{ \sum_{i=1}^m w_i \varphi_i \left(\sum_{j=1}^n w_{ij} \varphi_j + b_j \right) + b_s \right\}. \quad (2.12)$$

Santos (2005) define três aspectos básicos do projeto de uma rede neural MLP:

- i) a determinação do número de camadas ocultas;
- ii) a determinação do número de neurônio em cada uma das camadas; e
- iii) a especificação dos pesos sinápticos que interconectam os neurônios nas diferentes camadas da rede neural.

Figura 2.3 – Exemplo de estrutura de uma rede neural MLP



Segundo Haykin (2001), a camada oculta não faz parte nem da entrada ou saída da rede. A camada oculta e seus neurônios são responsáveis por fazer a rede aprender tarefas complexas extraindo progressivamente as características significativas dos padrões de entrada.

No entanto, segundo De Castro e De Castro (2001), uma desvantagem é que não há regras determinadas para especificar o número de camadas ocultas, tão pouco para determinar o número de neurônios em cada camada. Outra desvantagem relacionada às redes neurais MLP diz respeito ao alto grau de conectividade da rede, que impede que se extraiam as regras ou o vetor de pesos utilizado. Esta última restrição faz com que as redes neurais sejam classificadas, às vezes, de modelos caixa-preta.

Segundo Santos (2005), as etapas do *error backpropagation* (retropropagação do erro), proposto por Rumelhart *et al.* (1986), são:

- i) iniciar aleatoriamente os pesos das sinapses da rede neural;

- ii) apresentar à rede o conjunto de treinamento constituído das entradas e das saídas desejadas;
- iii) calcular a saída da rede neural, propagando a saída dos neurônios de cada camada da rede neural para os neurônios da próxima camada, passando por suas respectivas funções de ativação;
- iv) especificar as saídas desejadas e calcular os erros das camadas;
- v) ajustar os pesos da RN, baseando-se nos valores de erro, começando pelas unidades de saída e propagando-se em direção à primeira camada oculta;
- vi) repetir as etapas (iv) e (v) até que o erro obtido apresente um valor de tolerância aceitável (critério de parada de treinamento);
- vii) utilizar um conjunto de validação e/ou teste para analisar o desempenho da rede.

Algumas deficiências do algoritmo retropropagação do erro são discutidas por Guerra e Coelho (2002) e Krose e van der Smagt (1996), onde eles apontam que este algoritmo tem uma baixa velocidade de aprendizado e um acentuado custo computacional por consequência, além da possibilidade do algoritmo convergir para mínimos locais sem que o mínimo global seja atingido.

A alternativa para as deficiências apontadas é a utilização do algoritmo proposto por Levenberg e Marquardt, chamado de algoritmo Levenberg-Marquardt (Levenberg, 1944; Marquardt, 1963). Esse algoritmo utiliza a segunda derivada da superfície (taxa de mudança da curvatura) em detrimento da primeira derivada (a curvatura). Segundo Guerra e Coelho (2002), esse algoritmo é uma alternativa viável ao tradicional algoritmo de retropropagação do erro. As etapas podem ser separadas da seguinte forma:

- i) iniciar a rede neural MLP e apresentar todas as entradas e saídas;
- ii) calcular a resposta da rede neural;
- iii) calcular a soma dos erros quadráticos;
- iv) calcular a matriz Jacobiana;
- v) resolver a equação Gauss-Newton para a atualização dos pesos sinápticos (w);
- vi) parar quando um número de épocas foi realizado ou quando a soma do erro quadrático (MSE, *mean squared error*) encontrado está menor de um valor pré-determinado.

Freitas e Souza (2002) apontam um problema intrínseco da rede neural MLP, o *overfitting*, que é quando a rede especializa-se nos padrões de treinamento e perde sua

capacidade de generalização. É sugerida a técnica de *early stopping* para contornar o problema, que consiste em treinar a rede neural com uma determinada amostra de dados, denominada grupo de treinamento, e validar o desempenho da rede, periodicamente, utilizando outra amostra de dados chamada de grupo de validação. Se os resultados obtidos com a validação atingirem um nível satisfatório, o treinamento é interrompido. Existe ainda outro grupo de dados para avaliar se a rede neural é capaz de solucionar o problema, chamado grupo de teste.

2.3.2. Função de base radial

As RNs-RBF são de uma classe de redes neurais com uma camada oculta, onde cada neurônio é ativado por uma função de base radial, de onde é derivado seu nome. Os neurônios na camada oculta fornecem um conjunto de funções não-lineares, de base radial, que constituem uma base arbitrária para os padrões de entrada (SANTOS, 2005).

Esse tipo de rede foi originalmente desenvolvido para interpolação de dados em espaços multidimensionais, conforme diz Castro (2001). Esse problema pode ser formulado pela seguinte equação:

$$y_i = F(u_i), \quad (2.13)$$

onde y_i é um conjunto de escalares e u_i é um conjunto de vetores. Desta forma, se a função $F(\cdot)$ for definida analiticamente ela pode mapear valores do vetor u_i que não pertenciam ao conjunto original de treinamento, no conjunto de escalares y_i . Para que a função $F(u_i)$ seja uma solução analítica ela pode ser definida da seguinte forma:

$$F(u) = \sum_{i=1}^n w_i \varphi(\|u - u_i\|^2), \quad (2.14)$$

onde w_i é um coeficiente (o peso da conexão neste caso) e $\varphi(\|u - u_i\|^2)$ representa uma função radial. Existindo n funções radiais, o centro de cada função é dado por u_i . Desta forma, $\|u - u_i\|$ representa a distância Euclidiana entre o centro de cada função radial (u_i) e do vetor de entrada (u). Uma lista das principais funções de base radial é apresentada na tabela 2.4.

Haykin (2001) explica que quanto menor a distância Euclidiana entre o vetor de entrada e o centro de um neurônio da camada oculta, maior é a similaridade entre os dois. É por isso que, por exemplo, o resultado de uma função radial gaussiana é alto quando a

distância euclidiana é maior, assim como o resultado é baixo quando a distância Euclidiana é menor.

O parâmetro σ das funções de base radial vai determinar quão rapidamente o resultado da função radial usada se aproxima de zero à medida que a distância euclidiana aumenta. De Castro e De Castro (2001) entendem que, na função gaussiana, o parâmetro σ corresponde ao desvio padrão da respectiva função, sendo que o desvio padrão mede o espalhamento dos dados em torno do centro da função.

Tabela 2.4 – Principais funções de base radial. Fonte: Santos (2005).

Lâmina <i>spline</i> fina	$\varphi(u) = \frac{u - u_i}{\sigma^2} \log\left(\frac{u - u_i}{\sigma}\right), \sigma > 0$
Multi-quadrática	$\varphi(u) = \sqrt{(u - u_i)^2 + \sigma^2}, \sigma > 0$
Multi-quadrática inversa	$\varphi(u) = \frac{1}{\sqrt{(u - u_i)^2 + \sigma^2}}, \sigma > 0$
Gaussiana	$\varphi(u) = \exp\left(-\frac{(u - u_i)^2}{2\sigma^2}\right), \sigma > 0$

Antes de Broomhead e Lowe (1988), as tentativas de utilização de uma RN-RBF usavam tantas funções de base radial quantos fossem os padrões dos dados apresentados à camada de entrada da rede neural, com o objetivo de uma aproximação mais exata. No entanto isso ocasionava um alto custo computacional para a utilização da rede, além do problema de interpolação exata, indesejável em aproximação de funções. Broomhead e Lowe (1988) sugeriram então que uma RN-RBF tivesse menos funções de base radial na camada oculta do que o número de vetores de entrada, aumentando assim a generalização e diminuindo o custo computacional.

Na sua forma mais básica, as RNs-RBF são compostas por três camadas: camada de entrada, camada oculta e camada de saída. A primeira camada, de entrada, tem o objetivo de receber os sinais do ambiente exterior e passá-los para a próxima camada. A segunda camada, a camada oculta, é formada por um conjunto de neurônios, sendo que cada neurônio é associado a um vetor de parâmetros denominado de centro. Cada neurônio calcula a distância Euclidiana entre o respectivo centro e o valor passado pela camada de entrada com base em uma função de base radial. Como a função de base radial é não-linear, como a função gaussiana, a camada oculta passa a ser não-linear. A terceira camada é a camada de saída, que

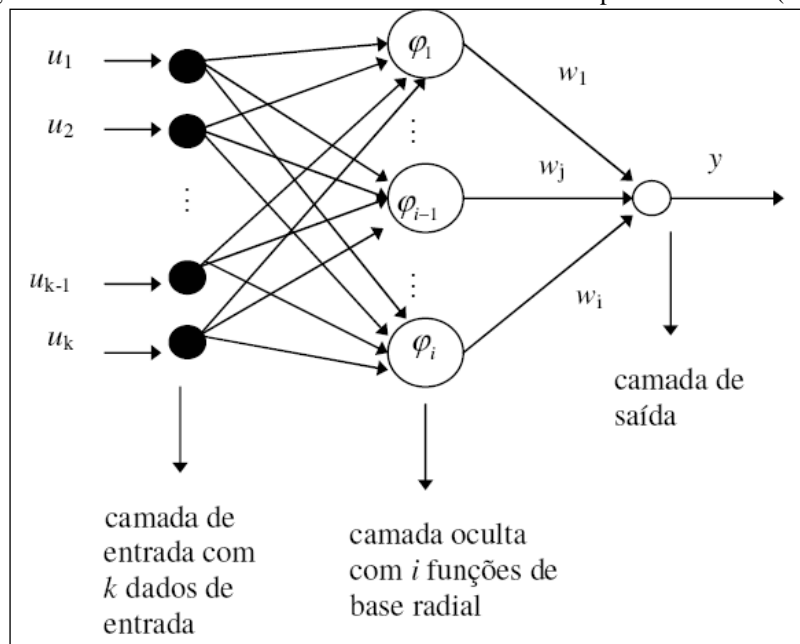
é a resposta linear do mapeamento feito pela camada oculta. É um mapeamento linear dos valores não-lineares da camada oculta. A figura 2.4 mostra a estrutura básica de uma RN-RBF (PAIVA, 1999).

A saída de uma rede neural RBF, usando uma função gaussiana na camada oculta, pode ser expressa na equação (2.15), tal que

$$y = \sum_{i=1}^n w_i \left(\exp \left(-\frac{\|u - u_i\|^2}{2\sigma^2} \right) \right), \quad (2.15)$$

onde w_i é o peso da sinapse entre a camada oculta e a camada de saída, $\exp \left(-\frac{\|u - u_i\|^2}{2\sigma^2} \right)$ é a função gaussiana presente na camada oculta e y representa a saída da rede.

Figura 2.4 – Estrutura básica de uma rede neural RBF. Adaptado de Santos (2005).



O processo de aprendizagem de uma RN-RBF é a determinação dos parâmetros livres da rede: a variância σ_i^2 , os centros u_i e os pesos sinápticos w_i . Essa determinação deve ser realizada de tal forma que a saída y seja a mais próxima da saída desejada.

Existem vários algoritmos para atualizar cada parâmetro livre de uma RN-RBF. Santos (2005) e Castro (2001) citam como exemplo os algoritmos genéticos, mapas de Kohonen, algoritmo de mínimos quadrados ortogonais, filtro de Kalman, *fuzzy c-means*, *k-means*, algoritmo de Moore-Penrose e Gradiente Estocástico.

É importante ressaltar que o aprendizado de uma RN-RBF é em parte não-supervisionado e em parte supervisionado. Os centros dos neurônios da camada oculta são determinados por métodos como *fuzzy c-means* ou *k-means*, métodos estes que não comparam ou esperam uma qual seria a saída esperada para determinar o centro, sendo desta forma um método não supervisionado. Enquanto os pesos das sinapses que ligam a camada oculta à camada de saída são ajustados por algoritmos como mínimos quadrados em batelada, que exige conhecer previamente qual será a saída esperada, sendo desta forma um método de aprendizagem supervisionado.

A tabela 2.5 mostra quais algoritmos podem ser aplicados a quais parâmetros livres.

Tabela 2.5 – Algoritmos de aprendizagem para ajuste de parâmetros livres. Adaptado de Castro (2001).

Variância	Centros	Pesos sinápticos
Constante: por conhecimento prévio e inferência a partir do conjunto de vetores de treino	Constante: por conhecimento prévio e inferência a partir do conjunto de vetores de treino	Gradiente Estocástico.
Gradiente Estocástico.	“Clusterização” pelo algoritmo <i>k-means</i> . Não-supervisionado.	Pseuda Inversa por decomposição em valores singulares.
<i>p-nearest neighbor</i>		

A combinação desses métodos para a determinação dos parâmetros livres de uma RN-RBF é que vai ser o processo de aprendizagem da rede. Conforme Santos (2005), os métodos mais comumente utilizados para o ajuste dos parâmetros livres são o *k-means* para o início e atualização dos centros e o mínimos quadrados para atualização dos pesos sinápticos. A seguir é descrito a utilização do método *k-means* e mínimos quadrados para ajuste dos centros e pesos sinápticos, respectivamente. O algoritmo *p-nearest neighbor*, sugerido por Moody e Darken (1989), é usado para calcular a variância de cada centro.

As etapas do procedimento *k-means*, segundo Santos (2005), são as seguintes:

- i) escolher aleatoriamente a posição inicial dos M_{ui} centros dos *clusters* (grupos) para $j=1, \dots, M_{ui}$, aleatoriamente, no conjunto de dados de entrada;
- ii) para $i=1, \dots, N$, onde N é o número de amostra do conjunto de dados, encontrar o centro u_k tal que a distância do vetor de entrada x_i seja mínima, ou seja, $\|x_i - u_k\| = \min \{ \|x_i - u_j\| \}$, para $j=1, \dots, M_{ui}$ e atribuir a amostra i para o *cluster* k ;

- iii) calcular a média dos pontos de dados pertencentes a cada *cluster*:

$$\bar{x}_j = \frac{1}{N_j} \sum_{i \in u_j} x_i, \text{ para } i=1, \dots, N, \quad (2.16)$$

onde N_j é o número de amostras pertencente ao *cluster* j . O novo centro de cada *cluster* é então atualizado como sendo a média de seus pontos de dados, ou seja,

$$u_j = \bar{x}_j;$$

- iv) repetir as etapas (ii) e (iii) até que os centros dos *clusters* e a pertinência converjam.

O algoritmo *p-nearest neighbor* é usado para determinar a variância σ_i^2 das funções de base radial. Depois de encontrados os centros de cada função, o algoritmo *p-nearest neighbor* calcula a largura das funções como sendo a média sobre todas as distâncias euclidianas entre o centro de cada unidade de processamento n_i e o centro da unidade n_j mais próximo:

$$\sigma_j = \frac{1}{p} \sum_{k=1}^p \|u_j - u_k\|. \quad (2.17)$$

Para o ajuste dos pesos sinápticos usa-se o estimador de mínimos quadrados ordinários, onde a matriz de interpolação pode ser descrita da seguinte forma:

$$\phi \cdot w = y. \quad (2.18)$$

E a equação (2.19) do estimador dos mínimos quadrados ordinários aplicada para ajustar os pesos sinápticos pode então ser descrita como:

$$w = (\phi^T \phi)^{-1} \phi \cdot y. \quad (2.19)$$

2.4. Índices de desempenho

Para Makridakis *et al.* (1998) em muitos casos o termo acuracidade refere-se ao quão bem um modelo de previsão se adapta a dados que já eram conhecidos, no entanto, para o consumidor das previsões, acuracidade representa o quão bem um modelo consegue minimizar os erros de uma previsão onde os dados ainda não são conhecidos.

Existe uma variedade de técnicas para testar a acuracidade de um modelo de previsão. Nesta seção serão apresentadas algumas dessas técnicas.

2.4.1. Medidas estatísticas padrões (ME, MAE, MSE e MAPE)

Makridakis *et al.* (1998) explica que o erro é definido pela diferença entre uma observação no tempo t e a previsão para o mesmo tempo t . Chega-se à seguinte equação:

$$e_t = Y_t - F_t. \quad (2.20)$$

Normalmente F_t é calculado usando-se os valores de Y_1 até Y_{t-1} , para os casos onde apenas um valor de previsão é gerado chama-se de previsão de um passo à frente. O erro também recebe o nome de erro da previsão de um passo à frente.

Para os casos onde existe previsão para n períodos à frente e também n observações à frente, poderão ser calculados n valores de erro, um para cada período. Para ambos os casos as seguintes equações de estatística padrão podem ser definidas:

$$\text{Mean Error (ME, Erro Médio)} = \frac{1}{n} \sum_{t=1}^n e_t, \quad (2.21)$$

$$\text{Mean Absolute Error (MAE, Erro Médio Absoluto)} = \frac{1}{n} \sum_{t=1}^n |e_t|, \text{ e} \quad (2.22)$$

$$\text{Mean Squared Error (MSE, Erro Médio Quadrático)} = \frac{1}{n} \sum_{t=1}^n e_t^2. \quad (2.23)$$

A medida de acuracidade dessas estatísticas depende da escala dos dados utilizada, desta maneira não facilitando a comparação entre séries temporais de diferentes áreas ou diferentes períodos.

Para comparar a acuracidade entre duas diferentes séries temporais ou intervalos de tempo é necessário usar uma medida do erro relativa ou percentual.

Primeiramente define-se um erro percentual através da seguinte equação:

$$\text{Percentage Error (PE, Erro Percentual)} = \left(\frac{Y_t - F_t}{Y_t} \right) \times 100. \quad (2.24)$$

A partir da equação (2.24) podem-se definir outras duas equações utilizadas:

$$\text{Mean Percentage Error (MPE, Erro Médio Percentual)} = \frac{1}{n} \sum_{t=1}^n PE_t, \text{ e} \quad (2.25)$$

$$\begin{aligned} &\text{Mean Absolute Percentage Error (MAPE, Erro Médio Percentual Absoluto)} \\ &= \frac{1}{n} \sum_{t=1}^n |PE_t|. \end{aligned} \quad (2.26)$$

É importante ressaltar que quando a série temporal contém observações de valor zero não é possível calcular o erro percentual, isso devido à divisão na equação (2.24).

Gujarati (2006) entende que o coeficiente R^2 vai dar uma idéia da qualidade do ajustamento entre a linha de regressão e o conjunto de dados. Aplicando-se nas séries temporais pode-se dizer que é quão bem os valores estimados se ajustam na série observada. O R^2 pode ser obtido através da equação (2.27).

$$R^2 = 1 - \frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2}. \quad (2.27)$$

Capítulo 3

Metodologia

Neste capítulo é apresentada caracterização da pesquisa, as séries temporais escolhidas, os modelos de previsão escolhidos, os índices de desempenho e a metodologia como foram aplicados os modelos de previsão nas séries temporais.

3.1. Caracterização da pesquisa

Segundo JUNG (2004), as pesquisas científicas podem ser classificadas quanto à natureza, objetivos, procedimentos, fonte de referências, além de ambiente e tempo de aplicação.

Em relação à natureza, esta pesquisa enquadra-se como tecnológica (JUNG, 2004), uma vez que conhecimentos básicos são aplicados e novos conhecimentos são gerados como resultado do processo de pesquisa.

Quanto aos objetivos, esta pesquisa apresenta-se como explicativa. As pesquisas explicativas, segundo JUNG (2004), visam à “identificação dos fatores que contribuem para a ocorrência dos fenômenos ou variáveis que afetam o processo”.

Referente aos procedimentos, essa pesquisa caracteriza-se como operacional. Segundo JUNG (2004), a pesquisa operacional “tem por princípio a investigação de forma sistemática e racional dos processos envolvidos” e utiliza ferramentas estatísticas e métodos matemáticos para obtenção da melhor solução. Dessa forma, observa-se que essa pesquisa utiliza análises estatísticas e métodos matemáticos para a avaliação dos resultados.

Com relação à fonte de referências, essa pesquisa é bibliográfica, pois segundo JUNG (2004), a pesquisa bibliográfica procura identificar as diferentes contribuições existentes na

literatura sobre o tema. Neste estudo, revisa-se a literatura a respeito de métodos lineares e não-lineares de previsão de séries temporais.

O ambiente de aplicação desta pesquisa é laboratorial. A pesquisa em laboratório, segundo JUNG (2004), “caracteriza-se pela possibilidade de se controlar as variáveis que possam interferir no experimento”.

Além disso, o fato de utilizar os preços de fechamento dos índices Bovespa, Dow Jones e Nasdaq dentro de um intervalo de tempo determinado, no caso entre Janeiro de 1995 e Agosto de 2008, faz com que esta pesquisa seja caracterizada quanto ao tempo de aplicação como longitudinal (JUNG, 2004).

3.2. Séries Temporais

As séries financeiras escolhidas foram:

- Ibovespa;
- Dow Jones; e
- Nasdaq.

Além disso, a série temporal Mackey-Glass foi selecionada a fim de servir como problema-teste.

A escolha dos índices baseou-se na importância que todos eles têm no cenário econômico nacional e internacional. Cada série temporal é apresentada detalhadamente nas próximas seções.

As séries temporais financeiras dos índices Bovespa, Dow Jones e Nasdaq passaram por uma normalização nos dados a fim de deixar os valores de cada série dentro do intervalo de 0 e 1. Procedimento necessário para ser possível a utilização de modelos de redes neurais na previsão. Para tanto cada série passou por uma transformação independente, onde todos os valores da série foram divididos pelo valor mais alto da série. Importante também ressaltar que os índices de desempenho foram calculados com base nos valores reais das séries.

3.2.1. Mackey-Glass

A saída do sistema Mackey e Glass (1977) é um dos estudos de caso mais conhecidos e utilizados na identificação de sistemas e previsões de séries temporais. Entre os vários autores que já usaram esta série, pode-se relacionar Jones *et al* (1990), Hsieh (1990), Gómez-

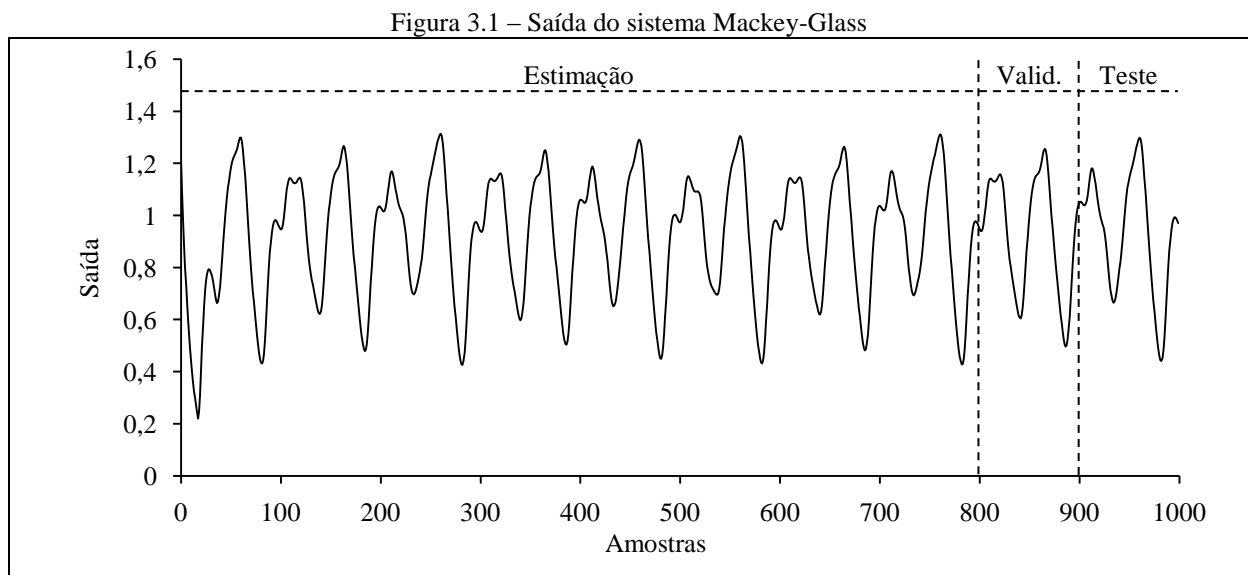
Ramírez *et al* (2007), Billings e Hong (1998), Paiva (1999), Boné e Crucianu (2002) e Jang(2003).

O sistema Mackey-Glass é um sistema caótico. Um sistema caótico também é conhecido como sistema dinâmico, que pode ser definido, segundo Alligood (2000), como um conjunto de estados possíveis, juntamente com uma regra que determina o estado presente em termos do estado passado, significando que o mesmo deve ser determinístico. Ser determinístico é poder determinar o estado presente unicamente baseado no estado passado. Outra característica dos sistemas dinâmicos é que eles são altamente sensíveis às condições iniciais.

A saída do sistema Mackey-Glass foi gerada nos mesmos moldes de Jang (1993). A equação diferencial com atraso (3.1) foi usada para tanto. A equação é dada por:

$$\dot{x}(t) = \frac{0,2x(t-\tau)}{1+x^{10}(t-\tau)} - 0,1x(t) \quad (3.1)$$

Onde t representa o tempo e τ o parâmetro da série caótica. Para obter os valores da saída do sistema a cada ponto inteiro foi aplicado o método Runge-Kutta de quarta ordem para encontrar uma solução numérica para a equação (4.1). O passo de tempo utilizado foi 0,1, a condição inicial $x(t) = 1,2$, $\tau = 17$ e $x(t)$ foi gerado para $0 \leq x \leq 999$. Foi também assumido $x(t) = 0$ quando $t < 0$. A figura 3.1 apresenta a saída do sistema Mackey-Glass.



Desta forma uma série temporal com 2000 amostras foi gerada, tendo-se selecionado apenas as amostras do tempo t 118 até 1117, ficando a série temporal com apenas 1000 amostras.

3.2.2. Índice Bovespa

Segundo a Bovespa (2008), o índice Bovespa é o mais importante indicador do desempenho médio das cotações do mercado de ações brasileiro. Sua relevância advém do fato do Índice Bovespa retratar o comportamento dos principais papéis negociados na Bovespa e também de sua tradição, pois o índice manteve a integridade de sua série histórica e não sofreu modificações metodológicas desde sua implementação em 1968.

O índice Bovespa representa o valor atual, em moeda corrente, de uma carteira teórica de ações constituída em 02/01/1968 (valor-base: 100 pontos), a partir de uma aplicação hipotética¹. Supõe-se não ter sido efetuado nenhum investimento adicional desde então, considerando-se somente os ajustes efetuados em decorrência da distribuição de proventos pelas empresas emissoras (tais como reinversão de dividendos recebidos e do valor apurado com a venda de direitos de subscrição, e manutenção em carteira das ações recebidas em bonificação). Dessa forma, o índice reflete não apenas as variações dos preços das ações, mas também o impacto da distribuição dos proventos, sendo considerado um indicador que avalia o retorno total de suas ações componentes.

Confiável e com uma metodologia de fácil acompanhamento pelo mercado, o índice Bovespa representa fielmente o comportamento médio das principais ações transacionadas, e o perfil das negociações à vista observadas nos pregões da Bovespa.

A finalidade básica do índice Bovespa é a de servir como indicador médio do comportamento do mercado. Para tanto, sua composição procura aproximar-se o mais possível da real configuração das negociações à vista (lote-padrão) na Bovespa.

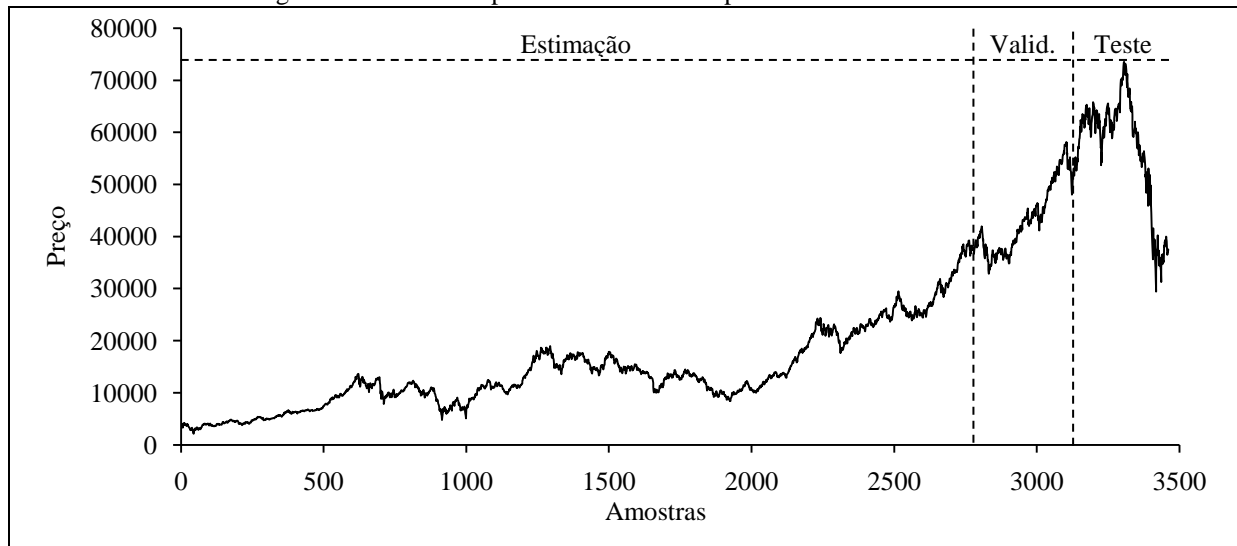
Em termos de liquidez, as ações integrantes da carteira teórica do índice Bovespa respondem por mais de 80% do número de negócios e do volume financeiro verificados no mercado à vista (lote-padrão) da Bovespa.

A Bovespa é responsável pela gestão, cálculo, difusão e manutenção do Índice Bovespa.

¹ Segundo a Bovespa (2008), o índice sofreu, unicamente para efeito de divulgação e sem prejuízo de sua metodologia de cálculo, as seguintes adequações: 1– divisão por 100, em 03/10/1983; 2– divisão por 10, em 02/12/1985; 3– divisão por 10, em 29/08/1988; 4– divisão por 10, em 14/04/1989; 5– divisão por 10, em 12/01/1990; 6– divisão por 10, em 28/05/1991; 7– divisão por 10, em 21/01/1992; 8– divisão por 10, em 26/01/1993; 9– divisão por 10, em 27/08/1993; 10– divisão por 10, em 10/02/1994; 11 – divisão por 10, em 03/03/1997.

A figura 3.2 apresenta a série temporal do índice Bovespa de 01/01/1995 até 31/12/2008.

Figura 3.2 – Série temporal do índice Bovespa utilizada neste trabalho



Para fazer a coleta dos dados utilizou-se a base de dados do *site* de finanças da Yahoo: www.finance.yahoo.com.

Foram coletados os dados relativos ao Índice Bovespa a partir da data 01/01/1995 até a data 31/12/2008. A periodicidade dos dados é diária e o preço utilizado é o de fechamento. Os dias onde não houve pregão, como feriados e fins de semana, foram removidos do universo.

Desta maneira o universo é compreendido pelos dias onde houve negociação entre o dia 01/01/1995 até o dia 31/12/2008, totalizando um montante de 3462 amostras.

A data de início em 01/01/1995 foi escolhida devido ao fato do plano Real ter entrado em vigor no ano anterior e alterado drasticamente a economia brasileira.

3.2.3. Índice Dow Jones

Segundo DowJones (2009), o índice Dow Jones é o mais difundido em jornais, na televisão e na internet. Também foi o primeiro índice a ser publicado nos Estados Unidos da América, ganhando fama depois da segunda guerra mundial, quando todas as atenções se voltaram para o desempenho excepcional da bolsa de valores norte americana. Além do fator da longevidade, outros dois fatores contribuem para sua popularidade: o índice é fácil de

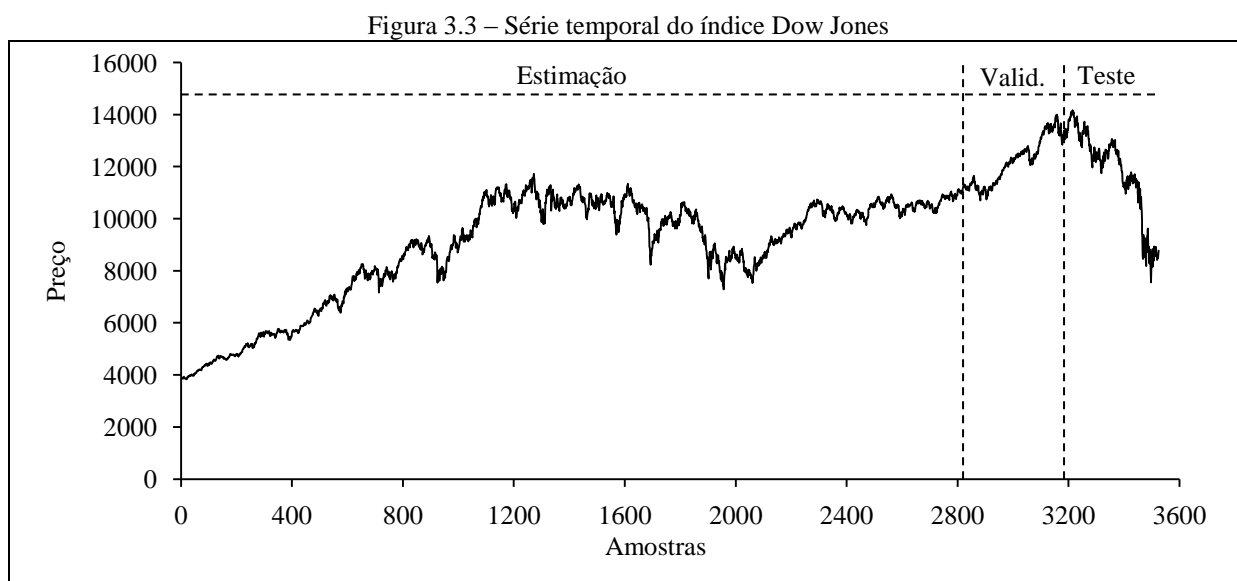
entender para a maioria das pessoas e representa com confiabilidade a tendência básica do mercado.

Ainda segundo DowJones (2009), o índice é composto por 30 empresas atualmente. Sendo dois terços das empresas indústrias manufatureiras e de bens de consumo. As empresas restantes representam serviços financeiros, de entretenimento, tecnologia da informação e outros setores.

Para fazer a coleta dos dados utilizou-se a base de dados do *site* de finanças da Yahoo: www.finance.yahoo.com.

Foram coletados os dados relativos ao Índice Dow Jones (^DJIA) a partir da data 01/01/1995 até a data 31/12/2008. A periodicidade dos dados é diária e o preço utilizado é o de fechamento. Os dias onde não houve pregão, como feriados e fins de semana, foram removidos do universo.

A figura 3.3 apresenta a série temporal do índice Dow Jones de 01/01/1995 até 31/12/2008.



O conjunto de dados é compreendido pelos dias, onde houve negociação entre o dia 01/01/1995 até o dia 31/12/2008, totalizando um montante de 3526 amostras.

As datas de início em 01/01/1995 e fim em 31/12/2008 foram escolhidas com base na série temporal do índice Bovespa.

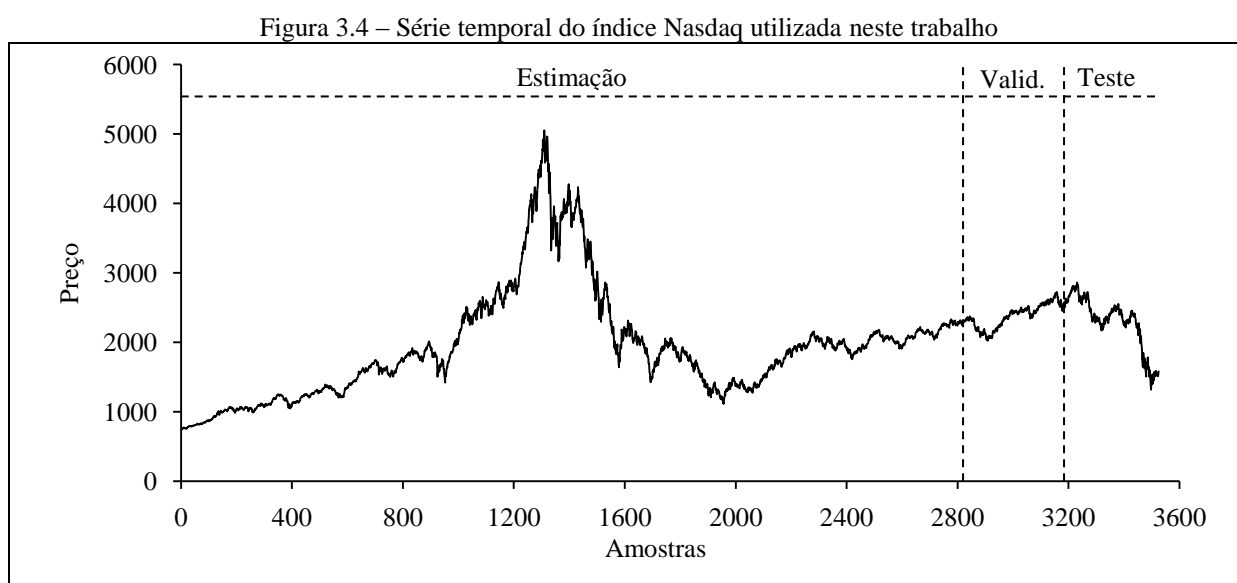
3.2.4. Índice Nasdaq

Segundo a Nasdaq (2009), o índice NASDAQ *Composite* compreende todas as ações comuns listadas na bolsa de valores NASDAQ, domésticas ou internacionais. O índice foi lançado no ano de 1971 com o valor base 100 e, atualmente, inclui mais de 3000 empresas, quase todas da área de tecnologia e biotecnologia. O índice representa a maior bolsa de valores eletrônica do mundo.

Para fazer a coleta dos dados utilizou-se a base de dados do *site* de finanças da Yahoo: www.finance.yahoo.com.

Foram coletados os dados relativos ao Índice NASDAQ *Composite* (^IXIC) a partir da data 01/01/1995 até a data 31/12/2008. A periodicidade dos dados é diária e o preço utilizado é o de fechamento. Os dias onde não houve pregão, como feriados e fins de semana, foram removidos do conjunto de dados.

A figura 3.4 apresenta a série temporal do índice Nasdaq de 01/01/1995 até 31/12/2008.



Desta maneira, o universo é compreendido pelos dias onde houve negociação entre o dia 01/01/1995 até o dia 31/12/2008, totalizando um montante de 3526 amostras.

As datas de início em 01/01/1995 e fim em 31/12/2008 foram escolhidas com base na série temporal do índice Bovespa.

3.3. Modelos de previsão

Os modelos de previsão escolhidos foram:

- Média móvel centrada com índice de sazonalidade;
- Holt-Winters;
- Rede neural MLP; e
- Rede neural RBF.

Essa escolha baseou-se em modelos de previsões já conhecidos do autor (MMC com índice de sazonalidade e Holt-Winters) e modelos comumente utilizados em outros trabalhos de previsão, sendo para séries financeiras ou não (RN-MLP e RN-RBF).

3.4. Índices de desempenho

A fim de comparar os resultados dos modelos de previsão para cada série temporal, foram selecionados três índices normalmente utilizados para análise de desempenho: MSE, MAPE e R^2 .

3.5. Aplicação da previsão

O *software* MATLAB, da MathWorks, foi utilizado para a codificação dos modelos de previsão bem como para a geração dos resultados, gráficos e posteriores cálculos dos índices de desempenho. De uma forma geral, um *script* desenvolvido especialmente para esta pesquisa lia os dados das séries temporais, que já tinham sido coletados em suas respectivas fontes em um passo anterior, e aplicava os modelos de previsão, armazenando os resultados em arquivos. Em seguida, os melhores parâmetros para cada modelo de previsão eram apresentados na tela, bem como os índices de desempenho encontrados para aqueles parâmetros.

As séries temporais foram divididas em três segmentos:

- Estimação, a fim de usar para a estimacão dos modelos: 80% dos dados da amostra;
- Validação, usado para encontrar o menor erro usando o modelo encontrado no segmento de estimacão: 10% dos dados da amostra; e
- Teste, usado para avaliar a capacidade de generalizacão e acuracidade do modelo encontrado: 10% dos dados da amostra.

Essa divisão ocorreu de forma seqüencial, considerando a ordem das observações da série temporal, ou seja, a estimação teve amostras do início da série até 80%, os 10% seguintes foram usados para validação e os 10% finais foram usados para teste. A tabela 3.1 mostra qual foi a distribuição de cada segmento para cada série temporal.

Tabela 3.1. – Distribuição das amostras para cada segmento

Série Temporal	Estimação	Validação	Teste
Mackey-Glass	1 até 800	801 até 900	901 até 1000
Índice Bovespa	1 até 2770	2771 até 3116	3117 até 3462
Índice Dow Jones	1 até 2821	2822 até 3174	3175 até 3526
Índice NASDAQ	1 até 2821	2822 até 3174	3175 até 3526

O *script* desenvolvido em MATLAB estimava o modelo com dados do primeiro segmento, a estimação. Em seguida fazia a previsão para o próximo segmento, o de validação, e gerava os índices de desempenho só para o segmento de validação. Os melhores parâmetros do modelo encontrados usando o segmento de validação eram então utilizados para gerar a previsão do segmento de teste.

Para os modelos de RNs a defasagem a ser usada funcionava como entradas na RN. Assim, quando o parâmetro dizia defasagem de 3 significa dizer que a rede recebia 3 entradas, uma para cada defasagem: entrada 1, observação no $t-1$; entrada 2, observação no $t-2$; e, entrada 3, observação no $t-3$. Onde t é o tempo da iteração atual para a previsão um passo à frente. Apenas as entradas relacionadas à defasagem participaram do conjunto de entradas dos modelos de RNs.

Os resultados então foram então analisados, tabulados e apresentados no capítulo 4.

Capítulo 4

Apresentação e Análise dos Resultados

4.1. Simulações para a saída do sistema Mackey-Glass

A saída do sistema Mackey-Glass foi escolhida para servir de comparativo entre os métodos validados neste trabalho e outros métodos anteriormente estudados por outros autores.

4.1.1. Previsão com MMC com índice de sazonalidade

Para a previsão usando MMC com índice de sazonalidade foram testadas sazonalidades de tamanho 1 até 400, sendo que a melhor sazonalidade encontrada para o período de estimação foi de 1. A figura 4.1 apresenta a previsão da saída do sistema Mackey-Glass usando o modelo MMC com índice de sazonalidade e também apresenta o erro quadrático para o segmento de validação e teste da mesma série.

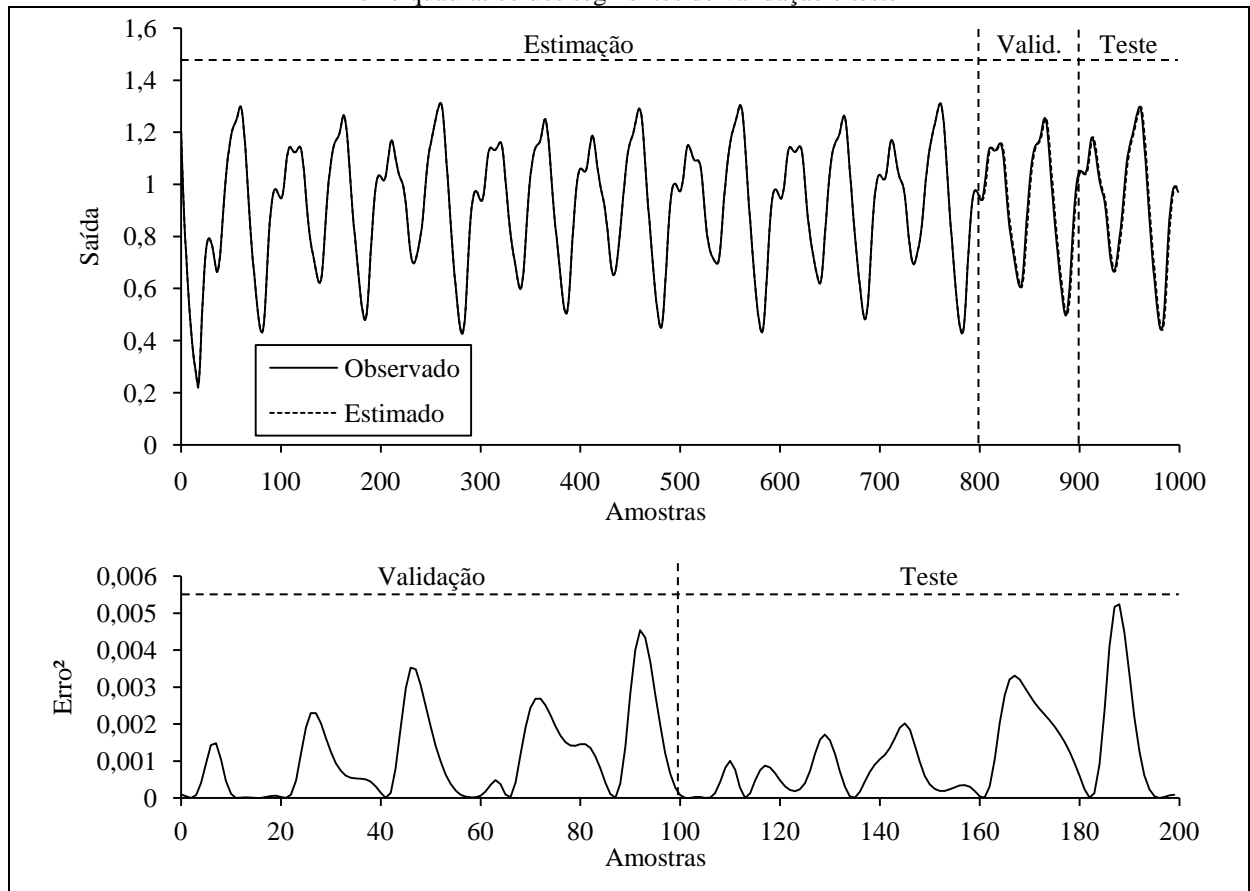
Considerando o MSE como índice de desempenho da fase de validação e buscando o menor valor para este índice, a melhor sazonalidade encontrada foi de 1.

Visto que os valores estimados na fase de estimação foram igualados aos valores observados da mesma fase, os índices de desempenho para esta fase não são considerados.

Na fase de validação, os valores do MSE, MAPE e R^2 foram, respectivamente, $1,0937 \times 10^{-3}$, 3,287300% e 0,99763.

Na fase de teste os valores do MSE, MAPE e R^2 foram, respectivamente, $1,0669 \times 10^{-3}$, 3,311100% e 0,97949.

Figura 4.1 – Previsão da saída do sistema Mackey-Glass usando o modelo MMC com índice de sazonalidade e o erro quadrático dos segmentos de validação e teste

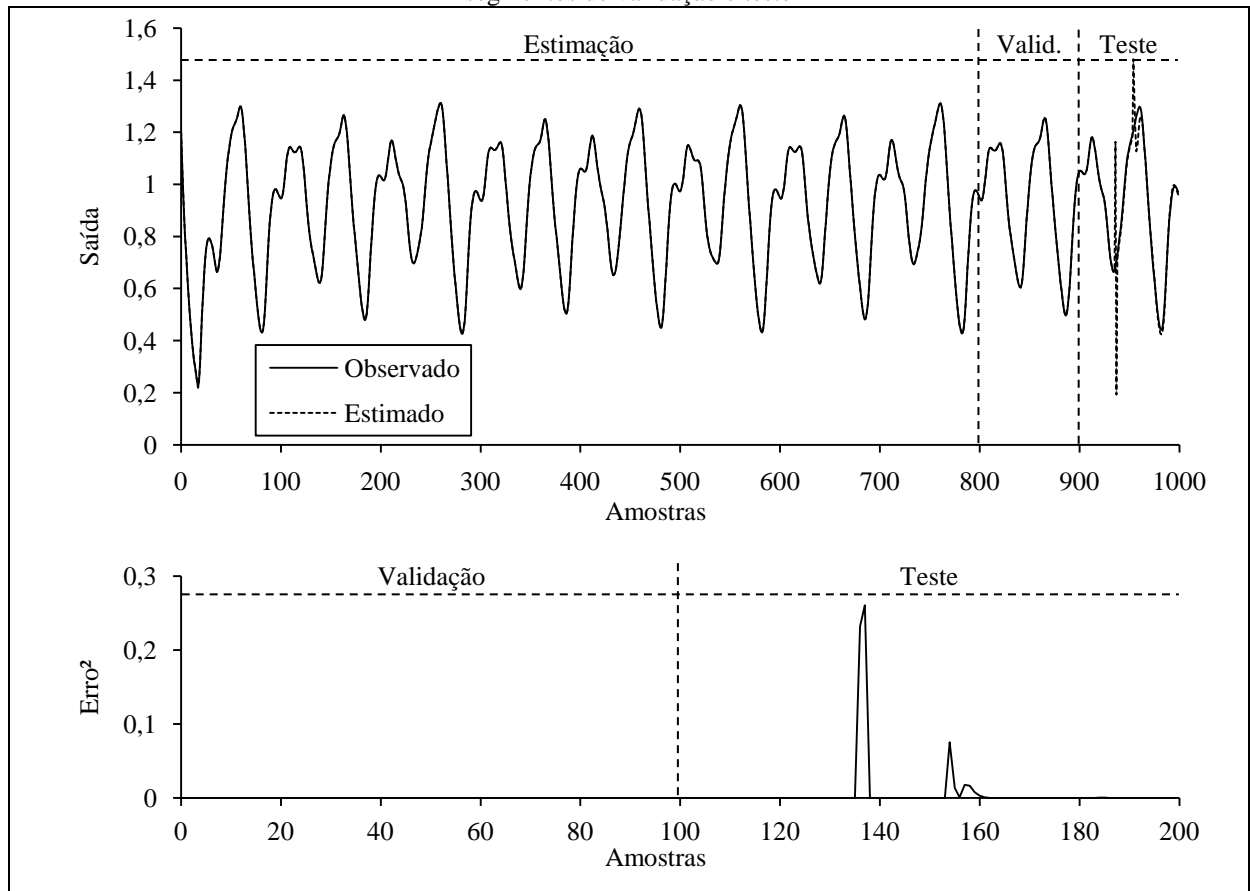


4.1.2. Previsão usando o método Holt-Winters

Para a previsão da saída do sistema Mackey-Glass usando Holt-Winters foram estimadas sazonalidades de tamanho 1 até 400, sendo que para cada sazonalidade estimada os parâmetros α , β e γ foram testados independentemente com valores entre 0 e 1 a um passo de 0,1 a cada etapa. A figura 4.2 apresenta a previsão da saída do sistema Mackey-Glass usando o modelo Holt-Winters e também apresenta o erro quadrático para o segmento de validação e teste da mesma série.

Considerando o MSE como índice de desempenho da fase de validação e buscando o menor valor para este índice, a melhor sazonalidade obtida foi de 234, enquanto os parâmetros α , β e γ encontrados foram, respectivamente, 1, 1 e 0,2.

Figura 4.2 – Previsão da saída do sistema Mackey-Glass usando o modelo Holt-Winters e o erro quadrático dos segmentos de validação e teste



Visto que os valores estimados na fase de estimação foram igualados aos valores observados da mesma fase, os índices de desempenho para esta fase não são considerados.

Na fase de validação os valores do MSE, MAPE e R^2 foram, respectivamente, $8,2605 \times 10^{-7}$, 0,082254% e 0,99999.

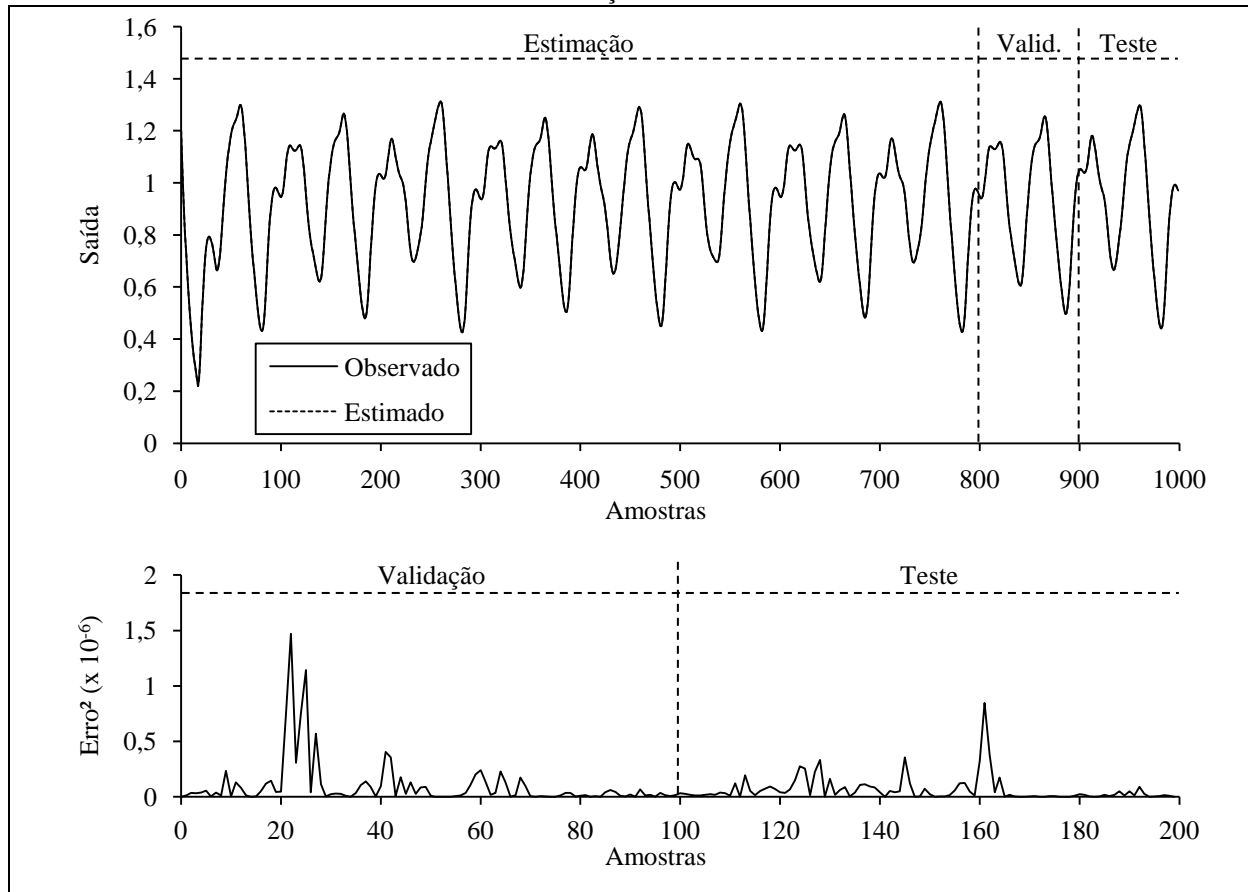
Na fase de teste os valores do MSE, MAPE e R^2 foram, respectivamente, $6,3728 \times 10^{-3}$, 2,956700% e 0,87748.

4.1.3. Previsão usando RNs-MLP

Foram realizadas previsões para a saída do sistema Mackey-Glass, usando RN-MLP, variando a defasagem de 1 até 10 e o número de neurônios na camada oculta de 1 até 25. O método de treinamento foi o Levenberg-Marquardt. A função de ativação da camada oculta foi usada a sigmóide e da camada de saída a linear. Foram 100 épocas de treinamento para cada iteração (defasagem por número de neurônios). A figura 4.3 apresenta o resultado da

melhor previsão encontrada usando RN-MLP para o sistema Mackey-Glass e também apresenta o erro quadrático para o segmento de validação e teste da mesma série.

Figura 4.3 – Previsão de saída do sistema Mackey-Glass usando RN-MLP e o erro quadrático dos segmentos de validação e teste



Foi buscado o menor MSE para a fase de validação da série temporal. Desta forma, o melhor valor para a defasagem foi de 5 e o melhor número de neurônios na camada oculta foi de 15.

Os índices de desempenho encontrados para esses dois parâmetros são apresentados na tabela 4.1.

Tabela 4.1 – Índices de desempenho para a previsão da saída do sistema Mackey-Glass usando RN-MLP

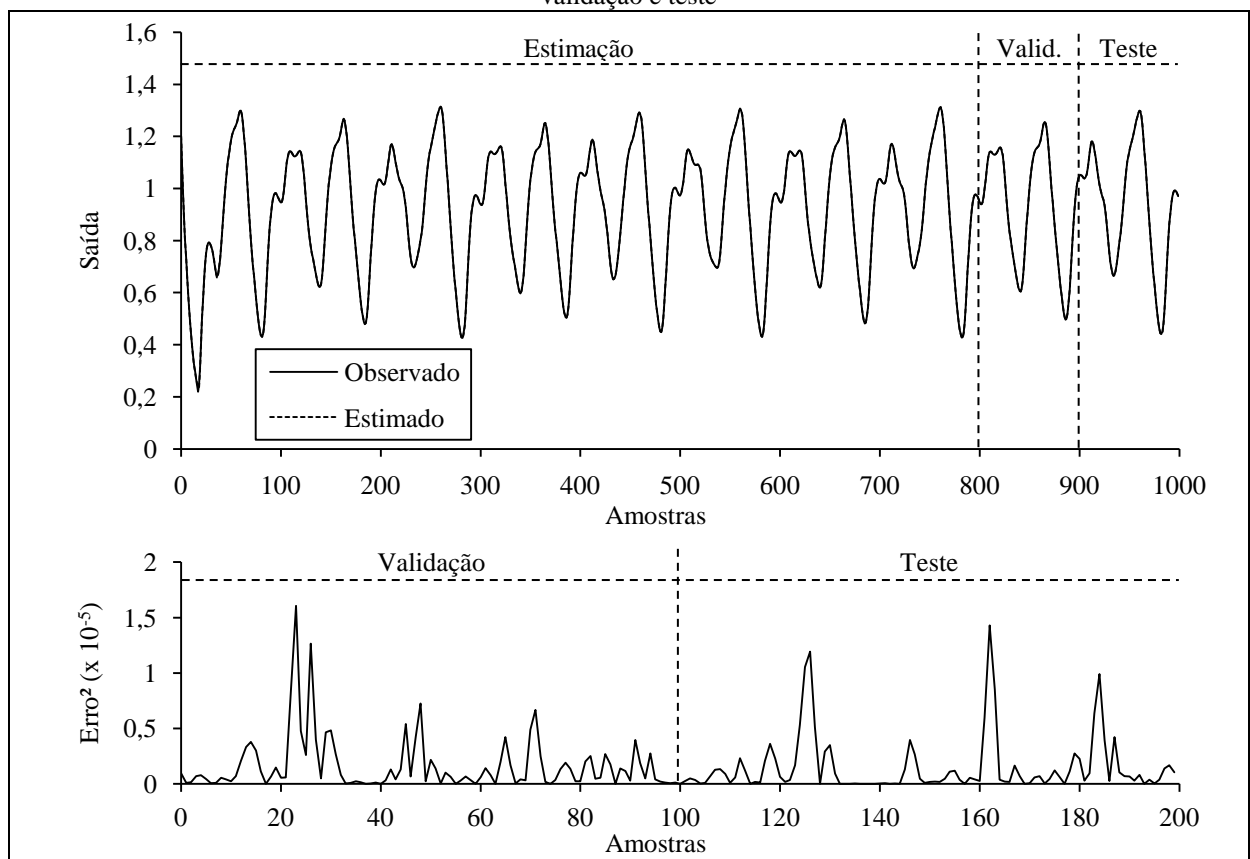
Fase	MSE	MAPE	R^2
Estimação	$1,1801 \times 10^{-7}$	0,026584%	0,99999
Validação	$1,0033 \times 10^{-7}$	0,025015%	0,99999
Teste	$6,4306 \times 10^{-8}$	0,020844%	0,99999

4.1.4. Previsão usando RNs-RBF

Foram realizadas previsões para a saída do sistema Mackey-Glass, usando RN-RBF, variando a defasagem de 1 até 50 e o número de centros de 2 até 50. Os *spreads* das funções de ativação da RN-RBF foram variados entre 1, 0,1, 0,01 e 0,001. Neste caso, a função de base radial adotada foi a Gaussiana. O ajuste dos centros foi realizado pelo método *k-means* e o ajuste dos pesos através do algoritmo de mínimos quadrados em batelada.

A figura 4.4 apresenta o resultado da melhor previsão encontrada usando RN-RBF para a saída do sistema Mackey-Glass e também apresenta o erro quadrático para o segmento de validação e teste da mesma série.

Figura 4.4 – Previsão da saída do sistema Mackey-Glass usando RN-RBF e o erro quadrático dos segmentos de validação e teste



Foi buscado o menor MSE variando-se a defasagem e o número de centros para a fase de validação da série temporal. Desta forma, o melhor valor para a defasagem foi de 24 e o melhor número de centros foi de 21. O *spread* que apresentou o melhor resultado foi 1.

Os índices de desempenho encontrados para esses dois parâmetros são apresentados na tabela 4.2.

Tabela 4.2 – Índices de desempenho para a saída do sistema da saída do sistema Mackey-Glass usando RN-RBF

Fase	MSE	MAPE	R^2
Estimação	$2,8395 \times 10^{-6}$	0,152370%	0,99995
Validação	$1,6646 \times 10^{-6}$	0,116250%	0,99999
Teste	$1,5106 \times 10^{-6}$	0,108780%	0,99997

4.1.5. Comentários sobre os resultados da saída do sistema Mackey-Glass

A tabela 4.3 apresenta, de forma tabulada, os resultados de cada modelo, para cada segmento da saída do sistema.

Tabela 4.3 – Comparativo dos resultados para a saída do sistema Mackey-Glass

Índice de desempenho	MMC com índice de sazonalidade	Holt-Winters	RN-MLP	RN-RBF
Estimação				
MSE			$1,1801 \times 10^{-7}$	$2,8395 \times 10^{-6}$
MAPE			0,026584%	0,152370%
R^2			0,99999	0,99995
Validação				
MSE	$1,0937 \times 10^{-3}$	$8,2605 \times 10^{-7}$	$1,0033 \times 10^{-7}$	$1,6646 \times 10^{-6}$
MAPE	3,287300%	0,082254%	0,025015%	0,116250%
R^2	0,99763	0,99999	0,99999	0,99999
Teste				
MSE	$1,0669 \times 10^{-3}$	$6,3728 \times 10^{-3}$	$6,4306 \times 10^{-8}$	$1,5106 \times 10^{-6}$
MAPE	3,311100%	2,956700%	0,020844%	0,108780%
R^2	0,97949	0,87748	0,99999	0,99997

Conforme citado anteriormente, a saída do sistema Mackey-Glass foi selecionada por servir de *benchmark* em vários estudos já realizados, e neste trabalho tem o sentido de posicionar os métodos selecionados de acordo com outros trabalhos realizados.

Dentre os modelos apresentados o que teve o pior resultado geral foi o modelo linear de médias móveis com índice de sazonalidade. É possível perceber que quase não houve diferença entre a fase de validação e a fase teste, pois os valores dos três índices analisados ficaram muito próximos. Esta pouca variação se deve ao fato de que a melhor sazonalidade encontrada para o modelo de MMC foi de 1, fazendo com que o valor estimado fosse igual ao valor observado no $t-1$. No que diz respeito ao modelo linear Holt-Winters, durante a fase de validação, que é onde o processo tenta encontrar o menor MSE, é possível observar que ficou em segundo lugar, com resultado inferior apenas para a RN-MLP, por uma estreita margem, e superando até mesmo a RN-RBF. No entanto, a fase de teste acabou ficando tão ruim quanto

o outro modelo linear estudado, o de médias móveis. É possível entender desta forma que o modelo Holt-Winters, para esta série, foi eficiente em estimar os parâmetros para um segmento conhecido da série, porém não se saiu bem quando o mesmo modelo foi aplicado em dados não conhecidos durante a estimação.

Ainda sobre o modelo Holt-Winters, é possível perceber através da figura 4.2 que quando o modelo foi aplicado na fase de teste os valores estimados ficaram longe dos valores observados em alguns momentos, enquanto em outros momentos os valores estimados acabaram ficando muito próximos dos valores observados. Isso explica porque o MSE do modelo Holt-Winters é maior que do modelo MMC com índice de sazonalidade, enquanto que o MAPE do modelo MMC com índice de sazonalidade é maior que o do modelo Holt-Winters.

No que dizem respeito aos modelos não-lineares, ambos mantiveram o mesmo desempenho durante a fase de estimação, validação e teste, mostrando a capacidade de generalização desses modelos na saída do sistema Mackey-Glass. Por fim, o melhor modelo encontrado foi o da RN-MLP, que além de ser melhor que a RN-RBF nas três fases, ainda obteve um resultado melhor na fase de teste do que na fase de validação, onde o processo procurava pelo menor MSE.

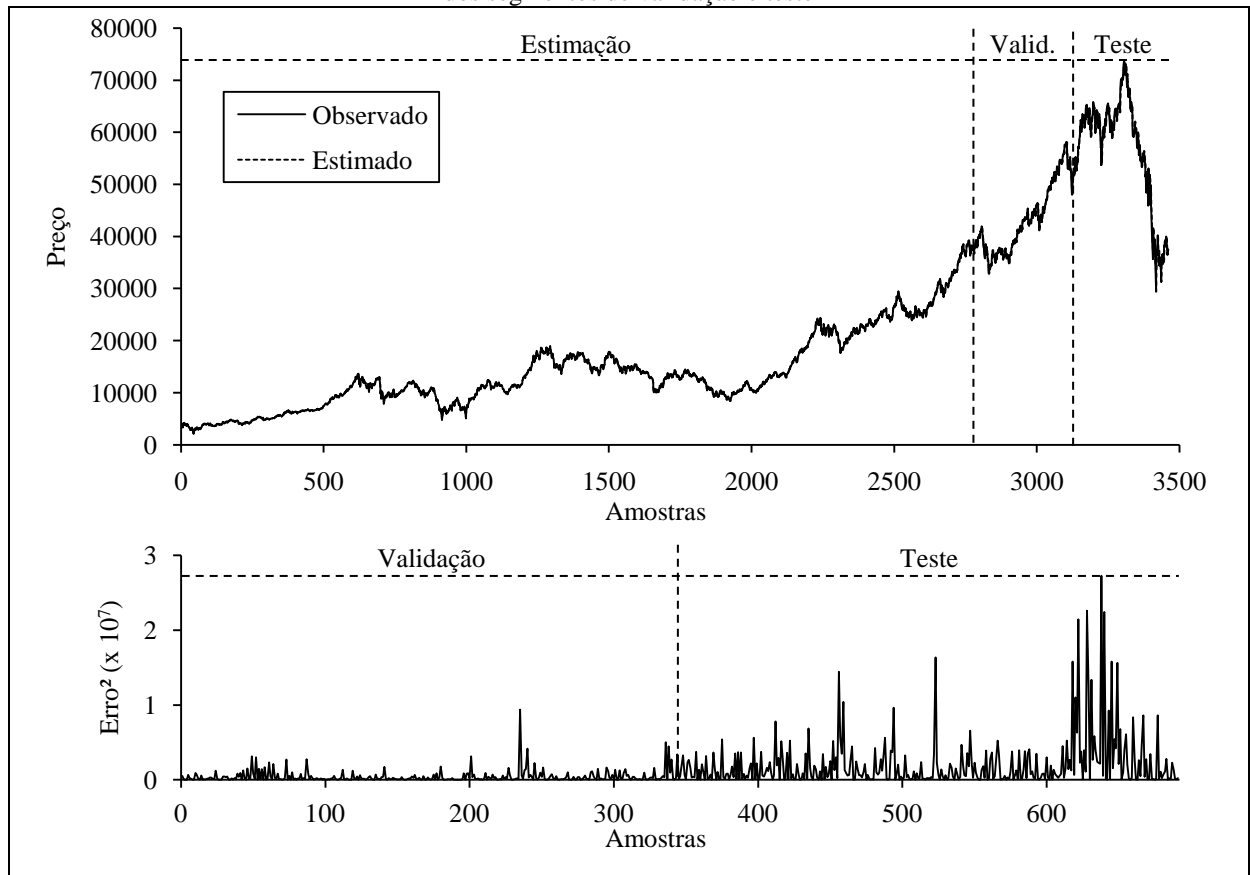
4.2. Simulações para a série Ibovespa

4.2.1. Previsão com média móvel centrada com índice de sazonalidade

Para a previsão usando média móvel centrada com índice de sazonalidade foram testadas sazonalidades de tamanho 1 até 400, sendo que a melhor sazonalidade encontrada para o período de estimação foi de 1. A figura 4.5 apresenta a previsão da série temporal do índice Bovespa usando MMC com índice de sazonalidade e também apresenta o erro quadrático para o segmento de validação e teste da mesma série.

Considerando o MSE como índice de desempenho da fase de validação e buscando o menor valor para este índice, a melhor sazonalidade encontrada foi de 1.

Figura 4.5 – Previsão da série Ibovespa usando o modelo MMC com índice de sazonalidade e o erro quadrático dos segmentos de validação e teste



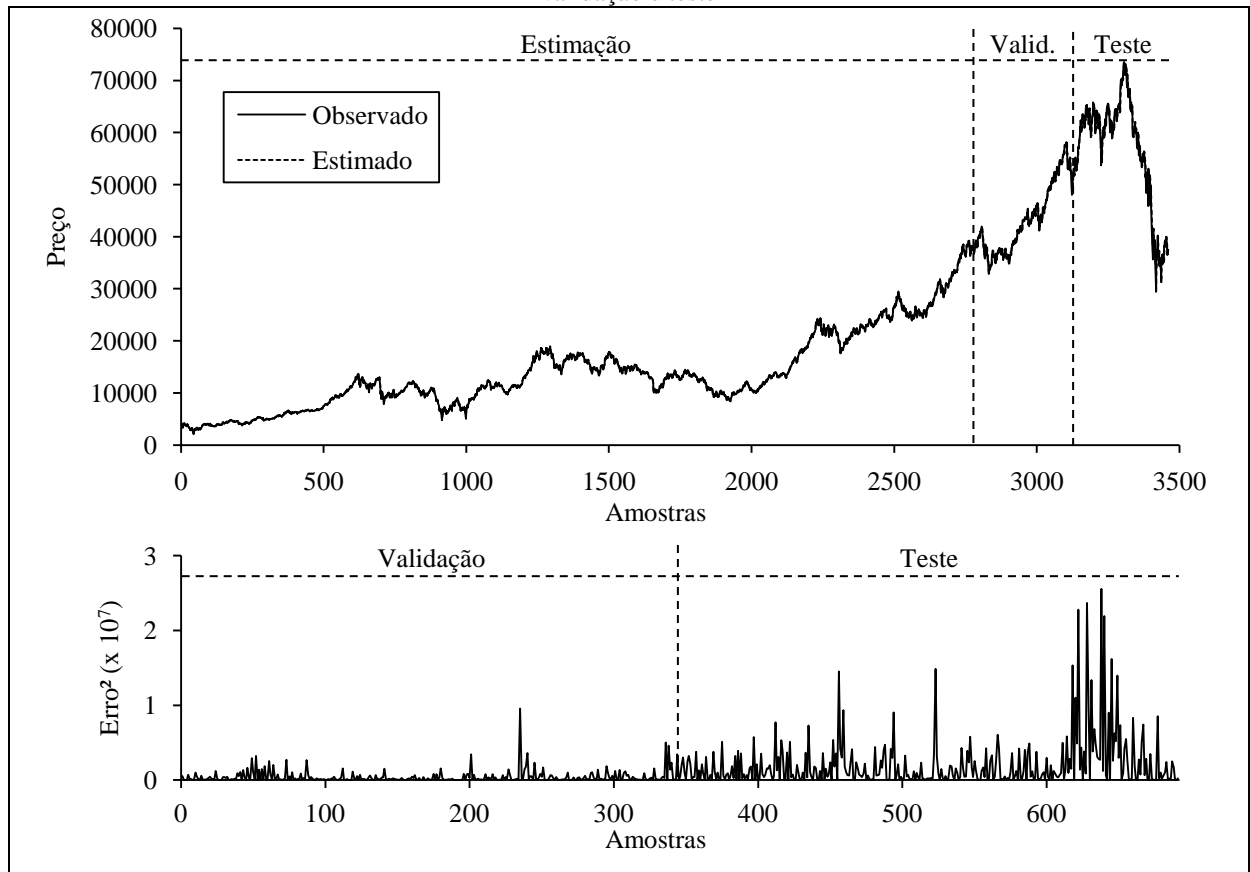
Dado que na fase de estimação os valores foram igualados aos valores observados da mesma fase, os índices de desempenho para esta fase não são considerados. Observa-se, no entanto, que os valores do MSE, MAPE e R^2 encontrados para a fase de validação foram, respectivamente, 427324,18, 1,1586% e 0,99994. Para a fase de teste os valores do MSE, MAPE e R^2 foram, respectivamente, 2057891,19, 2,0992% e 0,98119.

4.2.2. Previsão usando o método Holt-Winters

Para a previsão da série Ibovespa usando Holt-Winters foram estimadas sazonalidades de tamanho 1 até 400, sendo que para cada sazonalidade estimada os parâmetros α , β e γ foram testados independentemente com valores entre 0 e 1 a um passo de 0,1 a cada etapa. A figura 4.6 apresenta a previsão da série temporal do índice Bovespa usando o modelo Holt-Winters e também apresenta o erro quadrático para o segmento de validação e teste da mesma série.

Considerando o MSE como índice de desempenho da fase de validação e buscando o menor valor para este índice, a melhor sazonalidade encontrada foi de 1, enquanto os parâmetros α , β e γ encontrados foram, respectivamente, 0,8, 0 e 0,6.

Figura 4.6 – Previsão da série Ibovespa usando o modelo Holt-Winters e o erro quadrático dos segmentos de validação e teste

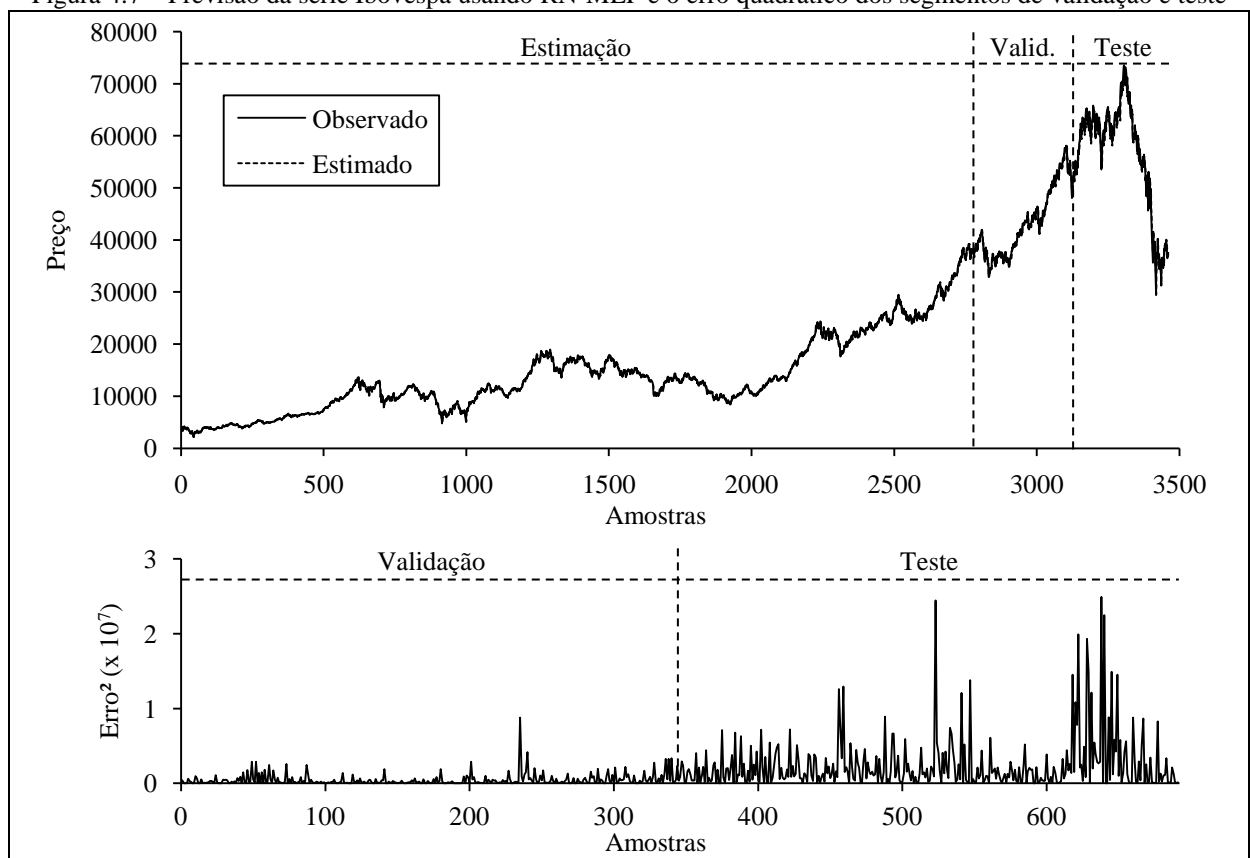


Uma vez que os valores estimados na fase de estimação foram igualados aos valores observados da mesma fase, os índices de desempenho para esta fase não são considerados. Já para a fase de validação os valores do MSE, MAPE e R^2 foram, respectivamente, 423746,24, 1,1480% e 0,99994. Enquanto para a fase de teste os valores do MSE, MAPE e R^2 foram, respectivamente, 2064644,51, 2,1094% e 0,98113.

4.2.3. Previsão usando RNs-MLP

Foram realizadas previsões para a série Ibovespa, usando RN-MLP, variando a defasagem de 1 até 10 e o número de neurônios na camada oculta de 1 até 25. O método de treinamento foi o Levenberg-Marquardt. A função de ativação da camada oculta foi usada a sigmóide e da camada de saída a linear. Foram 100 épocas de treinamento para cada iteração (defasagem por número de neurônios). A figura 4.7 apresenta a previsão da série temporal do índice Bovespa usando RN-MLP e também apresenta o erro quadrático para o segmento de validação e teste da mesma série.

Figura 4.7 – Previsão da série Ibovespa usando RN-MLP e o erro quadrático dos segmentos de validação e teste



Foi buscado o menor MSE para a fase de validação da série temporal. Desta forma, o melhor valor para a defasagem foi de 8 e o melhor número de neurônios na camada oculta foi de 1. Os índices de desempenho encontrados para esses dois parâmetros são apresentados na tabela 4.4.

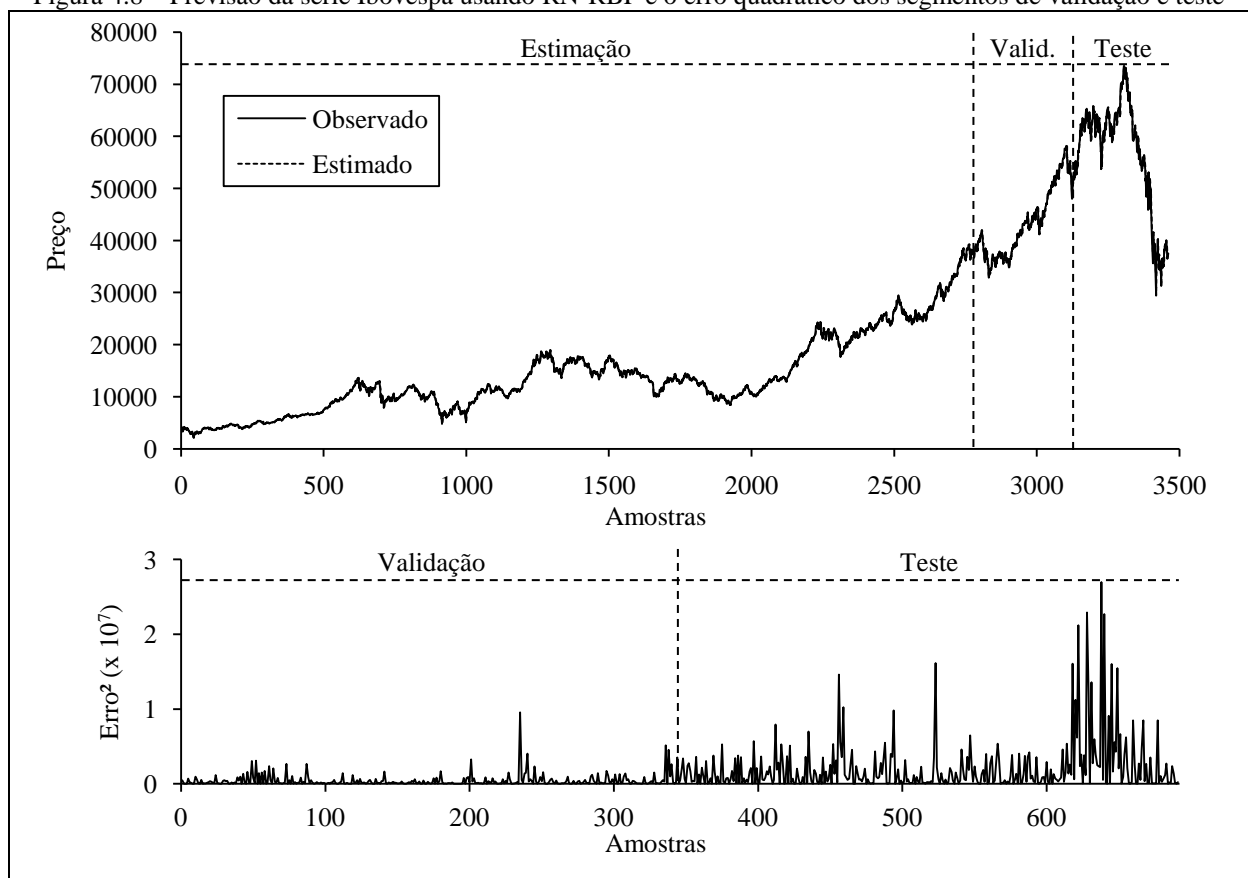
Tabela 4.4 – Índices de desempenho para a previsão da série Ibovespa usando RN-MLP

Fase	MSE	MAPE	R^2
Estimação	83178,33	1,6818%	0,99850
Validação	455112,73	1,2059%	0,99994
Teste	2406968,44	2,3217%	0,97800

4.2.4. Previsão usando RNs-RBF

Foram realizadas previsões para a série Ibovespa, usando RN-RBF, variando a defasagem de 1 até 50 e o número de centros de 2 até 50. Os *spreads* das funções da ativação da RN-RBF foram variados entre 1, 0,1, 0,01 e 0,001. A função de base radial usada foi a Gaussiana. O ajuste dos centros foi realizado pelo método *k-means* e o ajuste dos pesos através do algoritmo de mínimos quadrados. A figura 4.8 apresenta a previsão da série temporal do índice Bovespa usando RN-RBF e também apresenta o erro quadrático para o segmento de validação e teste da mesma série.

Figura 4.8 – Previsão da série Ibovespa usando RN-RBF e o erro quadrático dos segmentos de validação e teste



Foi buscado o menor MSE para a fase de validação da série temporal. Desta forma, o melhor valor para a defasagem foi de 1 e o melhor número de centros foi de 2. O *spread* que apresentou o melhor resultado foi 0,001.

Os índices de desempenho encontrados para esses dois parâmetros são apresentados na tabela 4.5.

Tabela 4.5 – Índices de desempenho para a previsão da série Ibovespa usando RN-RBF

Fase	MSE	MAPE	R^2
Estimação	83908,26	1,6941%	0,99849
Validação	425509,17	1,1558%	0,99994
Teste	2060447,57	2,0981%	0,98117

4.2.5. Comentários sobre os resultados da série Ibovespa

A comparação entre os índices de desempenho de cada modelo estudado para a série Ibovespa é apresentada na tabela 4.6.

Tabela 4.6 – Comparativo dos resultados para a série Ibovespa

Índice de desempenho	MMC com índice de saz.	Holt-Winters	RN-MLP	RN-RBF
Estimação				
MSE			83178,33	83908,26
MAPE			1,6818%	1,6941%
R^2			0,99850	0,99849
Validação				
MSE	427324,18	423746,24	455112,73	425509,17
MAPE	1,1586%	1,1480%	1,2059%	1,1558%
R^2	0,99994	0,99994	0,99994	0,99994
Teste				
MSE	2057891,19	2064644,51	2406968,44	2060447,57
MAPE	2,0992%	2,1094%	2,3217%	2,0981%
R^2	0,98119	0,98113	0,978	0,98117

Para a fase de estimação, é válido fazer o comparativo apenas entre os modelos não-lineares, visto que os modelos lineares tiveram seus valores estimados igualados aos valores observados.

A RN-MLP obteve uma pequena vantagem sobre a RN-RBF na fase de estimação, em todos os índices de desempenho.

Na fase de validação o modelo Holt-Winters foi o que obteve o melhor resultado nos índices MSE e MAPE. Todos os modelos obtiveram o mesmo índice R^2 . É interessante perceber que nesta fase o modelo RN-MLP foi o que obteve o pior resultado, ficando com uma diferença de mais de 4% para os índices MSE e MAPE dos outros modelos, entre eles os lineares.

Para a fase de teste, o modelo MMC com índice de sazonalidade apresentou o melhor MSE e o melhor R^2 . O modelo RN-RBF apresentou o melhor MAPE. É interessante observar que a diferença entre o MAPE da RN-RBF e do MMC com índice de sazonalidade foi

extremamente baixo. Para esta fase a RN-MLP apresentou o pior resultado para todos os índices de desempenho, sendo quase 10% a diferença para o terceiro pior índice MAPE e 15% para o terceiro pior índice MSE.

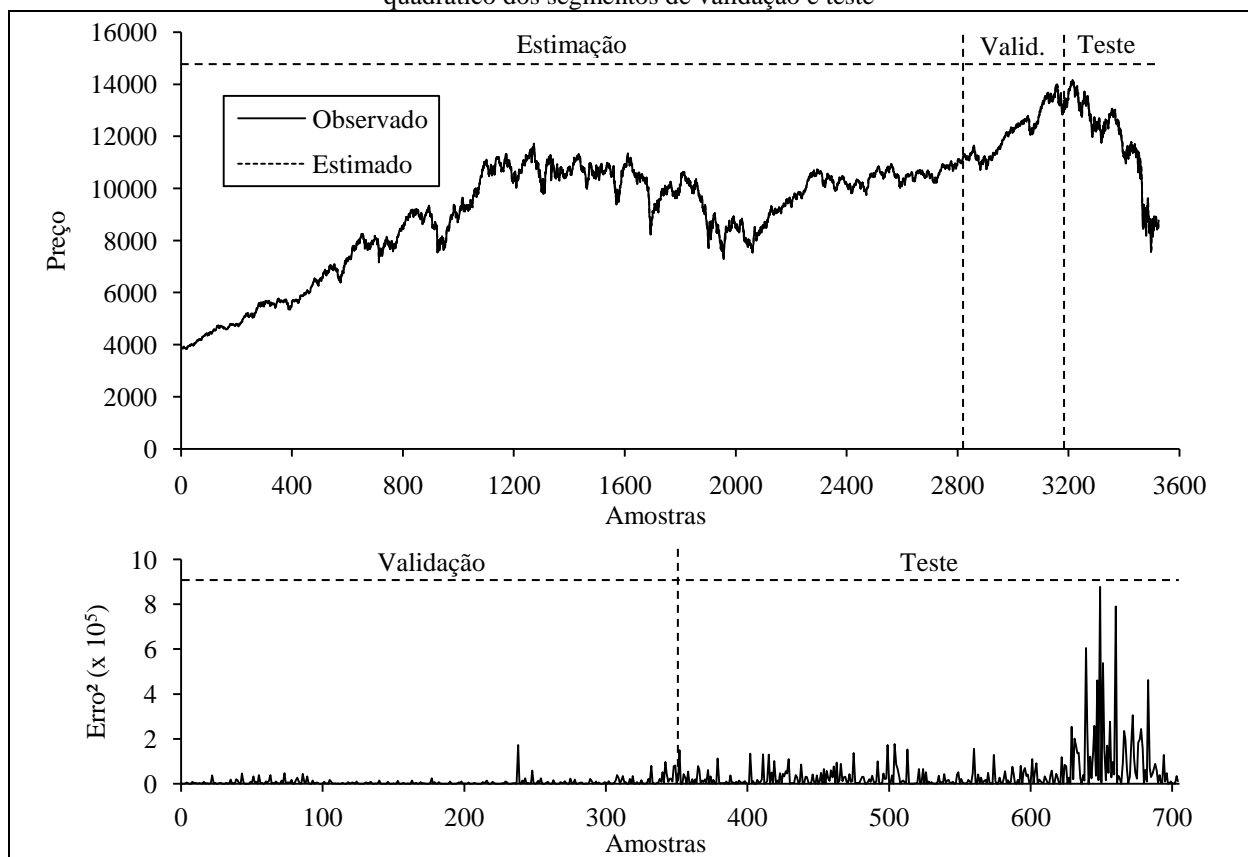
É possível concluir que, mesmo a RN-MLP tendo estimado melhor o modelo durante a fase de estimação, este modelo provou não ser tão eficiente na fase de validação onde o melhor modelo foi o Holt-Winters. No entanto, a fase de teste com dados fora da amostra, apresentou o modelo MMC com índice de sazonalidade como melhor resultado.

4.3. Simulações para a série Dow Jones

4.3.1. Previsão com média móvel centrada com índice de sazonalidade

Para a previsão usando MMC com índice de sazonalidade foram testadas sazonalidades de tamanho 1 até 400, sendo que a melhor sazonalidade encontrada para o período de estimação foi de 1.

Figura 4.9 – Previsão da série Dow Jones usando o modelo MMC com índice de sazonalidade e o erro quadrático dos segmentos de validação e teste



A figura 4.9 apresenta a previsão da série temporal do índice Dow Jones usando o modelo MMC com índice de sazonalidade e também apresenta o erro quadrático para o segmento de validação e teste da mesma série.

Considerando o MSE como índice de desempenho da fase de validação e buscando o menor valor para este índice, a melhor sazonalidade encontrada foi de 1.

Os índices de desempenho para a fase de estimação não foram considerados visto que os valores estimados nesta fase foram igualados aos valores observados. Durante a fase de validação os valores do MSE, MAPE e R^2 encontrados foram, respectivamente, 7476,20, 0,50123% e 0,99994. E Para a fase de teste os valores do MSE, MAPE e R^2 obtidos foram, respectivamente, 44932,00, 1,40340% e 0,98362.

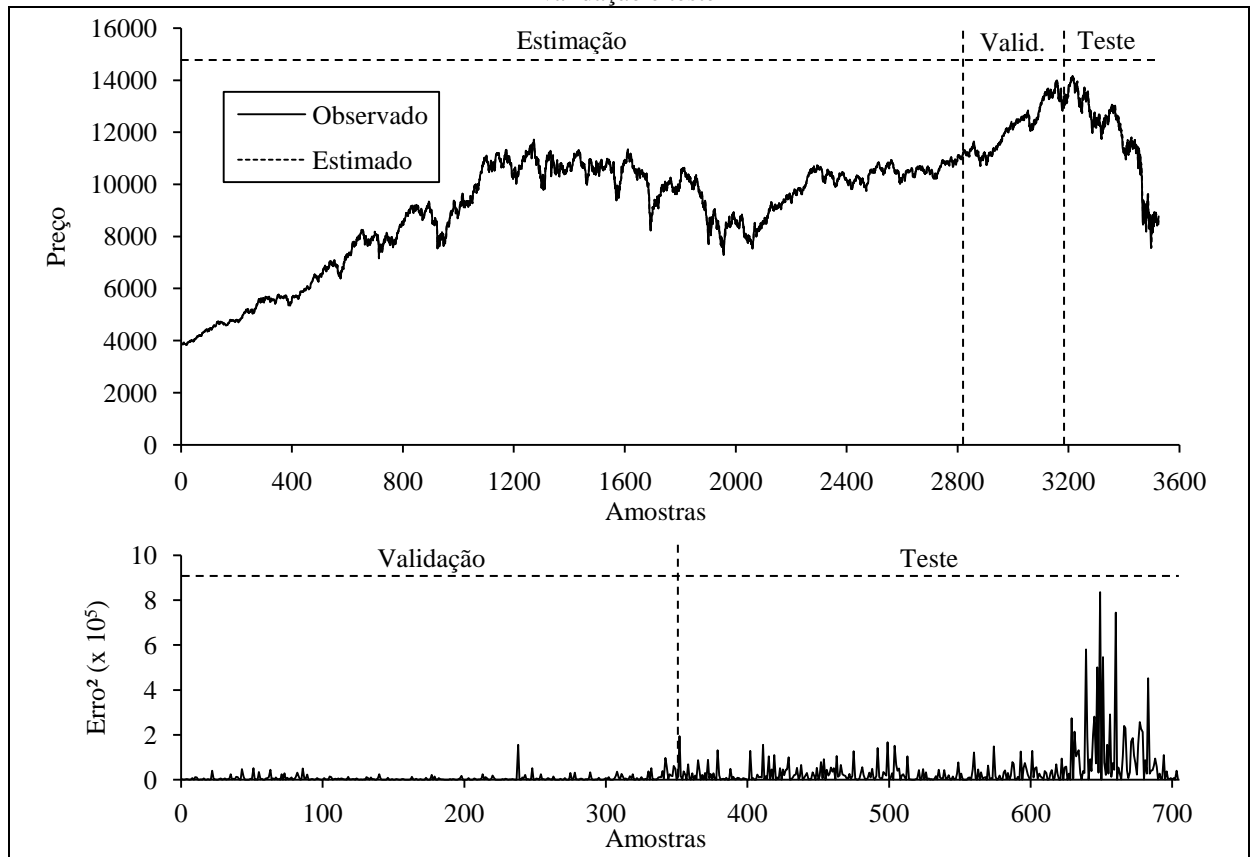
4.3.2. Previsão usando o método Holt-Winters

Para a previsão da série Dow Jones usando Holt-Winters foram estimadas sazonalidades de tamanho 1 até 400, sendo que para cada sazonalidade estimada os parâmetros α , β e γ foram testados independentemente com valores entre 0 e 1 a um passo de 0,1 a cada etapa. A figura 4.10 apresenta a previsão da série temporal do índice Dow Jones usando o modelo Holt-Winters e também apresenta o erro quadrático para o segmento de validação e teste da mesma série.

Considerando o MSE como índice de desempenho da fase de validação e buscando o menor valor para este índice, a melhor sazonalidade encontrada foi de 20, enquanto os parâmetros α , β e γ encontrados foram, respectivamente, 0,9, 0 e 0,3.

Visto que os valores estimados na fase de estimação foram igualados aos valores observados da mesma fase, os índices de desempenho para esta fase não são considerados. Para a fase de validação os valores do MSE, MAPE e R^2 foram, respectivamente, 7334,74, 0,50302% e 0,99994. Na fase de teste os valores do MSE, MAPE e R^2 foram, respectivamente, 44779,36, 1,41650% e 0,98367.

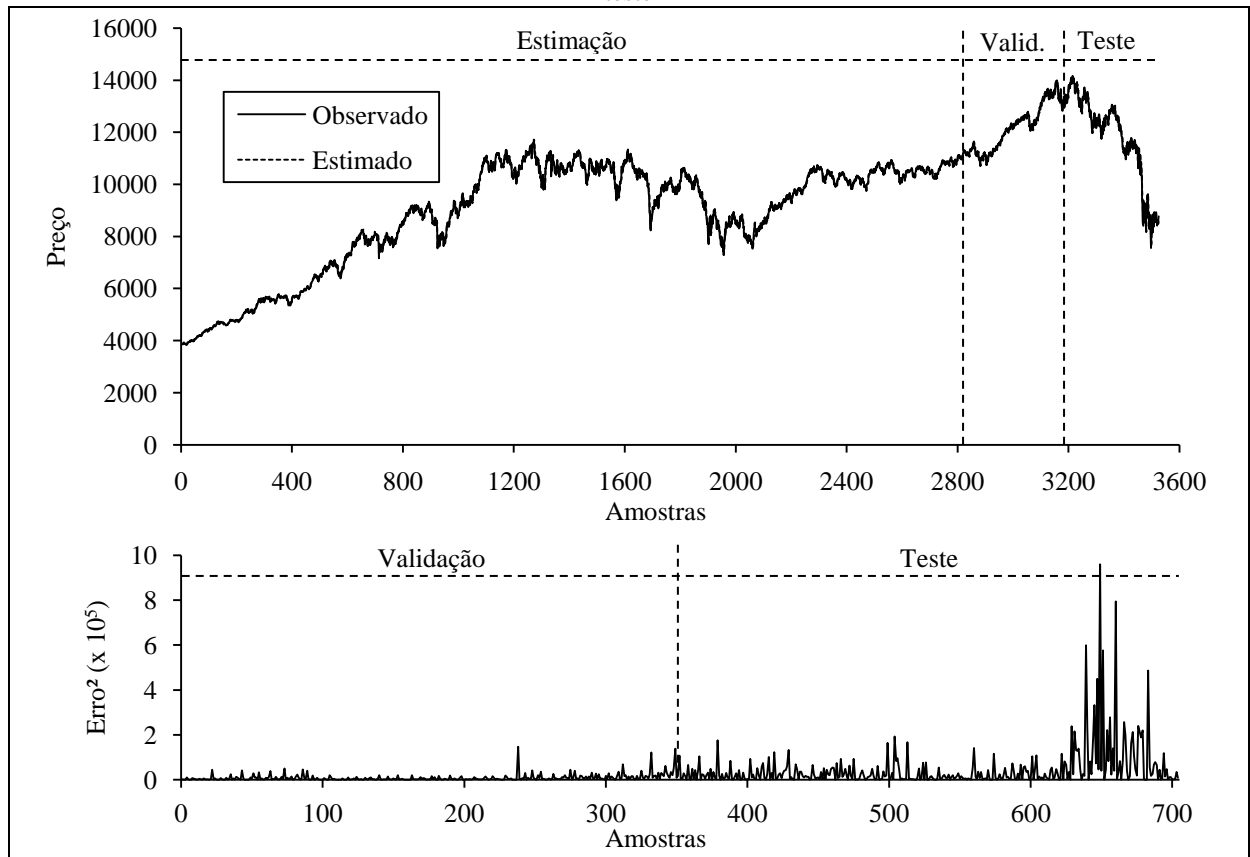
Figura 4.10 – Previsão da série Dow Jones usando o modelo Holt-Winters e o erro quadrático dos segmentos de validação e teste



4.3.3. Previsão usando RNs-MLP

Foram realizadas previsões para a série Dow Jones, usando RN-MLP, variando a defasagem de 1 até 10 e o número de neurônios na camada oculta de 1 até 25. O método de treinamento foi o Levenberg-Marquardt. A função de ativação da camada oculta foi usada a sigmóide e da camada de saída a linear. Foram 100 épocas de treinamento para cada iteração (defasagem por número de neurônios). A figura 4.11 apresenta a previsão da série temporal do índice Dow Jones usando RN-MLP e também apresenta o erro quadrático para o segmento de validação e teste da mesma série.

Figura 4.11 – Previsão da série Dow Jones usando RN-MLP e o erro quadrático dos segmentos de validação e teste



Foi buscado o menor MSE para a fase de validação da série temporal. Desta forma, o melhor valor para a defasagem foi de 5 e o melhor número de neurônios na camada oculta foi de 4.

Os índices de desempenho encontrados para esses dois parâmetros são apresentados na tabela 4.7.

Tabela 4.7 – Índices de desempenho para a previsão da série Dow Jones usando RN-MLP

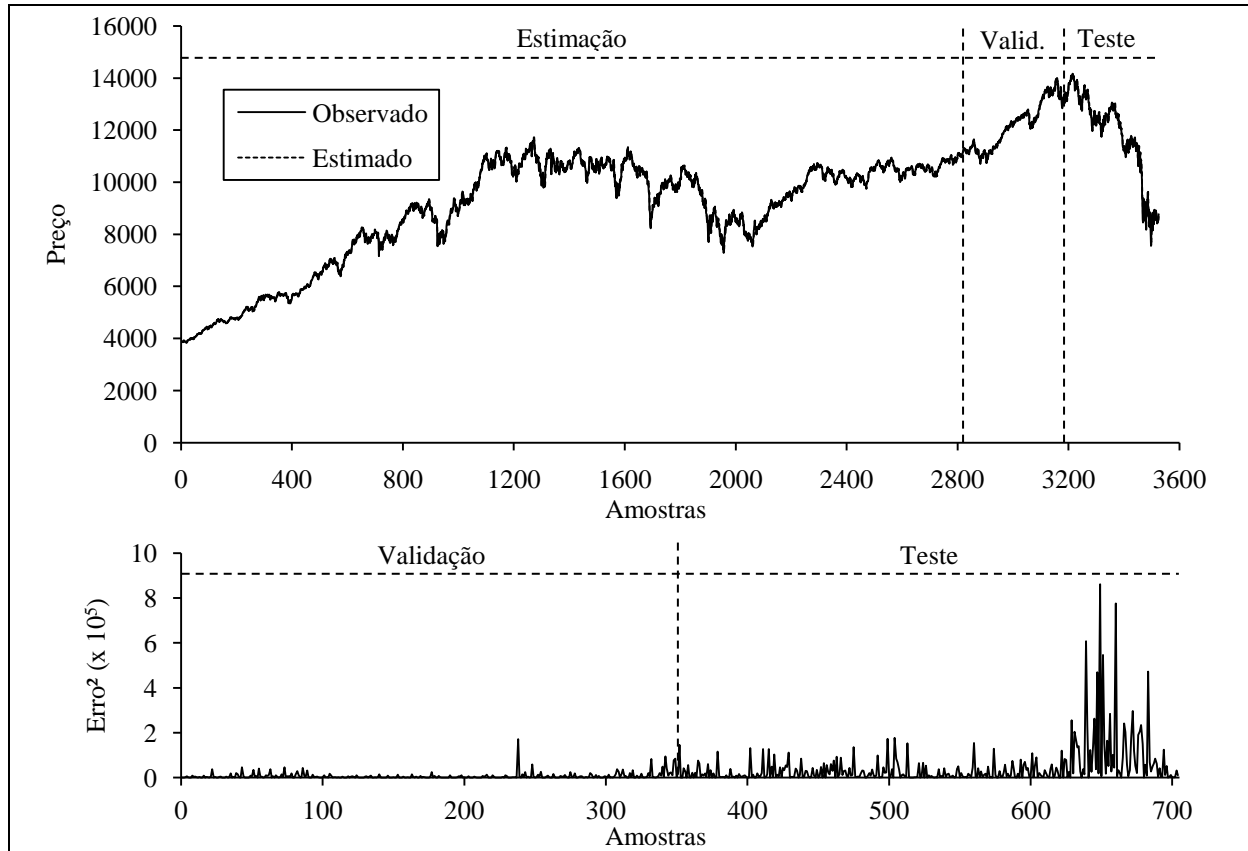
Fase	MSE	MAPE	R^2
Estimação	9388,04	0,77159%	0,99781
Validação	9390,53	0,60935%	0,99992
Teste	46305,46	1,44310%	0,98312

4.3.4. Previsão usando RNs-RBF

Foram realizadas previsões para a série Dow Jones, usando RN-RBF, variando a defasagem de 1 até 50 e o número de centros de 2 até 50. Os *spreads* das funções de ativação da RN-RBF foram variados entre 1, 0,1, 0,01 e 0,001. A função de base radial usada foi a Gaussiana. O ajuste dos centros foi realizado pelo método *k-means* e o ajuste dos pesos

através do algoritmo de mínimos quadrados. A figura 4.12 apresenta a previsão da série temporal do índice Dow Jones usando RN-RBF e também apresenta o erro quadrático para o segmento de validação e teste da mesma série.

Figura 4.12 – Previsão da série Dow Jones usando RN-RBF e o erro quadrático dos segmentos de validação e teste



Foi buscado o menor MSE para a fase de validação da série temporal. Desta forma, o melhor valor para a defasagem foi de 1 e o melhor número de centros foi de 8. O *spread* que apresentou o melhor resultado foi 0,001.

Os índices de desempenho encontrados para esses dois parâmetros são apresentados na tabela 4.8.

Tabela 4.8 – Índices de desempenho para a previsão da série Dow Jones usando RN-RBF

Fase	MSE	MAPE	R^2
Estimação	9543,65	0,78137%	0,99777
Validação	7468,81	0,50233%	0,99994
Teste	44901,99	1,40390%	0,98363

4.3.5. Comentários sobre os resultados da série Dow Jones

A tabela 4.9 apresenta os índices de desempenho de cada modelo estudado para a série Dow Jones.

Tabela 4.9 – Comparativo dos resultados para a série Dow Jones

Índice de desempenho	MMC com índice de saz.	Holt-Winters	RN-MLP	RN-RBF
Estimação				
MSE			9388,04	9543,65
MAPE			0,77159%	0,78137%
R^2			0,99781	0,99777
Validação				
MSE	7476,20	7334,74	9390,53	7468,81
MAPE	0,50123%	0,50302%	0,60935%	0,50233%
R^2	0,99994	0,99994	0,99992	0,99994
Teste				
MSE	44932,00	44779,36	46305,46	44901,99
MAPE	1,40340%	1,4165%	1,4431%	1,40390%
R^2	0,98362	0,98367	0,98312	0,98363

Na fase de estimação da série Dow Jones o método RN-MLP obteve um melhor resultado do que a RN-RBF, porém a diferença foi pequena.

O modelo Holt-Winters obteve o melhor valor para o MSE na fase de validação, enquanto o modelo de MMC com índice de sazonalidade apresentou o melhor MAPE. É interessante ressaltar que a diferença entre os modelos Holt-Winters, MMC com índice de sazonalidade e RN-RBF encontrada para o MAPE e o MSE foi de menos de 3%, enquanto o modelo RN-MLP ficou longe dos outros modelos apresentando uma diferença em torno de 20%. No que diz respeito ao índice R^2 apenas a RN-MLP ficou um pouco abaixo dos demais, não deixando de estar muito próximo dos outros, que empataram.

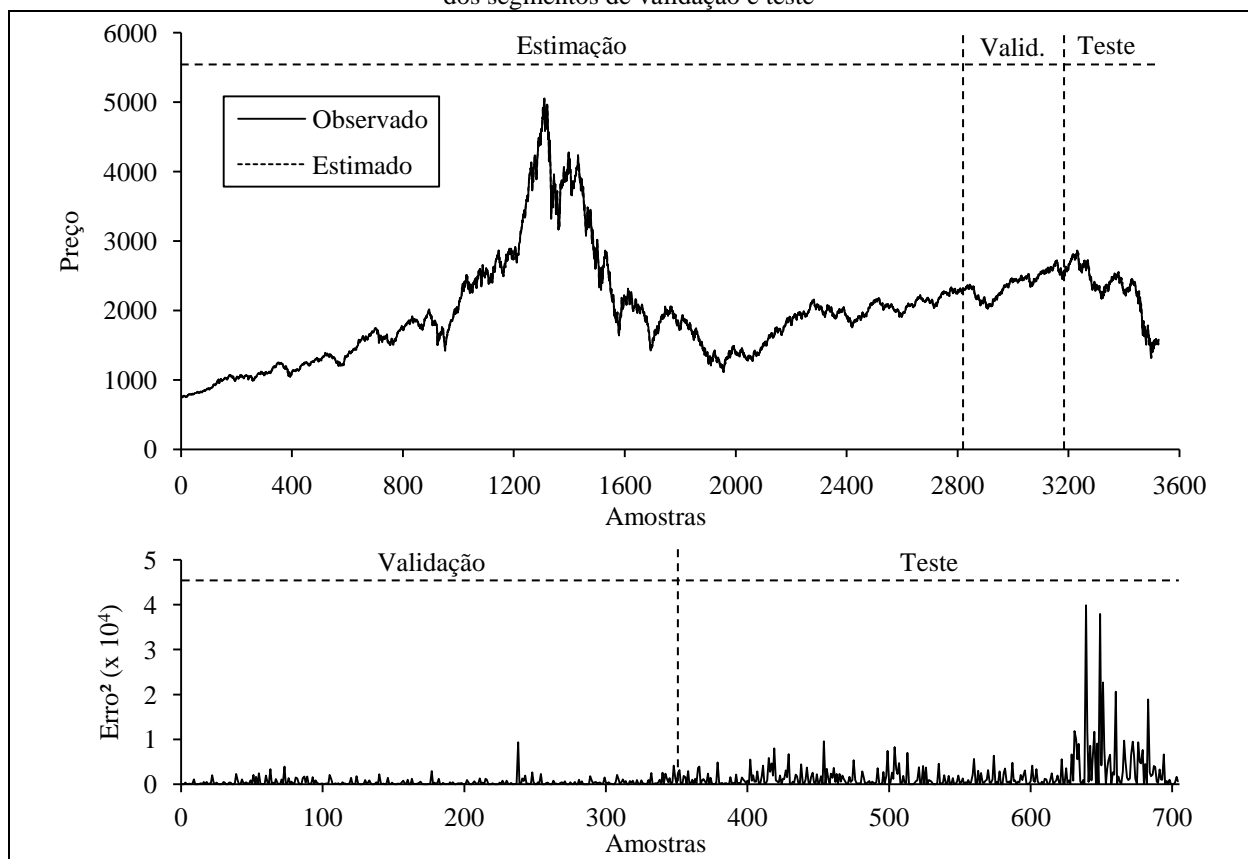
Na fase de teste, os modelos ficaram mais próximos uns dos outros, fazendo com que a diferença entre os maiores e os menores valores encontrados para o MSE, MAPE e R^2 ficasse em menos de 1%, com exceção da RN-MLP que apresentou diferença de 3% para o melhor colocado. De qualquer forma, o modelo MMC com índice de sazonalidade obteve melhor MAPE, enquanto o modelo Holt-Winters obteve o melhor MSE e o melhor R^2 .

4.4. Simulações para a série Nasdaq

4.4.1. Previsão com média móvel centrada com índice de sazonalidade

Para a previsão usando média móvel centrada com índice de sazonalidade foram testadas sazonalidades de tamanho 1 até 400, sendo que a melhor sazonalidade encontrada para o período de estimação foi de 1. A figura 4.13 apresenta a previsão da série temporal do índice Nasdaq usando o modelo MMC com índice de sazonalidade e também apresenta o erro quadrático para o segmento de validação e teste da mesma série.

Figura 4.13 – Previsão da série Nasdaq usando o modelo MMC com índice de sazonalidade e o erro quadrático dos segmentos de validação e teste



Considerando o MSE como índice de desempenho da fase de validação e buscando o menor valor para este índice, a melhor sazonalidade encontrada foi de 1.

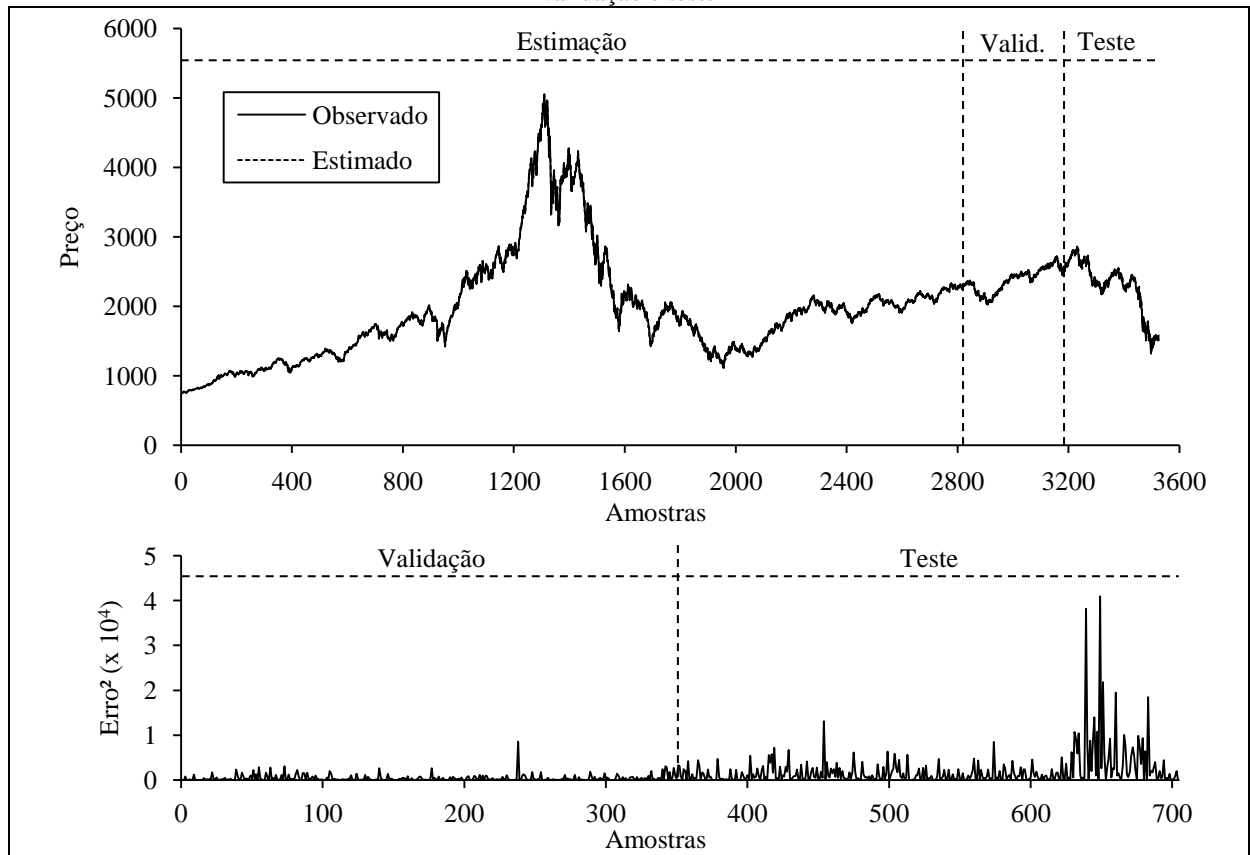
Durante a fase de estimação os valores estimados foram igualados aos valores observados desta mesma fase, por isso os índices de desempenho para esta fase não são considerados. Na fase de validação os valores do MSE, MAPE e R^2 obtidos foram,

respectivamente, 458,50, 0,68805% e 0,99993. Enquanto na fase de teste os valores do MSE, MAPE e R^2 foram, respectivamente, 2054,64, 1,60710% e 0,98513.

4.4.2. Previsão usando o método Holt-Winters

Para a previsão da série Nasdaq usando Holt-Winters foram estimadas sazonalidades de tamanho 1 até 400, sendo que para cada sazonalidade estimada os parâmetros α , β e γ foram testados independentemente com valores entre 0 e 1 a um passo de 0,1 a cada etapa. A figura 4.14 apresenta a previsão da série temporal do índice Nasdaq usando o modelo Holt-Winters e também apresenta o erro quadrático para o segmento de validação e teste da mesma série.

Figura 4.14 – Previsão da série Nasdaq usando o modelo Holt-Winters e o erro quadrático dos segmentos de validação e teste



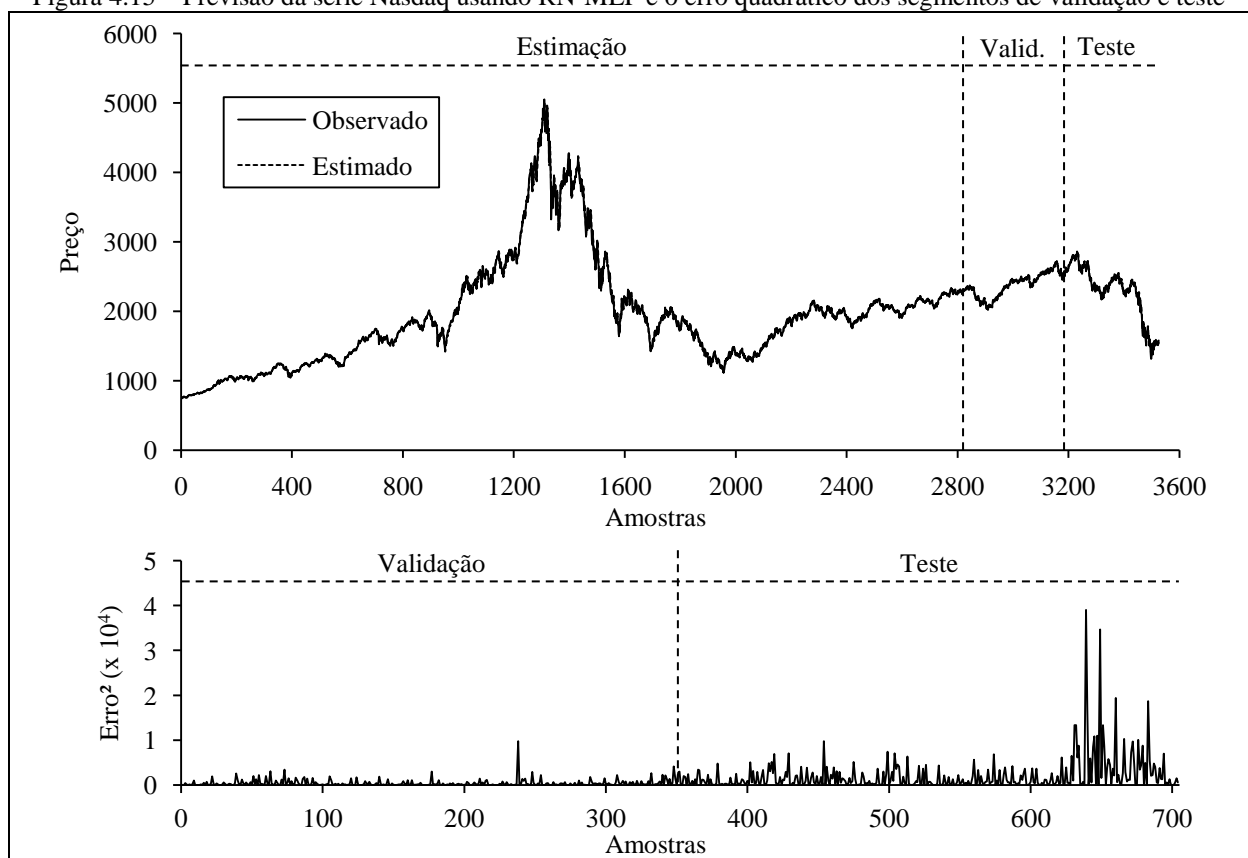
Considerando o MSE como índice de desempenho da fase de validação e buscando o menor valor para este índice, a melhor sazonalidade encontrada foi de 10, enquanto os parâmetros α , β e γ encontrados foram, respectivamente, 0,9, 0 e 0,3.

Os valores estimados na fase de estimação foram igualados aos valores observados da mesma fase, portanto os índices de desempenho para esta fase não foram considerados. Na fase de validação os valores do MSE, MAPE e R^2 obtidos foram, respectivamente, 453,59, 0,68990% e 0,99993. Enquanto na fase de teste os valores do MSE, MAPE e R^2 obtidos foram, respectivamente, 2081,37, 1,62590% e 0,98493.

4.4.3. Previsão usando RNs-MLP

Foram realizadas previsões para a série Nasdaq, usando RN-MLP, variando a defasagem de 1 até 10 e o número de neurônios na camada oculta de 1 até 25. O método de treinamento foi o Levenberg-Marquardt. A função de ativação da camada oculta foi usada a sigmóide e da camada de saída a linear. Foram 100 épocas de treinamento para cada iteração (defasagem por número de neurônios). A figura 4.15 apresenta a previsão da série temporal do índice Nasdaq usando RN-MLP e também apresenta o erro quadrático para o segmento de validação e teste da mesma série.

Figura 4.15 – Previsão da série Nasdaq usando RN-MLP e o erro quadrático dos segmentos de validação e teste



Foi buscado o menor MSE para a fase de validação da série temporal. Desta forma, o melhor valor para a defasagem foi de 3 e o melhor número de neurônios na camada oculta foi de 18.

Os índices de desempenho encontrados para esses dois parâmetros são apresentados na tabela 4.10.

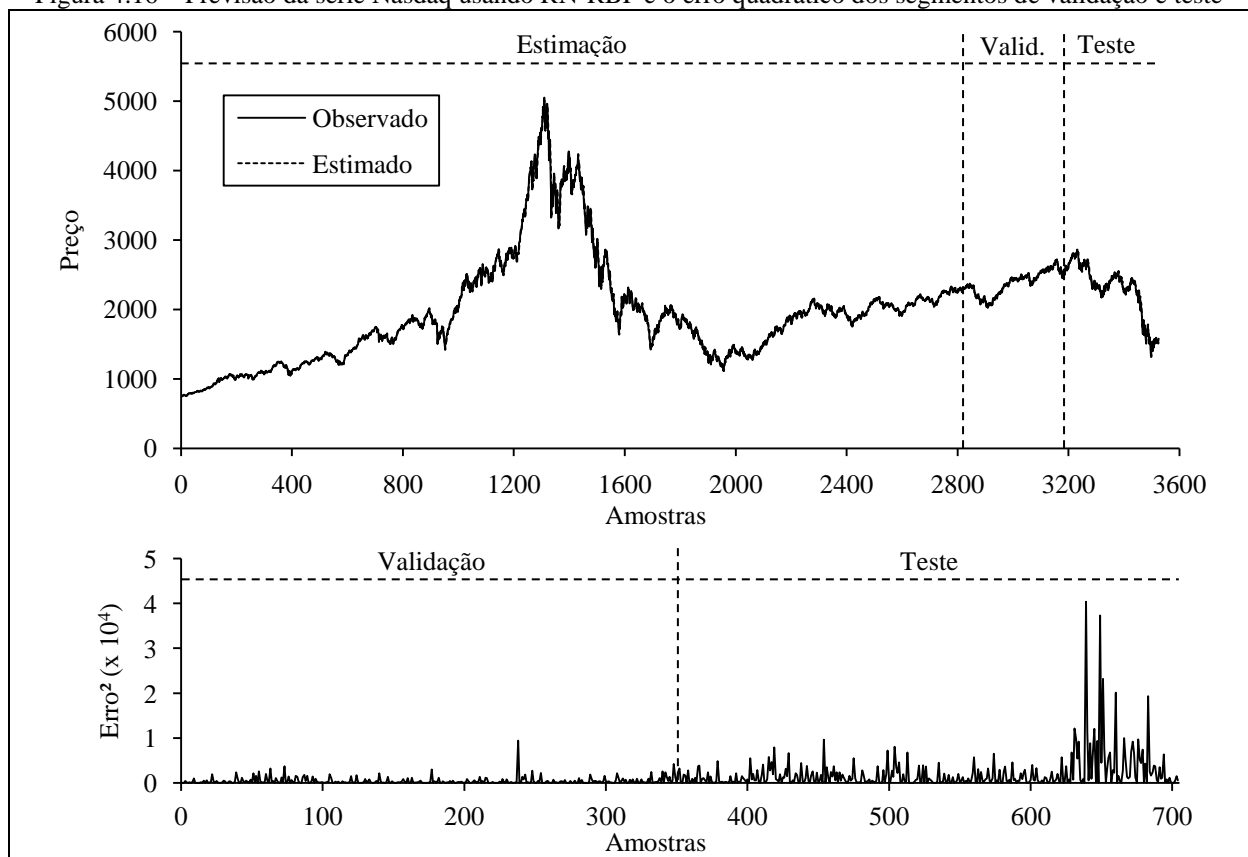
Tabela 4.10 – Índices de desempenho para a previsão da série Nasdaq usando RN-MLP

Fase	MSE	MAPE	R^2
Estimação	1796,83	1,22590%	0,99712
Validação	448,45	0,67599%	0,99993
Teste	2052,35	1,59980%	0,98514

4.4.4. Previsão usando RNs-RBF

Foram realizadas previsões para a série Nasdaq, usando RN-RBF, variando a defasagem de 1 até 50 e o número de centros de 2 até 50. Os *spreads* das funções de ativação da RN-RBF foram variados entre 1, 0,1, 0,01 e 0,001. A função de base radial usada foi a Gaussiana. O ajuste dos centros foi realizado pelo método *k-means* e o ajuste dos pesos através do algoritmo de mínimos quadrados.

Figura 4.16 – Previsão da série Nasdaq usando RN-RBF e o erro quadrático dos segmentos de validação e teste



A figura 4.16 apresenta a previsão da série temporal do índice Nasdaq usando RN-RBF e também apresenta o erro quadrático para o segmento de validação e teste da mesma série.

Foi buscado o menor MSE para a fase de validação da série temporal. Desta forma, o melhor valor para a defasagem foi de 1 e o melhor número de centros foi de 5. O *spread* que apresentou o melhor resultado foi 1.

Os índices de desempenho encontrados para esses dois parâmetros são apresentados na tabela 4.11.

Tabela 4.11 – Índices de desempenho para a previsão da série Nasdaq usando RN-RBF

Fase	MSE	MAPE	R^2
Estimação	1895,60	1,23110%	0,99697
Validação	457,40	0,68530%	0,99993
Teste	2060,65	1,60930%	0,98508

4.4.5. Comentários sobre os resultados da série Nasdaq

A tabela 4.12 apresenta os índices de desempenho de cada modelo estudado para a série Nasdaq.

Tabela 4.12 – Comparativo dos resultados para a série Nasdaq

Índice de desempenho	MMC com índice de saz.	Holt-Winters	RN-MLP	RN-RBF
Estimação				
MSE			1796,83	1895,60
MAPE			1,22590%	1,23110%
R^2			0,99712	0,99697
Validação				
MSE	458,50	453,59	448,45	457,40
MAPE	0,68805%	0,68990%	0,67599%	0,68530%
R^2	0,99993	0,99993	0,99993	0,99993
Teste				
MSE	2054,64	2081,37	2052,35	2060,65
MAPE	1,60710%	1,62590%	1,59980%	1,60930%
R^2	0,98513	0,98493	0,98514	0,98508

Para esta série temporal o modelo RN-MLP obteve o melhor desempenho em todos os índices, para todas as três fases da previsão.

Esta série, no entanto, apresenta uma característica diferente das demais estudadas: existe uma grande variação de preço entre a amostra 1000 e a 2000. Tal variação é referente à bolha das empresas “ponto com” nos Estados Unidos, e pouco foi sentido nas outras bolsas

estudadas neste trabalho. Ao analisar os parâmetros encontrados pelos modelos, percebe-se que o modelo RN-MLP conseguiu identificar melhor a série usando mais neurônios na camada oculta do que quando aplicado a outras séries temporais estudadas. O mesmo não aconteceu com a RN-RBF, que acabou encontrando o melhor modelo com defasagem igual a 1 e número de centros igual a 5.

4.5. Síntese conclusiva

Esta seção tem o objetivo de apresentar os resultados das previsões geradas pelos modelos para as séries em questão. Inicialmente uma descrição de cada série é apresentada a fim de deixar o leitor a par das características mais importantes.

Ao analisar as séries financeiras estudadas é possível observar que existiram períodos de crise, onde a variação de preço foi mais brusca. Esses períodos foram entre a amostra 1000 e 2000, que representa a crise das empresas “pontocom”, e mais atualmente, das últimas 500 amostras até o final da série, caracterizada pela crise mundial desencadeada pelo *subprime* americano. Também é possível perceber que a primeira crise, das empresas “pontocom”, afetou fortemente o índice Nasdaq, enquanto teve certa influência no índice Dow Jones e foi praticamente imperceptível no índice Ibovespa. A segunda crise que foi possível observar, desencadeada pelo *subprime*, afetou fortemente o índice Ibovespa e o índice Dow Jones, enquanto não foi tão forte para o índice Nasdaq.

A saída do sistema Mackey-Glass foi usada como *benchmark*, onde, considerando a fase de teste e o MSE como índice de desempenho, o melhor modelo de previsão foi a RN-MLP. O modelo Holt-Winters foi o pior modelo de previsão para esta série, enquanto a RN-RBF e o modelo MMC com índice de sazonalidade acabaram ficando em segundo e terceiro lugares, respectivamente.

Para as séries financeiras, Ibovespa, Dow Jones e Nasdaq, o modelo RN-MLP sempre superou o modelo RN-RBF na fase de estimação. Na fase de validação o modelo RN-MLP voltou a ser o melhor para a série Nasdaq, porém não apresentou resultado satisfatório para a série Dow Jones e Ibovespa. Enquanto isso os modelos MMC com índice de sazonalidade e Holt-Winters dividiram a liderança para as séries Dow Jones e Ibovespa.

Na fase de teste, onde foram utilizados dados fora da amostra, o modelo RN-MLP foi novamente o melhor para a série Nasdaq. O modelo Holt-Winters obteve o melhor MSE e R^2 para a série Dow Jones enquanto o modelo MMC com índice de sazonalidade superou a RN-

RBF como melhor MAPE para esta mesma série. Na série Ibovespa o modelo MMC com índice de sazonalidade superou os outros modelos nos índices MSE e R^2 , enquanto o modelo RN-RBF superou os demais no índice MAPE.

Os métodos, tanto lineares quanto não lineares, se alternaram como os melhores métodos entre uma série analisada e outra. Considerando somente as séries financeiras, e as fases de validação e teste, o modelo RN-MLP obteve o melhor desempenho para série Nasdaq, os modelos Holt-Winters, MMC com índice de sazonalidade e RN-RBF dividiram a liderança para série Dow Jones, e por fim os modelos lineares suplantaram os modelos não-lineares na série Ibovespa.

Uma possível explicação para o modelo RN-MLP ter sido, comparativamente com os outros modelos estudados, melhor para a série Nasdaq, mediano para a série Dow Jones e ruim para a série Ibovespa é que as amostras da fase de teste fazem parte de uma crise, onde os preços mudam abruptamente, e que durante a fase de estimação somente a série Nasdaq apresenta-se fortemente afetada por algum tipo de crise. Desta forma a possível explicação é que a RN-MLP aprendeu com a crise durante a fase de estimação e pode ter um desempenho melhor na fase de teste, onde outra crise estava em andamento.

Esta mesma idéia pode ser usada para explicar o porquê um método linear simples, como o MMC com índice de sazonalidade, obteve melhor desempenho que um não-linear mais complexo, como foi o caso da série Ibovespa. Como a primeira crise praticamente não afetou esta série, nenhum dos modelos mais complexos conseguiu prever com mais acurácia o que estava acontecendo na fase de teste.

Capítulo 5

Conclusão

Freqüentemente existe uma diferença de tempo entre um evento ou necessidade iminente e a ocorrência desse evento ou necessidade. Esta diferença de tempo é a principal razão para prever e planejar. A previsão é uma importante ajuda para um planejamento eficiente e eficaz (MAKRIDAKIS *et al.*, 1998).

A previsão de séries temporais é um problema que tem recebido especial atenção dos pesquisadores nos últimos anos. Prever o futuro, e em especial o comportamento de séries temporais, é fundamental em análises e tomada de decisões, e continua sendo um desafio para abordagens baseadas em estatística e computação (ABELÉM, 1994).

Com esta visão que a previsão de séries temporais é fundamental em análises e tomadas de decisões este trabalho teve como objetivo principal a comparação de alguns modelos de previsão, tanto lineares quanto não-lineares, para algumas séries temporais da área financeira, com foco num horizonte de curtíssimo prazo, ou seja, previsões de um passo à frente.

Foram utilizadas três séries temporais financeiras e um sistema com comportamento caótico para servir de *benchmark* com outros trabalhos. O sistema Mackey-Glass foi escolhida para servir de *benchmark* devido sua utilização para o mesmo fim em outros trabalhos. Nas séries financeiras o índice Ibovespa, o índice Dow Jones e o índice Nasdaq foram escolhidas como estudo de caso. Todas tiveram o mesmo período, de 1 de Janeiro de 1995 até 31 de Dezembro de 2008, utilizado. No entanto, o número de amostras acabou variando entre 3462 para a série do Ibovespa e 3526 para as séries Dow Jones e Nasdaq.

O desempenho dos modelos estudados foi analisado pelos índices de desempenho MSE, MAPE e R^2 .

Os modelos de previsão lineares estudados foram o de MMC com índice de sazonalidade e o Holt-Winters. Sendo o primeiro modelo mais simples e o segundo modelo mais robusto e também mais utilizado em outros trabalhos. No que diz respeito aos modelos não-lineares dois modelos de redes neurais foram escolhidos, a RN-MLP e a RN-RBF.

Os modelos foram submetidos a uma busca pelos melhores parâmetros, sendo cada série temporal tratada separadamente. Para cada parâmetro de cada modelo foi fixado uma faixa de valores e todas as combinações de parâmetros foram testadas para encontrar o MSE, que indicou qual foi o melhor conjunto de parâmetros usados.

Os métodos, tanto lineares quanto não lineares, se alternaram como os melhores métodos entre uma série analisada e outra. Considerando somente as séries financeiras, e as fases de validação e teste, o modelo RN-MLP obteve o melhor desempenho para série Nasdaq, os modelos MMC com índice de sazonalidade e Holt-Winters dividiram a liderança para série Dow Jones, e por fim os modelos lineares suplantaram os modelos não-lineares na série Ibovespa.

O modelo RN-MLP foi, comparativamente com os outros modelos estudados, melhor para a série Nasdaq, mediano para a série Dow Jones e ruim para a série Ibovespa. Uma possível resposta para esse comportamento é que as amostras da fase de teste fazem parte de uma crise, onde os preços mudam bruscamente, e que durante a fase de estimação somente a série Nasdaq apresenta-se fortemente afetada por algum tipo de crise. Desta forma, a RN-MLP aprendeu com a crise durante a fase de estimação e pode apresentar um desempenho melhor na fase de teste, onde outra crise estava em andamento.

Esta mesma linha também explicaria o motivo de um método linear simples, como o MMC com índice de sazonalidade, obter melhor desempenho que um não-linear mais complexo, como foi o caso da série Ibovespa. Como a primeira crise praticamente não afetou esta série, nenhum dos modelos mais complexos conseguiu prever com mais acurácia o que estava acontecendo na fase de teste.

Para trabalhos futuros, algumas modificações que seria interessante considerar: adicionar um componente de tendência no modelo MMC com índice de sazonalidade; adicionar um segundo componente de sazonalidade no modelo Holt-Winters, o chamado ciclo duplo; utilizar mais de uma camada oculta na RN-MLP, bem como outras funções de ativação do neurônio, a utilização de outros métodos de treinamento além do Levenberg-Marquardt e a adição de uma variável de entrada com o valor da tendência; no modelo RN-RBF o teste de

outras funções radiais, como multi quadrática, bem como outros métodos de treinamento e detecção de centros. Para as séries temporais também seria interessante estudar as mesmas séries com frequências maiores, como 15 minutos e 1 hora, ao invés de final do dia. Além de tudo isso, é possível que a utilização do volume negociado e os preços de abertura e média possam direcionar a resultados melhores de previsão.

Referências Bibliográficas

- ABELÉM, A. J. G. **Redes neurais artificiais na previsão de séries temporais.** 1994. Dissertação de Mestrado, Pós-Graduação em Engenharia Elétrica, PUCRJ, Rio de Janeiro, RJ.
- ALLIGOOD, R. T.; SAUER, T. D.; YORKE, J. A. **Chaos: an introduction to Dynamical Systems.** New York, NY, USA: Springer, 2000.
- AMJADY, N.; FARSHID K.. **Mid-term load forecasting of power systems by a new prediction method.** Energy Conversion and Management, doi: 10.1016/j.enconman.2008.04.008, 2008.
- ANÉ, T.; URECHE-RANGAU L.; GAMBET J.; BOUVEROT J. **Robust outlier detection for Asia-Pacific stock index returns.** Journal of International Financial Markets, Institutions and Money, vol. 18, n. 4, pp. 326-343, 2008.
- BARBIERO, C. C. M. **Séries temporais: um estudo de previsão para a receita operacional da ECT – Empresa Brasileira de Correios e Telégrafos.** 2003. Dissertação (Mestrado em Engenharia de Produção), Universidade Federal de Santa Catarina, Florianópolis, SC.
- BARBOSA, A. **Análise da demanda do álcool utilizando os métodos de suavização exponencial.** 2005. Monografia (Graduação em Estatística), Departamento de Estatística, Universidade Estadual de Maringá, Maringá, PR.
- BILLINGS, S.; HONG, X. **Dual-orthogonal radial basis function networks for nonlinear time series prediction.** Neural Networks, vol. 11, n.3, pp. 479-493. 1998.

- BONÉ, R.; CRUCIANU, M. **Multi-step-ahead Prediction with Neural Networks: a review**. In: *Approches Connexionnistes en Sciences Économiques et en Gestion*, pp. 97-106. Boulogne sur Mer, França, 2002.
- BOVESPA. **Índice Bovespa**. 2008. Disponível em: < <http://www.bovespa.com.br/Mercado/RendaVariavel/Indices/FormConsultaApresentacaoP.asp?Indice=Ibovespa>>, acessado em 24/08/2008.
- BRAGA, P. C. A. S. **Previsão do IBOVESPA utilizando modelos híbridos**. 2006. Dissertação de Mestrado, Programa de Pós-Graduação de Engenharia Elétrica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ.
- BRESSAN, A. A. **Tomada de decisão em futuros agropecuários com modelos de previsão de séries temporais**. *Revista de Administração de Empresas*. vol. 3, n. 1, Art. 9, jan./jun. 2004. São Paulo, SP: Fundação Getúlio Vargas, 2004.
- BROOMHEAD, D.; LOWE, D. **Multivariable functional interpolation and adaptative networks**. *Complex Systems*, vol. 2, n. , pp. 321–355, 1988.
- CASTRO, M. C. F. **Predição não-linear de séries temporais usando redes neurais RBF por decomposição em componentes principais**. 2001. Tese de Doutorado, Programa de Pós-Graduação em Engenharia Elétrica, Faculdade de Engenharia Elétrica e Computação, UNICAMP, Campinas, SP.
- CORRAR, L. J.; THEÓPHILO, C. R. **Pesquisa operacional para decisão em contabilidade e administração: contabilometria**. São Paulo, SP: Atlas, 2004.
- CORTEZ, P. A. R. **Algoritmos genéticos e redes neuronais na previsão de séries temporais**. 1997. Dissertação (Mestrado em Informática), Departamento de Informática, Universidade do Minho, Braga, Portugal.

- CORTEZ, P. A. R. **Modelos inspirados na natureza para previsão de séries temporais**. 2002. Tese (Doutorado em Informática), Departamento de Informática, Universidade do Minho, Guimarães, Portugal.
- DAVIS, M.; AQUILANO, N.; CHASE, R. **Fundamentos da administração da produção**. Porto Alegre, RS: Bookman, 2001.
- DE CASTRO, F., DE CASTRO, M. **Redes neurais artificiais**. Porto Alegre, RS: Pontifícia Universidade Católica do Rio Grande do Sul, 2001.
- DOWJONES. **Dow Jones industrial average – Overview**. 2009. Disponível em: <<http://www.djaverages.com/?view=industrial&page=overview>>, acessado em 25/06/2009.
- EHLERS, R.S. **Análise de séries temporais**. Disponível em <<http://leg.ufpr.br/~ehlers/notas>>. 2007. Acesso em: 20/07/2008.
- FALCO, G. P. **Técnicas univariadas aperfeiçoadas para a previsão de curtíssimo prazo a partir de dados horários**. 2005. Dissertação (Mestrado em Engenharia Elétrica), Programa de Pós-Graduação em Engenharia Elétrica, PUCRJ, Rio de Janeiro, RJ.
- FREITAS, A. A. C. de. **Previsão de séries temporais via seleção de variáveis, reconstrução dinâmica, ARMA-GARCH e redes neurais artificiais**. 2007. Tese (Doutorado em Engenharia Elétrica), Faculdade de Engenharia Elétrica e de Computação, UNICAMP, Campinas, SP.
- FREITAS, S., SOUZA, A. **Utilização de um modelo baseado em redes neurais para a precificação de opções**. In: Encontro Nacional de Administração – ENANPAD. Anais do XX Encontro Nacional de Administração – ENANPAD. Salvador, BA, 2002.
- GAITHER, N.; FRAZIER, G.. **Administração da produção e operações**. São Paulo: Thomson, 2002. 598 p.

- GÓMEZ-RAMÍREZ, E.; NAJIM, K.; IKONEN, E. **Forecasting time series with a new architecture for polynomial artificial neural network.** Applied Soft Computing, vol. 7, n. 4, pp. 1209-1216. 2007.
- GOOIJER, J. G.; HYNDMAN, R. J. **25 years of time series forecasting.** International Journal of Forecasting, vol. 22, n. 3, pp. 443-473, 2006.
- GUERRA, F., COELHO, L. S.. **Identificação de sistema dinâmico caótico usando rede neural *perceptron* multicamadas.** In: Simpósio Sul-brasileiro de Matemática e Informática. Curitiba, 2002.
- GUJARATI, D. N. **Econometria Básica.** Rio de Janeiro, RJ: Elsevier, 2006.
- HAMZAÇEBI, C. **Improving artificial neural networks' performance in seasonal time series forecasting.** Information Sciences, vol. 178, n. 23, pp. 4550-4559. 2008.
- HAYKIN, S. **Redes neurais: princípios e prática.** 2. Ed. Porto Alegre, RS: Bookman, 2001.
- HSIEH, D. **Chaos and Nonlinear Dynamics: Application to Financial Markets.** Journal of Finance, vol. 46, n. 5, pp. 1839-1916. 1990.
- JANG, J. S. R. **ANFIS: Adaptive network-based fuzzy inference system.** IEEE Transactions on Systems, Man and Cybernetics, vol. 23, n. 3, pp. 665-685. 1993.
- JONES, R. D.; LEE, Y. C.; BARNES, C. W.; FLAKE, G. W.; LEE, K.; LEWIS, P. S.; QIAN, S. **Function approximation and time series prediction with neural network.** Proceedings of the International Joint Conference on Neural Networks, vol. 1, pp. 649-665. 1990.
- JUNG, C. F. **Metodologia Para a Pesquisa & Desenvolvimento.** Rio de Janeiro, RJ: Axcel Books do Brasil Editora, 2004.

- LEVENBERG, K. **A method for the solution of certain non-linear problems in least squares.** Quart. Applied Mathematics, vol. 2, n. 2, pp. 164-168, 1944.
- LEWIS, C. D. **Demand forecasting and inventory control: a computer aided learning approach.** New York, USA: John Wiley & Sons, 1997.
- LIMA, F. G. **Um método de análise e previsão de sucessões cronológicas unidimensionais lineares e não-lineares.** 2004. Tese (Doutorado em Administração), Programa de Pós-Graduação em Administração, Universidade de São Paulo, São Paulo, SP.
- LOPES, R. D. **Previsão de autopeças: estudo de caso em uma concessionária de veículos.** 2002. Dissertação (Mestrado em Engenharia de Produção), Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, SC.
- MACKEY, M. C.; GLASS L. **Oscillation and chaos in physiological control systems.** Science, vol. 197, n. 4300, pp. 287-289. 1977.
- MAKRIDAKIS, S. G.; WHEELWRIGHT, S. C.; HYNDMAN, R. J. **Forecasting: methods and applications.** 3^a. ed. New York, USA: Wiley, 1998.
- MARTINS, P. G.; LAUGENI, F. P. **Administração da produção.** 2^a. ed. São Paulo, SP: Saraiva, 2005.
- MARQUARDT, D. W. **An algorithm for least-squares estimation of nonlinear parameters.** SIAM Journal of Applied Mathematics. vol. 11, n. 2, pp. 431-441, 1963.
- MCCULLOCH, W. S.; PITTS, W. **A logical calculus of the ideas immanent in nervous activity.** Bulletin of Mathematical Biophysics, vol. 5, n. 4, pp. 115-133. 1943.

- MIRANDA, C. V. C. **Previsão de dados de alta frequência para carga elétrica usando Holt-Winters com dois ciclos.** 2007. Dissertação de Mestrado, Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ.
- MOODY, J.; DARKEN, C. **Fast learning in networks of locally-tuned processing units.** Neural Computation, vol.1, n. 2, pp. 281-294, 1989.
- MORETTIN, P. A.; TOLOI, C. M. **Séries temporais.** 2^a. ed., São Paulo, SP: Atual, 1987.
- NASDAQ. **NASDAQ Composite Index Methodology.** 2009. Disponível em: <<http://www.nasdaqtrader.com/content/home/help/indexmethod/CompIndexMethod.pdf>>, acessado em 27/06/2009.
- PAIVA, R. P. P. C. **Identificação neuro-difusa: aspectos de interpretabilidade.** 1999. Dissertação (Mestrado em Engenharia Elétrica), Departamento de Engenharia Informática, Universidade de Coimbra, Coimbra, Portugal.
- ROSENBLATT, F. **The Perceptron: a probabilistic model for information storage and organization in the brain.** Psychological Review, vol. 65, n. 6, pp. 386-408, 1958.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. **Learning internal representations by error propagation.** Em: Parallel distributed processing – MIT Press, vol. 1, pp. 318-362, 1986.
- SANTOS, A. A. P. **Previsão não-linear da taxa de câmbio real/dólar utilizando redes neurais e sistemas nebulosos.** 2005. Dissertação (Mestrado em Economia), Programa de Pós-Graduação em Economia, Universidade Federal de Santa Catarina, Florianópolis, SC.

- SANTOS, A. A. P.; COSTA JUNIOR, N. C. A.; COELHO, L. S. **Computational intelligence approaches and linear models in case studies of forecasting exchange rates.** *Expert Systems with Applications*, vol. 33, n. 4, pp. 816-823, 2007.
- SELIM, H.. **Determinants of house prices in Turkey: hedonic regression versus artificial neural network.** *Expert Systems with Applications*, vol. 36, n. 2, pp. 2843-2852, 2009.
- SOTO, C. P. **Redes neurais temporais para o tratamento de sistemas variantes no tempo.** 1999. Dissertação (Mestrado em Engenharia Elétrica), Departamento de Engenharia Elétrica, PUCRJ, Rio de Janeiro, RJ.
- SOUZA, A. L. C. **Uso do Excel e do *Crystal Ball Predictor* para análise e previsão de preços de commodities do setor sucroalcooleiro.** 2005b. Monografia (Graduação em Engenharia de Produção), Universidade Federal de Itajubá, Itajubá, MG.
- SOUZA, G. P. **Previsão do consumo industrial de energia elétrica no estado de Santa Catarina: uma aplicação da combinação de previsões entre modelos univariados e de regressão dinâmica.** 2005a. Dissertação (Mestrado em Engenharia de Produção), Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, SC.
- SOUZA, G. P. ; SAMOHYL, R. W. ; MEURER, R. **Previsão do consumo de energia elétrica do setor industrial em Santa Catarina - um estudo comparativo entre diferentes métodos de previsão através de suas discrepâncias.** In: XXXVI Simpósio Brasileiro de Pesquisa Operacional - O Impacto da Pesquisa Operacional nas Novas Tendências Multidisciplinares, São João Del Rei, MG. 2004.
- SOUZA, R. C. ; BARROS, M.; MIRANDA, C. V. C. **Short term load forecasting using double seasonal exponential smoothing and interventions to account for holidays and temperature effects.** In: TLAIO II - 2 do Taller Latino Iberoamericano de Investigación de Operaciones, Acapulco, México. 2007.

- TAYLOR, J. W. **Exponential Smoothing with a damped multiplicative trend.** International Journal of Forecasting, vol. 19, n. 4, 715-725, 2003.
- TEIXEIRA, L. C. A. **Análises do padrão de resposta da frequência cardíaca pelos métodos de séries temporais e semiparamétrico e de sua variabilidade na determinação do limiar de anaerobiose.** 2003. Dissertação (Mestrado em Bioengenharia), Programa de Pós-Graduação Interunidade Bioengenharia, Universidade Federal de São Carlos, São Carlos, SP.
- TUBINO, D. F.. **Manual de planejamento e controle da produção.** 2^a. ed., São Paulo, SP: Atlas, 2000.
- WINTERS, P. R. **Forecasting sales by exponentially weighted moving averages.** Management Science, vol. 6, n. 3, pp. 324-342, 1960.
- ZHANG, G. P. **An investigation of neural networks for linear time-series forecasting.** Computers & Operations Research, vol. 28, n. 12, pp. 1183-1202, 2001.
- ZHANG, G. P.; QI, M. **Neural network forecasting for seasonal and trend time series.** European Journal of Operational Research, vol. 160, n. 2, pp. 501-514, 2005.
- ZHANG, G; PATUWO, B. E.; HU, M.Y. **Forecasting with artificial neural networks: the state of the art.** International Journal of Forecasting, Ohio, USA, vol. 14, n. 1, pp. 35-62, 1998.
- ZOU, H. F.; XIA, G. P.; YANG, F. T.; WANG, H. Y. **An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting.** Neurocomputing, vol. 70, n. 16-18, pp. 2913-2923, 2007.