

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ  
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO E  
SISTEMAS**

**MÉTODOS BOX-JENKINS E NEURO-NEBULOSO ANFIS APLICADOS À  
PREVISÃO DE SÉRIES TEMPORAIS**

**CURITIBA  
2009**

**ALINE PURCOTE**

**MÉTODOS BOX-JENKINS E NEURO-NEBULOSO ANFIS APLICADOS À  
PREVISÃO DE SÉRIES TEMPORAIS**

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.

Orientador: Prof. Dr. Leandro dos Santos Coelho

**CURITIBA**

**2009**

## **AGRADECIMENTOS**

Agradeço primeiramente a Deus, por ter me concedido as graças necessárias para a realização deste sonho.

Aos meus pais Leila Aparecida Ferreira Purcote e Érico Purcote pelo imenso carinho e pelo verdadeiro amor dedicado ao meu crescimento, educação e formação.

Ao meu noivo Marcelis Quinsler pela sua paciência, apoio e amor imensurável e a minha irmã Zuleika Purcote pelo seu companheirismo durante esta jornada.

Ao meu orientador Leandro dos Santos Coelho agradeço a positiva influência que exerceu em minha formação acadêmica, por sua compreensão e pela sua impecável orientação a este trabalho.

Aos demais parentes, amigos, professores e alunos que das mais diferentes formas, participaram desta história.

## SUMÁRIO

<b>LISTA DE FIGURAS</b>	<b>6</b>
<b>LISTA DE TABELAS</b>	<b>8</b>
<b>LISTA DE ABREVIATURAS</b>	<b>10</b>
<b>RESUMO</b>	<b>11</b>
<b>ABSTRACT</b>	<b>12</b>
<b>1 INTRODUÇÃO</b>	<b>13</b>
1.1 OBJETIVOS	16
1.2 JUSTIFICATIVA	17
1.3 TRABALHOS CORRELATOS	18
1.4 ORGANIZAÇÃO DO TRABALHO	21
<b>2 SÉRIES TEMPORAIS</b>	<b>22</b>
2.1 CONSIDERAÇÕES GERAIS	22
2.2 ANÁLISE DE SÉRIES TEMPORAIS	23
2.3 PROCESSO ESTOCÁSTICO E SÉRIE TEMPORAL	24
2.4 ESTACIONARIEDADE	25
2.5 RUÍDO BRANCO	26
2.6 MODELOS E PROCEDIMENTOS DE PREVISÃO	26
2.7 ÍNDICES DE DESEMPENHO	27
2.8 ESTUDO DE CASO	29
<b>3 REVISÃO DA LITERATURA SOBRE MÉTODOS DE PREVISÃO</b>	<b>30</b>
3.1 MÉTODO DE HOLT-WINTERS	30
3.2 MODELOS BOX-JENKINS	32
3.2.1 Identificação dos Modelos	35
3.2.2 Estimação dos Parâmetros	37
3.2.3 Verificação	38

3.2.4	Previsão	41
3.3	REDES NEURAIS ARTIFICIAIS	41
3.3.1	Modelo de um Neurônio	45
3.3.2	Arquiteturas de Redes Neurais	47
3.4	SISTEMAS NEBULOSOS	50
3.4.1	Sistema de Inferência Nebuloso	55
3.5	SISTEMA NEURO-NEBULOSO DO TIPO ANFIS	58
3.5.1	Arquitetura ANFIS	59
<b>4</b>	<b>OBTENÇÃO E ANÁLISE DOS RESULTADOS</b>	<b>63</b>
4.1	SIMULAÇÕES PARA A SÉRIE INSUMO X	65
4.1.1	Simulações usando Modelo ARIMA	65
4.1.2	Simulações usando Modelo Holt-Winters	69
4.1.3	Simulações usando Modelo ANFIS	73
4.1.4	Análise comparativa dos modelos para o Insumo X	77
4.2	SIMULAÇÕES PARA A SÉRIE INSUMO Y	78
4.2.1	Simulações usando Modelo ARIMA	78
4.2.2	Simulações usando o Modelo Holt-Winters	83
4.2.3	Simulações usando Modelo ANFIS	87
4.2.4	Análise comparativa dos modelos para o Insumo Y	91
<b>5</b>	<b>CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS</b>	<b>93</b>
	<b>REFERÊNCIAS</b>	<b>94</b>

## LISTA DE FIGURAS

Figura 2.1 – Representação esquemática de um sistema linear	23
Figura 3.1 – Ciclo iterativo de um modelo ARIMA	35
Figura 3.2 - Espectro acumulado do ruído branco	40
Figura 3.3 – Representação do sistema nervoso	42
Figura 3.4 – Modelo não-linear de um neurônio	46
Figura 3.5 – Rede alimentada adiante com uma única camada de neurônios	48
Figura 3.6 – Rede alimentada adiante totalmente conectada com uma camada oculta e uma camada de saída	49
Figura 3.7 – Rede recorrente	50
Figura 3.8 – Exemplo de Função de pertinência triangular	53
Figura 3.9 – Exemplo de Função de pertinência Gaussiana	54
Figura 3.10 – Exemplo de Função de pertinência sino	55
Figura 3.11 – Sistema de inferência nebuloso	56
Figura 3.12 – Modelo nebuloso de Takagi-Sugeno	58
Figura 3.13 – Arquitetura ANFIS do tipo Takagi-Sugeno	60
Figura 3.14 – Modelo nebuloso do tipo Sugeno de primeira ordem	60
Figura 4.1 – Dados do Insumo X	63
Figura 4.2 – Dados do Insumo Y	64
Figura 4.3 – Dados do Insumo X versus previsão do modelo ARIMA (Simulação 1)	66
Figura 4.4 – Dados do Insumo X versus previsão do modelo ARIMA (Simulação 2)	66
Figura 4.5 – Erros do modelo ARIMA Insumo X (Simulação 1)	67
Figura 4.6 – Erros do modelo ARIMA Insumo X (Simulação 2)	68
Figura 4.7 – Insumo X versus previsão do modelo HW	69
Figura 4.8 – Erros do modelo HW Insumo X	70
Figura 4.9 – Insumo X versus previsão do modelo HW (Simulações 3 e 4)	71
Figura 4.10 – Erros do modelo HW Insumo X (Simulações 3 e 4)	72
Figura 4.11 – Insumo X versus previsão do modelo ANFIS (Simulação 1)	74
Figura 4.12 – Insumo X versus previsão do modelo ANFIS (Simulação 2)	75
Figura 4.13 – Insumo X versus previsão do modelo ANFIS (Simulação 3)	76

Figura 4.14 – Insumo Y versus previsão do modelo ARIMA (simulação 1)	79
Figura 4.15 – Insumo Y versus previsão do modelo ARIMA (simulação 2)	80
Figura 4.16 – Erros do modelo ARIMA para o Insumo Y (Simulação 1)	81
Figura 4.17 – Erros do modelo ARIMA para o Insumo Y (Simulação 2)	82
Figura 4.18 – Insumo Y versus previsão do modelo HW	83
Figura 4.19 – Erros do modelo HW para o Insumo Y	84
Figura 4.20 – Insumo Y versus previsão do modelo HW (Simulações 3 e 4)	85
Figura 4.21 – Erros do modelo HW para o Insumo Y (Simulações 3 e 4)	86
Figura 4.22 – Insumo Y versus previsão do modelo ANFIS (Simulação 1)	88
Figura 4.23 – Insumo Y versus previsão do modelo ANFIS (Simulação 2)	89
Figura 4.24 – Insumo Y versus previsão do modelo ANFIS (Simulação 3)	90

## LISTA DE TABELAS

Tabela 4.1- Índices de Desempenho do modelo ARIMA para o Insumo X	68
Tabela 4.2- Índices de Desempenho do modelo HW para o Insumo X	71
Tabela 4.3- Índices de desempenho do modelo HW para o Insumo X (Simulações 3 e 4)	72
Tabela 4.4 - Resultados estrutura ANFIS para a Simulação 1 usando dados do Insumo X	73
Tabela 4.5 – Índices de desempenho do modelo ANFIS para o Insumo X (Simulação 1)	74
Tabela 4.6 - Resultados da estrutura ANFIS para a simulação 2 usando dados do Insumo X	75
Tabela 4.7 – Índices de Desempenho do modelo ANFIS para o Insumo X (Simulação 2)	75
Tabela 4.8 - Resultados estrutura ANFIS para a Simulação 3 usando dados do Insumo X	76
Tabela 4.9 – Índices de desempenho do modelo ANFIS para o Insumo X (simulação 3)	77
Tabela 4.10 – Comparativo de resultados para o Insumo X	78
Tabela 4.11 - Índices de Desempenho do modelo ARIMA para o Insumo Y (Simulação 1)	82
Tabela 4.12 - Índices de Desempenho do modelo ARIMA para o Insumo Y (Simulação 2)	82
Tabela 4.13- Índices de desempenho do modelo HW para o Insumo Y	85
Tabela 4.14- Índices de Desempenho do modelo HW para o Insumo Y (simulações 3 e 4)	86
Tabela 4.15 - Resultados estrutura ANFIS para a simulação 1 usando dados do Insumo Y	87
Tabela 4.16 – Índices de desempenho do modelo ANFIS para o Insumo Y (simulação 1)	88
Tabela 4.17 - Resultados estrutura ANFIS para a simulação 2 usando dados do Insumo Y	88

Tabela 4.18 – Índices de desempenho do modelo ANFIS para o Insumo Y (Simulação 2)	89
Tabela 4.19 - Resultados estrutura ANFIS para a Simulação 3 usando dados do Insumo Y	90
Tabela 4.20 – Índices de desempenho do modelo ANFIS para o Insumo Y (simulação 3)	90
Tabela 4.21 – Comparativo de resultados para o Insumo Y	91

## LISTA DE ABREVIATURAS

ANFIS - *Adaptive Neuro-Fuzzy Inference System*

AR - Auto Regressivo (do inglês, *Auto Regressive*)

ARIMA – Auto Regressivo Integrado Médias Móveis (do inglês, *Auto Regressive Integrated Moving Average*)

BJ - Box-Jenkins

CAT - *Criterion Autoregressive Transfer function*

FIS – Sistema de Inferência *fuzzy* ( do inglês, *Fuzzy Inference Systems*)

HW - Holt-Winters

I - Integração

MA - Médias Móveis (do inglês, *Moving Average*)

MAPE - Erro Percentual Médio Absoluto (do inglês, *Mean Absolute Percentual Error*)

MSE - Erro Médio Quadrático (do inglês, *Mean Squared Error*)

RMSE - Raiz Quadrada do Erro Médio Quadrático (do inglês, *Rooted Mean Squared Error*)

RNAs - Redes Neurais Artificiais

$R^2$  - Coeficiente de determinação múltipla

VLSI - *Very-Large-Scale-Integrated*

## RESUMO

O planejamento e abordagens de apoio à tomada de decisão podem ter a previsão como um dos seus subsídios, uma vez que fornece informações que possibilitam o planejamento com antecedência e, conseqüentemente, permite que os recursos produtivos estejam disponíveis na quantidade, momento e qualidade adequados. Previsões acuradas auxiliam no desenvolvimento de estratégias, identificação de prioridades e alocação de recursos, além de permitir que as organizações ofereçam elevados níveis de serviço aos clientes, planejem expansões de capacidade e evitem perdas nas vendas e estoques. No desejo de compreender o passado e prever o futuro, pode-se determinar o comportamento futuro de um sistema a partir de observações do passado. Nestas situações, tem-se a Previsão de Séries Temporais como potencial ferramenta. Esta dissertação aborda dentre os modelos clássicos de previsão de séries temporais, os modelos Box-Jenkins e Holt-Winters. Os modelos tradicionais de previsão são de fácil desenvolvimento e implementação, mas não são capazes, muitas vezes, de capturar relações não-lineares contidas nos dados. Neste contexto é também abordado, neste trabalho, o modelo neuro-nebuloso *Adaptive Neuro-Fuzzy Inference System* (ANFIS), pois é um modelo matemático que consegue mapear complexas relações não-lineares contidas em dados. O ANFIS combina os fundamentos das redes neurais artificiais e dos sistemas nebulosos em um sistema híbrido neuro-nebuloso, que pode ser utilizado para a previsão de séries temporais. O objetivo geral desta dissertação é verificar a qualidade de previsão um passo à frente usando modelos lineares e não-lineares. Os modelos matemáticos analisados foram Box-Jenkins, Holt-Winters e ANFIS. Estes modelos foram validados em séries de consumo mensal de insumos da empresa paranaense Beta, considerando como horizonte de tempo o período de novembro de 2003 até maio de 2009.

**PALAVRAS-CHAVE:** Box-Jenkins, Holt-Winters, Sistema neuro-nebuloso, Séries Temporais, Previsão.

## **ABSTRACT**

The planning and approaches to support decision-making can be forecasting as one of its subsidies, as it provides information that enable the planning in advance and hence allows productive resources that are available in quantity, time and quality. Assist in developing accurate forecasts of strategies, identifying priorities and allocating resources and allowing organizations to offer high levels of service to customers, plan expansions of capacity and avoid losses in sales and inventories. The desire to understand the past and predict the future, you can determine the future behavior of a system from observations observed in the past. In these situations, it has been time series forecast as a potential tool. This dissertation addresses among classical models for time series forecasting, Box-Jenkins and Holt-Winters models. The traditional forecasting models are easy to develop and implement, but are unable, often, to capture nonlinear relationships contained in the data. In this context is also addressed, in this work, the Adaptive Neuro-Fuzzy Inference System (ANFIS), it is a mathematical model that can map complex nonlinear relationships contained in data. The ANFIS combines the fundamentals of artificial neural networks and fuzzy systems in a hybrid neuro-fuzzy systems, which can be used for time series predicting. The object of this study is to assess the quality of forecasting one step ahead using linear and nonlinear models. The mathematical models were analyzed Box-Jenkins, Holt-Winters and ANFIS. These models were validated in a series of monthly consumption of inputs of the Beta paranaense company, considering the time horizon as the period from November 2003 until May of 2009.

**KEYWORDS:** Box-Jenkins, Holt-Winters, Fuzzy System, Time Series, Forecasting.

## 1 INTRODUÇÃO

As condições econômicas e empresariais variam ao longo do tempo, diante disto, os gestores necessitam encontrar maneiras de se manterem informados quanto aos efeitos que essas mudanças têm sobre suas decisões. Uma técnica que pode ser utilizada para ajudar no planejamento de futuras necessidades operacionais é a previsão.

Apesar dos numerosos métodos de previsão propostos, muitos deles possuem um objetivo em comum: fazer previsões de eventos futuros para que estas projeções possam então ser incorporadas no apoio à tomada de decisão.

Os procedimentos de previsão utilizados na prática variam, podendo ser simples e intuitivos, onde pouca ou nenhuma análise de dados é envolvida, ou mais quantitativos e complexos, onde a análise pode ser considerável.

Vários modelos estatísticos já foram propostos para previsão de séries temporais, a maioria baseados em métodos lineares. Um modelo linear para previsão é uma função que relaciona os valores da série com valores anteriores (auto-regressão), seus erros (médias móveis) ou com outras séries ou variáveis (regressão múltipla). Dentre os modelos clássicos têm-se os modelos Box-Jenkins, conhecidos como modelos ARIMA (Auto-Regressivo Integrado Médias Móveis, do inglês *Autoregressive Integrated Moving Average*) e o modelo Holt-Winters.

Os modelos Box-Jenkins visam captar o comportamento da correlação seriada ou autocorrelação entre os valores da série temporal, e com base nesse comportamento realizam previsões (KIRCHNER, 2006). A modelagem baseia-se na construção de funções fundamentadas em um ciclo iterativo que utiliza os próprios dados da série para determinar uma classe de modelos que permita fazer previsões. Inicialmente, é proposta uma classe de modelos que permite a obtenção de um modelo com base em critérios específicos, em seguida os parâmetros são estimados e, através da análise dos resíduos, o modelo ajustado é avaliado; caso não seja adequado, o ciclo é repetido (MARTÍNEZ e ZAMPROGNO, 2003). O sucesso da metodologia Box-Jenkins deve-se ao fato de que diferentes modelos podem, entre eles, conter o comportamento de diversos tipos de séries sem a necessidade de um acentuado número de parâmetros a serem estimados à escolha do modelo (GOOIJER e HYNDMAN, 2006). De maneira geral, são necessárias pelo menos 50

observações para o desenvolvimento de um modelo aceitável de Box-Jenkins. Este fato pode impossibilitar a obtenção dos modelos em situações onde não existem muitas observações disponíveis (LOPES, 2007).

O método Holt-Winters é um dos mais utilizados para a previsão de curto prazo, devido à sua simplicidade, baixo custo computacional, precisão e capacidade de ajustamento automático e rápido a mudanças na série (SOUTO, BALDEÓN e RUSSO, 2006). Este método é utilizado em cenários em que as séries temporais apresentam, geralmente, padrão de demanda com tendência linear e sazonalidade. O método procura determinar as três componentes da decomposição da série: nível, tendência e sazonalidade. Morettin e Tolo (1987) atribuem como desvantagens deste método, a dificuldade de se determinar os melhores valores para as constantes de suavização, as dificuldades de estudo de propriedades estatísticas, tais como a média e a variância de previsão, que dificultam a construção do intervalo de confiança para os valores previstos.

Os modelos clássicos de previsão são de fácil implementação, mas não são capazes de capturar muitas relações não-lineares contidas nos dados. Assim, as Redes Neurais Artificiais (RNAs) e o *Adaptive Neuro-Fuzzy Inference System* (ANFIS) constituem uma alternativa promissora para a previsão, pois são modelos versáteis que podem mapear relações não-lineares presentes nos dados.

Uma RNA consiste de um processador paralelamente distribuído constituído de unidades de processamento simples (neurônios), que tem a propensão de armazenar conhecimento e torná-lo disponível para o uso. As RNAs são inspiradas no funcionamento do cérebro humano em dois aspectos: o conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem e as forças de conexão entre os neurônios denominados pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido. É evidente que uma RNA extrai seu poder computacional através de sua estrutura paralelamente distribuída e de sua habilidade de aprender e, portanto de generalizar. A generalização se refere ao fato da RNA gerar saídas adequadas para entradas que não estavam presentes durante o seu treinamento (HAYKIN, 2001).

As RNAs têm a capacidade de aprender padrões subjacentes presentes nos conjuntos de dados, apresentando desempenho geralmente superior aos modelos lineares tradicionais, principalmente quando o sistema regente dos dados é desconhecido, não-linear e/ou não-estacionário. Por esta razão, as RNAs podem

representar uma contribuição ao estudo das séries temporais resultantes de tais sistemas (HAYKIN, 2001).

As RNAs cujos parâmetros podem ser alterados ao longo do processo de aprendizado podem ser empregadas diretamente em uma série de aplicações envolvendo modelagem de sistema, apoio à tomada de decisão, processamento de sinais, controle de processos, entre outras áreas. Estas redes são chamadas de redes adaptativas e existe uma classe que são funcionalmente equivalentes a sistemas de inferência nebulosa. Um dos modelos mais difundidos é o modelo proposto por Roger Jang denominado *Adaptive Neuro-Fuzzy Inference System* (ANFIS).

As RNAs aparecem como uma alternativa para o projeto de sistemas nebulosos, pois permitem que parte do conhecimento do especialista seja obtida via treinamento, resultando em sistemas neuro-nebulosos com alto poder de adaptação. O sistema híbrido conhecido como ANFIS consiste na estruturação da base de regras nebulosas por meio de uma RNA.

Os sistemas nebulosos possuem um papel fundamental nas mais diversas áreas do processamento de informação e representam uma ponte de ligação entre o processamento simbólico e o numérico. A teoria de sistemas nebulosos utiliza símbolos, termos lingüísticos, com os quais estão associadas semânticas bem definidas que, após serem convertidas em funções de pertinência de conjuntos nebulosos, possibilitam o processamento numérico destes símbolos ou conceitos. Em sistemas baseados em regras nebulosas, informações imprecisas e incompletas, em analogia às informações sensoriais recebidas pelos órgãos e interpretadas pelo cérebro, são tratadas de forma lingüística. A estrutura básica de um sistema nebuloso possui três componentes conceituais: uma base de regras, que contém o conjunto de regras nebulosas; uma base de dados, que define as funções de pertinência usadas nas regras nebulosas e um mecanismo de raciocínio, que realiza um procedimento de inferência (raciocínio nebuloso) para derivar a saída ou conclusão, baseado nas regras e fatos conhecidos. A seleção de uma base de regras nebulosas, associada à escolha de semânticas e operadores apropriados, abre caminho para a criação de sistemas nebulosos que modelam e processam o conhecimento do especialista em aplicações específicas (DELGADO, 2002).

Os sistemas nebulosos representam uma importante ferramenta de modelagem baseada na teoria dos conjuntos nebulosos e têm sido aplicados com

sucesso em diversas áreas, como: controle automático, classificação e reconhecimento de padrões, tomada de decisão, sistemas inteligentes, previsão de séries temporais e robótica (DELGADO, 2002).

O ANFIS, proposto por Jang (1992), é uma das primeiras arquiteturas neuro-nebulosas. Esta arquitetura possui aplicação nas áreas de previsão de funções não-lineares e controle inteligente. O ANFIS processa variáveis lingüísticas, como o processamento encontrado nos sistemas de inferência nebulosa. Apesar desta característica, é implementado sob o paradigma de redes neurais; por conseqüência, está categorizado como um sistema híbrido incorporado. A principal vantagem de sua implementação sob o domínio da arquitetura de RNAs está em incorporar a capacidade de aprendizado. O ANFIS é um sistema que pode processar as variáveis lingüísticas com a capacidade de aprendizado de uma rede neural (SANDMANN, 2006).

## 1.1 OBJETIVOS

O objetivo geral deste trabalho é verificar a qualidade de previsão um passo à frente usando modelos lineares e não-lineares.

Dentre os modelos clássicos, os modelos Box-Jenkins (BJ) e Holt-Winters (HW) são analisados. Já o modelo não-linear analisado é o modelo ANFIS.

Para este estudo são utilizadas duas séries de consumo mensal de insumos entre os meses de novembro de 2003 e maio de 2009 disponibilizados pela empresa paranaense Beta, totalizando 67 observações.

Os objetivos específicos são:

- Estudar e aplicar modelos clássicos utilizados na previsão de séries temporais, incluindo os modelos de Box-Jenkins (BJ) e Holt-Winters (HW);
- Estudar e validar o modelo ANFIS para aplicação em previsão de séries temporais;
- Analisar comparativamente o desempenho dos modelos de previsão utilizando os índices de desempenho  $R^2$  (coeficiente de determinação múltipla), o Erro Médio Quadrático (MSE, do inglês *Mean Squared Error*), o Erro Percentual Médio Absoluto (MAPE, do inglês *Mean Absolute Percentual Error*) e Raiz Quadrada do Erro Médio Quadrático (RMSE, do inglês *Rooted Mean Squared Error*).

## 1.2 JUSTIFICATIVA

Independentemente do tamanho da empresa é importante um plano que direcione as ações do presente como meio de atingir as metas para o futuro. O planejamento da empresa requer organização, controles e ferramentas que auxiliem na tarefa de tomada de decisão.

Para facilitar e respaldar o processo de apoio à tomada de decisão pode-se utilizar metodologias matemáticas e estatísticas, as quais servem de suporte para transformar dados reais em processos de modelagem, ajudando no processo decisório.

A previsão constitui um meio de fornecer informações e subsídios para uma conseqüente tomada de decisão. A previsão constitui a base para as decisões estratégicas, táticas e operacionais e tem sido utilizada, com sucesso em planejamento financeiro, *marketing*, produtivo, pessoal, entre outros (OLIVEIRA, 2007). De forma crescente, as organizações têm reconhecido a importância da precisão das previsões em suas operações efetivas e na melhoria da competitividade. Para a produção, ela é fundamental para a tomada de decisões periódicas envolvendo seleção de processo, planejamento de capacidade, melhorias de *layout*, decisões contínuas sobre planejamento da produção, programação, compra dos insumos utilizados em seus processos produtivos e gestão de seus estoques.

A rapidez nas mudanças torna necessária a disponibilização imediata de informações relevantes e confiáveis para decisões eficazes. O mercado atual é altamente competitivo e seletivo, exigindo não somente eficiência, mas principalmente eficácia.

A previsão tem sido dominada por modelos lineares, por muitas décadas. Os modelos lineares são de fácil desenvolvimento e implementação, além de compreensão e interpretações relativamente simples. Entre os modelos mais populares, destacam-se os modelos de Box-Jenkins (ARIMA) para a previsão de séries temporais. O fato dos modelos de Box-Jenkins serem modelos lineares pode ser um fator limitante se o fenômeno em estudo for governado por dinâmica não-linear.

Por outro lado, os sistemas neuro-nebulosos estão entre os sistemas híbridos mais pesquisados, por associarem vantagens de duas técnicas de modelagem, as

RNAs e os sistemas nebulosos. Esses sistemas combinam a capacidade de aprendizado das RNAs com o poder de interpretação lingüístico dos sistemas de inferência *fuzzy* (CAMPOS, JESUS, MENDES, 2007). O modelo ANFIS é um exemplo e implementa um modelo híbrido capaz de transformar conhecimento ou experiência em base de regras, que através de um algoritmo de aprendizagem adequado, é capaz de construir automaticamente um conjunto de regras nebulosas, a partir de um conjunto de dados.

Os sistemas baseados em lógica nebulosa permitem que o comportamento do sistema possa ser descrito apenas com relações do tipo SE-ENTÃO. O ANFIS surge como uma ferramenta flexível em um ambiente dinâmico e possui a capacidade de aprender rapidamente padrões complexos, tendências presentes nos dados e se adaptar a mudanças, características estas que são desejáveis em se tratando de previsão de séries temporais.

### 1.3 TRABALHOS CORRELATOS

A seguir é apresentada uma revisão da literatura sobre os modelos matemáticos analisados, que incluem o Box-Jenkins (ARIMA), o Holt-Winters e o ANFIS.

Zandonade (1993) associou previsão de séries temporais a RNAs e aplicou o algoritmo de retropropagação do erro (*error backpropagation*) na previsão de quatro séries temporais: uma série ruidosa, uma série com tendência, uma série sazonal e uma série relativa ao Consumo de Energia Elétrica da cidade de Uruguaiana no Rio Grande do Sul. Os resultados obtidos foram comparados com os modelos ARIMA e um modelo com intervenção. Para todas as séries foi obtida a previsão pontual superior aos modelos ARIMA. Para a série com tendência, a RNA captou a tendência utilizando os dados diferenciados. A série sazonal mostrou a capacidade da rede de aprender a componente periódica.

Abelém (1994) investigou a utilização das RNAs na previsão de séries temporais. A motivação para o uso da RNA foi determinar melhores resultados que os apresentados pelos procedimentos estatísticos convencionais, como por exemplo, o modelo Box-Jenkins. Os procedimentos estatísticos convencionais, na sua maioria, são lineares possuindo dificuldades no tratamento de dados reais compostos, muitas vezes, de complexas relações não-lineares. O estudo de caso

apresentado validou a capacidade das RNAs na previsão de séries temporais e os resultados obtidos comparados com o modelo Box-Jenkins comprovaram a superioridade das RNAs no tratamento de dados não-lineares e ruidosos.

Oliveira (2002) propôs uma metodologia para ajudar na solução do controle dos estoques nas empresas em geral, contribuindo para a otimização de seus processos, reduzindo custos e liberando espaços ociosos. Utilizando o modelo Box-Jenkins, com base em dados passados, a demanda dos produtos foram projetadas para o sistema de controle de estoques. Como resultado final foi apresentado um *software* de previsão de demanda e controle de estoques para metodologia Box-Jenkins. O projeto baseado em modelos ARIMA forneceu resultados promissores tanto na fase de simulação quanto em casos reais utilizados.

Zhang (2003) propôs uma metodologia híbrida que combina os modelos ARIMA e as RNAs. O modelo linear ARIMA e o modelo não-linear de RNAs são utilizados conjuntamente visando capturar as diferentes formas de relacionamento das séries temporais. Os resultados da simulação indicaram que a combinação dos modelos pode ser uma forma eficaz de melhorar a precisão da previsão obtida por qualquer um dos dois modelos, quando utilizados separadamente.

Martínez e Zamprogno (2003) compararam o desempenho das metodologias de previsão de séries temporais Box-Jenkins, alisamento exponencial e RNAs para diferentes horizontes de previsão. Foram analisadas duas séries de acordo com estas metodologias e suas previsões são comparadas sob diferentes medidas de ajuste. Dos modelos avaliados, o algoritmo de Holt-Winters com sazonalidade multiplicativa foi o que apresentou os melhores resultados de previsão e, também, o menor custo computacional do que os modelos de Box-Jenkins, onde o modelo selecionado de RNAs apresentou um desempenho superior para horizontes maiores de previsão.

Brusaferro (2006) buscando entender melhor o comportamento do mercado automotivo brasileiro anual propôs analisar e prever o mercado nacional de veículos utilizando três modelos: modelo de série temporal, modelo função de transferência com a previsão suplementar das variáveis macroeconômicas e modelo neuro-nebuloso. Para o modelo de série temporal foi utilizado o modelo ARIMA(1,1,0) e para o estudo de função de transferência foram utilizadas as variáveis PIB brasileiro e renda média. Utilizaram-se dois modelos neuro-nebuloso: um simples, com os anos como entrada e outro complexo que utiliza como entrada valores anteriores do

mercado. Cada modelo foi aplicado a duas amostras, uma amostra de 48 amostras (dados de 1957 a 2004) e outra de 28 amostras (dados de 1977 a 2004). Comparando os modelos estudados, para a amostragem de 48 observações o melhor modelo, baseado no menor somatório dos quadrados dos erros, foi o modelo neuro-nebuloso, já para a amostra de 28 amostras o modelo com maior aderência foi o modelo de função de transferência.

Kirchner (2006) buscando conhecer o comportamento do faturamento de uma empresa do ramo agrícola e de suas filiais realizou previsões utilizando a metodologia Box-Jenkins. Ao realizar as previsões e compará-las com os dados reais disponibilizados pela empresa, constatou a eficiência dos modelos utilizados. Com base nos resultados obtidos, pode-se afirmar, que a metodologia Box-Jenkins possibilita realizar previsões, se constituindo em uma ferramenta eficaz ao gestor, a qual favorece o planejamento das atividades da empresa, infra-estrutura e, conseqüentemente, contribui para a manutenção, crescimento e sucesso da mesma.

Sandmann (2006) aplicou a arquitetura ANFIS na predição de funções que geram séries temporais. O objetivo principal foi apresentar uma técnica em sistemas de arquitetura neuro-nebulosa para predição de uma função utilizando duas séries temporais: uma função logística e o mapa de Mackey-Glass. Com este trabalho foi possível comprovar a eficiência do ANFIS para a predição de funções, mesmo que utilizando dados de entrada com funções de pertinência de espaço de variáveis geradas automaticamente.

Campos, Jesus e Mendes (2007) apresentaram duas abordagens de previsão de consumo de energia elétrica de curto prazo: o modelo nebuloso lingüístico de Mamdani e uma abordagem do sistema ANFIS. Os resultados obtidos mostraram que a previsão usando o sistema ANFIS apresentou uma boa generalização quanto a previsão e um melhor resultado quando comparado ao modelo Mamdani. No modelo Mamdani a previsão realizada foi de apenas um passo à frente, enquanto que no modelo ANFIS esse horizonte de previsão foi maior, apresentando erros menores de previsão. O modelo ANFIS também apresenta um modelo adaptativo para a predição, enquanto o modelo baseado em Mamdani necessita de conhecimento humano especialista para operação e escolha do modelo correto para cada tipo de situação.

Oliveira (2007) propôs um estudo comparativo entre modelos lineares e as RNAs, avaliando as tecnologias geradoras de previsão de valores, para identificação

dos métodos que garantiam confiança e índices elevados de acerto. O estudo foi realizado na gestão de risco de instituições financeiras. Os modelos foram aplicados em duas séries temporais: uma de valores mensais e outra de valores diários. Na série de valores mensais, o modelo Auto Regressivo (AR) obteve os melhores índices de precisão entre os modelos lineares; na série de valores diários o modelo que combinou os modelos AR com regressão múltipla obteve os melhores índices. Em todas as séries analisadas, o modelo de RNAs que utilizou os erros do modelo linear como entrada da RNA apresentou os melhores resultados.

Valenzuela *et al.* (2008) propuseram uma hibridização de técnicas inteligentes tais como RNAs, sistemas nebulosos e algoritmos evolutivos, de modo que o modelo híbrido ARIMA combinado a RNA superou a qualidade da previsão desses modelos quando utilizados separadamente. Foi proposta a utilização de regras nebulosas para obter a ordem dos modelos ARMA ou ARIMA bem como a utilização de um modelo híbrido de ARIMA e RNAs que combina as vantagens dos modelos ARIMA com o poder computacional das RNAs. As simulações realizadas mostraram que a sinergia dos diferentes paradigmas e técnicas utilizadas produz excelentes resultados para a construção de sistemas inteligentes visando à previsão de séries temporais.

#### 1.4 ORGANIZAÇÃO DO TRABALHO

O restante desta dissertação está estruturada da seguinte forma: no Capítulo 2 é apresentada a descrição do problema, detalhando os principais conceitos relacionados a séries temporais, seus modelos de previsão e índices de desempenho adotados. No Capítulo 3, os conceitos dos métodos de previsão de séries temporais Box-Jenkins, Holt-Winters e ANFIS são detalhados. No Capítulo 4 são apresentados os resultados obtidos e as análises comparativas entre os modelos analisados. No capítulo final são apresentadas as conclusões, limitações da pesquisa, além das perspectivas de trabalhos futuros.

## 2 SÉRIES TEMPORAIS

### 2.1 CONSIDERAÇÕES GERAIS

Uma série temporal é um conjunto de observações de uma dada variável, ordenado segundo o parâmetro tempo  $t$ , geralmente em intervalos eqüidistantes. Se  $Z_t$  representa o valor da variável aleatória  $Z$  no instante  $t$ , denota-se a série temporal por  $Z_1, Z_2, \dots, Z_N$ , onde  $N$  é o tamanho da série ou número de observações seriais da variável. As séries temporais podem ser classificadas em (SOUZA e CAMARGO, 2004):

- a) Discretas: quando o conjunto de observações for finito ou infinito numerável, isto é,  $Z = \{1, 2, \dots, t\}$ ;
- b) Contínuas: quando o conjunto for infinito não-numerável, isto é,  $N = \{t: 0 < t < N\}$ ;
- c) Determinísticas: quando uma função matemática pode ser usada para estabelecer exatamente os valores futuros da série;
- d) Estocásticas: quando os valores futuros da série só podem ser estabelecidos em termos probabilísticos;
- e) Multivariadas: quando a série temporal é representada por um vetor  $\{Z(t), t \in N\}$  de ordem  $rx1$ ;
- f) Multidimensional: quando se tem  $\{Z(t), t \in N\}$  e  $t$  é um vetor de ordem  $px1$ .

Há, basicamente, dois enfoques usados na análise de séries temporais. Em ambos, o objetivo é construir modelos para as séries, com propósitos determinados. No primeiro enfoque, a análise é realizada no domínio temporal, mede-se a magnitude do evento que ocorre em determinado instante. A análise no domínio temporal é baseada em um modelo paramétrico (número finito de parâmetros), utilizam-se as funções de autocovariância e autocorrelação. No segundo enfoque, a análise é conduzida no domínio da freqüência e os modelos propostos são modelos não paramétricos, onde se destaca a análise espectral.

## 2.2 ANÁLISE DE SÉRIES TEMPORAIS

Os objetivos da realização da análise de séries temporais estão vinculados à:

- a) Descrição: inicia-se a análise de séries temporais com a construção de um gráfico que mostra como a série evolui no tempo. Obtêm-se também medidas descritivas de suas principais características;
- b) Explicação: tendo-se um conjunto de observações de duas ou mais variáveis, pode ser possível explicar o comportamento de uma em função das demais. Se as observações estão indexadas no tempo, tem-se a teoria de função de transferência presente em sistemas lineares. Um sistema linear converte uma entrada em uma saída através de uma operação de filtragem, conforme apresentado na Figura 2.1.

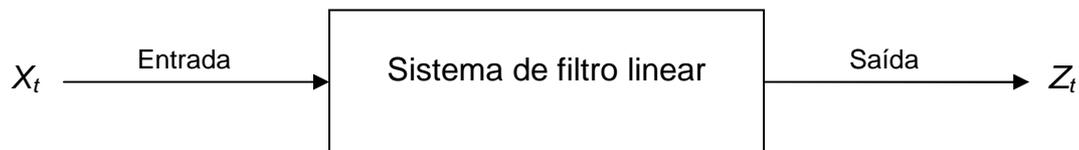


Figura 2.1 – Representação esquemática de um sistema linear

Na figura 2.1,  $X_t$  representa a entrada de um sistema  $S$  e  $Z_t$  representa a saída. A operação que relaciona a série  $Z_t$  à série  $X_t$  é denominada filtro. A função que relaciona a série de entrada à da saída é a função de transferência. Se a série  $Z_t$  puder ser escrita como uma combinação linear de termos da série  $X_t$ , tem-se um filtro linear. Frequentemente, a filtragem linear é apenas uma combinação linear das observações passadas e presentes da série de entrada.

- c) Previsão: a previsão do comportamento futuro da variável constitui um dos principais objetivos da análise de séries temporais;
- d) Controle: quando uma série temporal mede qualidade, o objetivo da análise pode ser de controlar o processo gerador.

O interesse principal deste trabalho é a previsão de séries temporais e seus métodos de previsão.

### 2.3 PROCESSO ESTOCÁSTICO E SÉRIE TEMPORAL

Um modelo que descreve a estrutura de probabilidade de uma seqüência de observações é denominado processo estocástico. Os processos estocásticos são sistemas que evoluem no tempo e/ou no espaço de acordo com leis probabilísticas (SOUZA e CAMARGO, 2004).

O propósito da teoria dos processos estocásticos é o estudo daqueles mecanismos dinâmicos que proporcionam meios de análise de uma seqüência de observações, vistas conjunta e independentemente em uma sucessão de momentos de tempo, as quais são influenciadas por fatores aleatórios. Ao se observar o comportamento de uma série temporal, verifica-se que esta pode ser considerada como uma particular realização de uma seqüência de observações produzidas por um mecanismo probabilístico. Dessa maneira, uma série temporal pode ser abordada como uma realização de um processo estocástico (KIRCHNER,2006).

Um processo estocástico é uma família  $Z=\{Z(t), t \in N\}$  tal que para cada  $t \in R$ ,  $Z(t)$  é uma variável aleatória. Se  $N \equiv Z = \{1, \dots, t\}$ , diz-se que o processo é de parâmetro discreto, denotado por  $Z_t$ . Se  $N \in R$ , diz-se que o processo é de parâmetro contínuo, denotando-se por  $Z(t)$  (SOUZA e CAMARGO, 2004).

Considerando-se que  $Z_t$  origina-se de um experimento que pode ser repetido sob condições idênticas; a cada experimento obtém-se um registro de valores de  $Z_t$  no tempo, sendo que cada um deles é uma realização do processo. Assim, uma série temporal pode ser observada como uma parte da trajetória ou de uma realização parcial do processo estocástico, isto é, uma amostra finita de observações no tempo.

O conceito de processo estocástico proporciona uma relevante ferramenta à análise probabilística de séries temporais. Seja uma série temporal  $Z_t, t = 1, \dots, N$  com  $N$  observações sucessivas, pode-se ver a série como sendo extraída de uma distribuição de probabilidade conjunta  $P(Z_1, \dots, Z_N)$ . A série é então observada como uma realização amostral dentre todas as séries possíveis de tamanho  $N$ , que poderiam ter sido geradas por um mesmo mecanismo subjacente, denominado processo estocástico.

Um processo estocástico está estatisticamente determinado quando se conhece suas funções de distribuição até a  $n$ -ésima ordem. Na prática, ocorrem situações problemáticas por não se conhecerem todas as funções de distribuição até

a  $n$ -ésima ordem e por ter-se apenas uma realização do processo estocástico, a partir da qual se deseja inferir características do mecanismo gerador da série. Para superar essas dificuldades assumem-se duas restrições: estacionariedade e ergodicidade (SOUZA e CAMARGO, 2004).

## 2.4 ESTACIONARIEDADE

Uma das suposições mais freqüentes que se faz a respeito de uma série temporal é a de que ela é estacionária, ou seja, ela se desenvolve no tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio estável.

Um processo estacionário satisfaz as seguintes condições:

- i)  $E[Z_t] = E[Z_{t+k}] = \mu(t), \forall t;$
- ii)  $E[(Z_t - \mu)^2] = \sigma^2(t) = \sigma^2, \sigma^2 < \infty, \forall t;$
- iii)  $Cov[Z_t, Z_{t+k}] = Cov[Z_{t+m}, Z_{t+k+m}], \forall m.$

onde  $\mu$  é o nível da série e  $Cov$  é a função de autocovariância.

A importância do conhecimento da série ser ou não estacionária reside no fato de que, quando se trabalha com uma série estacionária, se está em presença de uma função amostral do processo que tem a mesma forma em todos os instantes do tempo ( $t \in N$ ), o que acarreta possibilidades de obtenção de estimativa das características do processo de forma simples.

Como a maioria dos procedimentos de análise estatística de séries temporais supõem que estas sejam estacionárias, é necessário transformar os dados originais, se estes não formam uma série estacionária. Uma transformação simples consiste em tomar diferenças sucessivas da série original, até se obter uma série estacionária. A primeira diferença de  $Z_t$  é definida por (2.1):

$$\Delta Z_t = Z_t - Z_{t-1}, \quad (2.1)$$

a segunda diferença é dada por (2.2)

$$\Delta^2 Z_t = \Delta[\Delta Z_t] = \Delta[Z_t - Z_{t-1}], \quad (2.2)$$

ou seja,

$$\Delta^2 Z_t = Z_t - 2Z_{t-1} + Z_{t-2}. \quad (2.3)$$

De modo geral, a  $n$ -ésima diferença de  $Z_t$  é definida por (2.4).

$$\Delta^n Z_t = \Delta[\Delta^{n-1} Z_t]. \quad (2.4)$$

Em situações normais, será suficiente tomar uma ou duas diferenças para que a série se torne estacionária.

## 2.5 RUÍDO BRANCO

É uma seqüência de variáveis aleatórias independentes e identicamente distribuídas, não necessariamente, mas usualmente com distribuição normal  $N(0, \sigma_a)$  com média zero e variância constante  $\sigma_a^2$ . Assim, o processo  $a_t$ ,  $t = 0, 1, 2, \dots$  é um ruído branco se:

$$a_t \equiv \text{Dist} [0, \sigma_a^2], \text{ com Média: } E[a_t] = 0. \quad (2.5)$$

$$\text{Autocovariância: } E[a_t, a_{t+k}] = \gamma_k = 0; \text{ para } k = \pm 1, \pm 2, \dots \text{ e } \gamma_k = \sigma_a^2 \text{ para } k = 0. \quad (2.6)$$

$$\text{Autocorrelação: } \rho_k = 1 \text{ para } k = 0 \text{ e } \rho_k = 0 \text{ para } k = \pm 1, \pm 2, \dots \quad (2.7)$$

$$\text{Autocorrelação parcial: } \phi_{kk} = 1 \text{ para } k = 0 \text{ e } \phi_{kk} = 0 \text{ para } k = \pm 1, \pm 2, \dots \quad (2.8)$$

$$\text{Densidade espectral: } f(w) = \frac{1}{2\pi} \quad (2.9)$$

$$\text{Densidade espectral (não normalizada): } h(w) = \frac{\sigma_a^2}{2\pi}. \quad (2.10)$$

## 2.6 MODELOS E PROCEDIMENTOS DE PREVISÃO

Um modelo pode ser considerado uma descrição probabilística de uma série temporal e cabe ao usuário decidir como utilizar esse modelo (MORETTIN e TOLOI, 1987). O objetivo de um modelo é clarificar a estrutura e a dinâmica do fenômeno e também, ajudar a fazer previsões ou mesmo controlar o sistema.

Um modelo que descreve uma série não conduz, necessariamente, a um procedimento de previsão. Será necessário especificar uma função objetivo, além do modelo, para se chegar ao procedimento (MORETTIN e TOLOI, 1987).

A construção de um modelo se faz em três etapas:

- i) Indução: construir o modelo a partir de observações;
- ii) Dedução: estudo das implicações do modelo;
- iii) Verificação: experimentos com o modelo e testes de previsões.

Suponha-se que existam observações disponíveis até o instante  $t$  e deseja-se prever o valor da série no instante  $t+h$ . Utilizando a notação descrita em MORETTIN e TOLOI (1987) tem se:

$t$ : origem;

$h$ : horizonte de previsão;

$Z(t+h)$ : observação real no instante  $t+h$ ;

$\hat{Z}_t(h)$ : previsão para  $Z(t+h)$ .

Os horizontes de previsão são classificados, geralmente, em: prazo imediato (menor que um mês), curto (de um a três meses), médio (de três a seis meses) e longo prazo (um ano ou mais) (SOUZA e CAMARGO, 2004). Neste trabalho é utilizado como horizonte de previsão o curto prazo.

Vale a pena ressaltar, que a previsão não constitui um fim em si, mas apenas um meio de fornecer informações para uma conseqüente tomada de decisão, visando determinados objetivos (MORETTIN e TOLOI, 1987).

Inúmeras abordagens de previsão de séries temporais são encontradas na literatura. Estas abordagens podem ser organizadas em: técnicas descritivas, modelos lineares e modelos não-lineares.

Dentre os modelos apresentados na literatura, são analisados neste trabalho os modelos Box-Jenkins, Holt-Winters e ANFIS. Esses modelos são apresentados, em detalhes, no capítulo 3.

## 2.7 ÍNDICES DE DESEMPENHO

Existem muitas medidas de desempenho para a análise da qualidade de previsão. Entretanto, a maioria descreve a diferença entre os valores reais e os valores previstos da série, conhecidos como erro ou resíduo.

Para avaliação da acurácia da previsão será adotado, nesta dissertação, o Coeficiente de Determinação Múltipla, também denominado  $R^2$ , o Erro Médio

Quadrático (MSE), o Erro Percentual Médio Absoluto (MAPE) e a Raiz Quadrada do Erro Médio Quadrático (RMSE).

O  $R^2$  mede o grau de associação entre os valores observados e os calculados; é um número no intervalo [0;1] e calculado pela equação (2.11).

$$R^2 = 1 - \frac{\sum_{t=1}^N (y(t) - \hat{y}(t))^2}{\sum_{t=1}^N (y(t) - \bar{y})^2}, \quad (2.11)$$

onde  $y$  é o valor real,  $\hat{y}$  é o valor estimado, e  $\bar{y}$  é a média das  $N$  amostras medidas. Quando o valor de  $R^2$  é igual a 1, tem-se uma aproximação perfeita do modelo para os dados medidos do sistema.

O MSE consiste em calcular a média dos erros ao quadrado, elevando os erros de previsão ao quadrado, somando e dividindo-os pelo número de períodos.

Sendo assim, considerando-se que  $y_i$  sejam os valores reais,  $y_i'$  sejam os valores estimados e  $N$  seja o número de períodos, o cálculo do MSE é descrito pela equação (2.12).

$$MSE = \frac{\sum_{t=1}^n (y_i(t) - y_i'(t))^2}{N}. \quad (2.12)$$

Devido ao cálculo da média do erro ao quadrado, os grandes erros se destacam, quando comparados aos erros de menor magnitude. Valores de MSE iguais a zero indicam uma exata adequação do modelo para os dados medidos do sistema.

Com o objetivo de avaliar a magnitude do erro com relação à série histórica, o MAPE calcula os erros de previsão em percentuais absolutos.

Considerando  $y_i$  os valores reais da série,  $y_i'$  os valores previstos e  $N$  o número de períodos de previsão, a fórmula para cálculo do MAPE é dada pela equação (2.13).

$$MAPE = \frac{\sum_{i=1}^n \frac{|y_i - y_i'|}{y_i}}{N}. \quad (2.13)$$

O RMSE avalia o tamanho do erro de previsão e é calculada através da média de erros de cada previsão em relação ao valor real, elevados ao quadrado. Quanto mais próximo de zero é o RMSE melhor é o modelo de previsão. O cálculo do RMSE é descrito pela equação (2.14).

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - y'_i)^2}{N}}. \quad (2.14)$$

## 2.8 ESTUDO DE CASO

A empresa Beta é uma empresa paranaense líder nacional em capacidade operacional e segmento de mercado, considerada uma das maiores da América Latina, instalada numa área de 60 mil m<sup>2</sup> na cidade de Curitiba (PR). Empresa fundada em 1972 atua em todas as regiões brasileiras e possui atuação internacional, atendendo as maiores empresas brasileiras e clientes na América Latina, Estados Unidos da América, Europa e Japão. Possui um centro de distribuição em São Paulo, além de representações em todo o Brasil e Mercosul. Reconhecida pelo mercado por seu comprometimento socioambiental e sua proposta de serviços com qualidade, tecnologia e bom atendimento, sendo à primeira empresa brasileira em seu segmento a obter as certificações ISO 14001 e 9001, recebendo diversos prêmios nacionais e internacionais.

Inicialmente foi realizada uma pesquisa na base de dados da empresa para analisar o consumo mensal das matérias-primas básicas utilizadas em seu processo produtivo. Após avaliação, foram escolhidos dois diferentes insumos utilizados em processos produtivos distintos, considerando o volume de dados disponíveis e variações no consumo mensal dos insumos. As duas séries possuem como horizonte de tempo o período de novembro de 2003 até maio de 2009, totalizando, em cada série, 67 observações mensais. As séries históricas selecionadas são analisadas através de métodos quantitativos de previsão com o objetivo de gerar previsões um passo à frente e realizar uma análise comparativa dos modelos utilizados. A aderência dos modelos aos dados é avaliada através dos indicadores de desempenho  $R^2$ , MSE, MAPE e RMSE, os quais fornecem o modelo mais adequado às séries analisadas.

### 3 REVISÃO DA LITERATURA SOBRE MÉTODOS DE PREVISÃO

#### 3.1 MÉTODO DE HOLT-WINTERS

Em 1960, Winters estendeu o modelo de Holt, incluindo uma nova equação que possibilitasse ser acrescentada nas previsões o comportamento da componente sazonal dos dados sob análise, propondo assim o Método de Holt-Winters.

O método de Holt-Winters procura avaliar três componentes determinísticas da decomposição da série: nível, tendência e sazonalidade. Dependendo das características da série considerada, existem dois modelos que estão relacionados com as relações entre as componentes, ou seja, um modelo aditivo e outro multiplicativo, sendo o último mais usual segundo MORETTIN e TOLOI (1981). Tais procedimentos são baseados em três equações com constantes de suavização diferentes, que são associadas a cada uma das componentes do padrão da série.

No modelo aditivo, a amplitude da variação sazonal é constante ao longo do tempo, ou seja, a diferença entre o maior e menor valor da série permanece relativamente constante no tempo. Já no modelo multiplicativo, a amplitude da variação sazonal aumenta ou diminui em função do tempo.

O modelo sazonal multiplicativo em um período de sazonalidade  $s$  considera um fator de sazonalidade  $F_t$  como multiplicativo e conforme MORETTIN e TOLOI (1981) têm-se (3.1).

$$Z_t = \mu_t F_t + T_t + a_t, \quad t = 1, \dots, n, \quad (3.1)$$

onde  $\mu_t$  é o nível da série e pode ser estimado por  $\tilde{Z}_t$ .

As equações de suavização são definidas em função das constantes de suavização  $A$ ,  $C$  e  $D$  e representam, respectivamente, os fatores de nível, tendência e sazonalidade conforme equações 3.2, 3.3 e 3.4.

$$\tilde{Z}_t = A \left( \frac{Z_t}{\hat{F}_{t-s}} \right) + (1 - A)(\hat{Z}_{t-1} + \hat{T}_{t-1}) \quad 0 < A < 1, \quad t = s+1, \dots, n \quad (3.2)$$

$$\hat{T}_t = C (\bar{Z}_t - \bar{Z}_{t-1}) + (1 - C)\hat{T}_{t-1} \quad 0 < C < 1, \quad t = s+1, \dots, n \quad (3.3)$$

$$\hat{F}_t = D \left( \frac{Z_t}{\bar{Z}_t} \right) + (1 - D)\hat{F}_{t-s} \quad 0 < D < 1, \quad t = s+1, \dots, n. \quad (3.4)$$

A previsão para o procedimento multiplicativo utiliza as equações (3.5) e (3.6) em função do horizonte:

- Previsões dentro do período sazonal

$$\hat{Z}_t(h) = (\bar{Z}_t + h\hat{T}_t)\hat{F}_{t+h-s}, \quad h = 1, 2, \dots, s. \quad (3.5)$$

- Previsões entre o primeiro e o segundo período sazonal "s"

$$\hat{Z}_t(h) = (\bar{Z}_t + h\hat{T}_t)\hat{F}_{t+h-2s}, \quad h = s+1, s+2, \dots, 2s. \quad (3.6)$$

Este procedimento de previsão, por semelhança, pode ser estendido aos demais horizontes.

Os valores iniciais para as equações recorrentes são obtidos a partir das relações (3.7), (3.8) e (3.9).

$$F_j = \frac{Z_j}{\left(\frac{1}{s}\right)\sum_{k=1}^s Z_k}, \quad j = 1, 2, \dots, s \quad (3.7)$$

$$\bar{Z}_s = \left(\frac{1}{s}\right)\sum_{k=1}^s Z_k, \quad (3.8)$$

$$\hat{T}_s = 0. \quad (3.9)$$

O modelo sazonal multiplicativo pode ser modificado para tratar situações onde o fator sazonal é aditivo, conforme equação (3.10).

$$Z_t = \mu_t + T_t + F_t + a_t. \quad (3.10)$$

As estimativas do fator sazonal, nível e tendência da série são dadas pelas equações (3.11), (3.12) e (3.13).

$$\hat{F}_t = D(Z_t - \bar{Z}_t) + (1 - D)\hat{F}_{t-s}, \quad 0 < D < 1 \quad (3.11)$$

$$\bar{Z}_t = A(Z_t - \hat{F}_{t-s}) + (1 - A)(\bar{Z}_{t-1} + \hat{T}_{t-1}), \quad 0 < A < 1 \quad (3.12)$$

$$\hat{T}_t = C(\bar{Z}_t - \bar{Z}_{t-1}) + (1 - C)\hat{T}_{t-1}, \quad 0 < C < 1 \quad (3.13)$$

onde  $A$ ,  $C$  e  $D$  são as constantes de suavização.

As previsões dos valores futuros da série são definidas pelas equações (3.14) e (3.15).

$$\hat{Z}_t(h) = \bar{Z}_t + h\hat{T}_t + \hat{F}_{t+h-s}, \quad h = 1, 2, \dots, s, \quad (3.14)$$

$$\hat{Z}_t(h) = \bar{Z}_t + h\hat{T}_t + \hat{F}_{t+h-2s}, \quad h = s+1, \dots, 2s, \quad (3.15)$$

onde  $\bar{Z}_t$ ,  $\hat{T}_t$  e  $\hat{F}_t$  são dados por (3.12), (3.13) e (3.11), respectivamente.

A aplicação do método de Holt-Winters não exige que os dados originais respeitem uma distribuição normal e/ou sejam estacionários, tratando-se, portanto,

de um método determinístico, o que o torna popular, por ser considerado uma vantagem, mas, no entanto, isto também contribui para que, em algumas situações, torne limitada sua aplicabilidade.

MORETTIN e TOLOI (1981) atribuem como vantagens deste método a adequação às séries com comportamento mais geral, a flexibilidade em função da escolha das constantes de alisamento e ser de fácil entendimento e aplicação. São citadas como desvantagens para este método, a dificuldade de se determinar os melhores valores para as constantes de suavização, as dificuldades de estudo de propriedades estatísticas como média e a variância de previsão que dificultam a construção do intervalo de confiança para os valores previstos.

A determinação das constantes de suavização ( $A$ ,  $C$ ,  $D$ ) é realizada de modo a tornar mínimo o MSE.

### 3.2 MODELOS BOX-JENKINS

A filosofia da modelagem Box-Jenkins se utiliza de duas idéias: o princípio da parcimônia e a construção de modelos através de um ciclo iterativo. O princípio de parcimônia estabelece que, para uma adequada representação matemática, deve-se escolher um modelo com o menor número de parâmetros possíveis. Já o ciclo iterativo é uma estratégia de seleção de modelos a ser empreendida até que tenhamos um modelo satisfatório (SOUZA e CAMARGO, 2004).

Os modelos de Box-Jenkins, também conhecidos por ARIMA, são modelos que visam captar o comportamento da correlação seriada ou autocorrelação entre os valores da série temporal, e com base nesse comportamento realizar previsões futuras (KIRCHNER, 2006).

A metodologia consiste em ajustar modelos auto-regressivos integrados médias móveis aos dados disponíveis da série temporal. Esses modelos são resultantes da combinação de três componentes, chamados de filtros: o componente auto-regressivo (AR), o filtro de integração (I) e o componente de médias móveis (MA). O termo auto-regressivo corresponde a defasagem da série transformada, o termo média móvel refere-se às defasagens dos erros aleatórios e o termo integrado refere-se ao processo de diferenciar a série original para torná-la estacionária.

A idéia básica destes modelos é a de que os mesmos são compostos de uma parte sistemática e a outra consiste de ruído branco com distribuição normal, média zero e uma determinada variância (OLIVEIRA, 2002).

Um modelo não sazonal ARIMA é denotado por um modelo ARIMA( $p, d, q$ ) onde

- $p$  é o número de parâmetros auto-regressivos;
- $d$  é o número de diferenças necessárias para tornar a série estacionária;
- $q$  é o número de parâmetros médias móveis.

A estrutura geral ARIMA( $p, d, q$ ) é expressa pela equação (3.16).

$$\phi(B) \nabla^d Z_t = \theta(B) a_t, \quad (3.16)$$

onde

- $B$  é o operador de translação para o passado: representa uma defasagem de  $k$  períodos de tempo para trás e é definido pela equação (3.17).

$$B^k Z_t = Z_{t-k}. \quad (3.17)$$

- $\phi(B)$  é o polinômio característico auto-regressivo de ordem  $p$ , expresso por (3.18).

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p. \quad (3.18)$$

- $\theta(B)$  é o polinômio característico de médias móveis de ordem  $q$ , definido por (3.19).

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q. \quad (3.19)$$

- $\nabla^d$  é o operador diferença de ordem  $d$ : é uma transformação nos dados que consiste em tomar diferenças sucessivas da série original e é definido por (3.20) e (3.21).

$$Z_t - Z_{t-1} = Z_t - B Z_t = (1 - B) Z_t = \nabla Z_t; \quad (3.20)$$

$$\nabla^d = (1 - B)^d. \quad (3.21)$$

- $a_t$  é o ruído branco:

$$a_t \sim N(0, \sigma_a^2). \quad (3.22)$$

Pode-se ainda dizer que:

$$\omega_t = \nabla^d Z_t, \quad (3.23)$$

onde  $\omega_t$  é uma série estacionária.

Então uma forma mais explícita para esta série temporal pode ser dada pela equação (3.24).

$$\omega_t = \phi_1 \omega_{t-1} + \phi_2 \omega_{t-2} + \dots + \phi_p \omega_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}, \quad (3.24)$$

que é o modelo básico geral não sazonal utilizado, correspondente a equação (3.16).

A estratégia para a construção do modelo é baseada em um ciclo iterativo, no qual a escolha do modelo é baseada nos próprios dados. Os estágios do ciclo iterativo são (MORETTIN e TOLOI, 2006):

- Especificação: onde uma classe geral de modelos é considerada para análise;
- Identificação: onde uma estrutura de modelos é selecionada, com base na análise de autocorrelações, autocorrelações parciais e outros critérios. Identificam-se as ordens  $p$ ,  $d$  e  $q$  da estrutura. É uma das fases críticas do método;
- Estimação: fase em que os parâmetros do componente auto-regressivo, do componente de médias móveis e da variância, da estrutura identificada, são estimados;
- Verificação ou diagnóstico: onde através de uma análise de resíduos o modelo ajustado é considerado adequado ou não para os fins em vista;
- Previsão: é a última etapa da metodologia e o objetivo principal. É realizada somente quando a etapa anterior for satisfatória.

Caso o modelo não seja adequado, o ciclo é repetido, voltando-se à fase de identificação. Um procedimento que, muitas vezes, é utilizado é identificar não só um único modelo, mas alguns modelos que serão então estimados e verificados (MORETTIN e TOLOI, 2006).

O ciclo iterativo de um modelo ARIMA é ilustrado pelo fluxograma apresentado na figura 3.1.

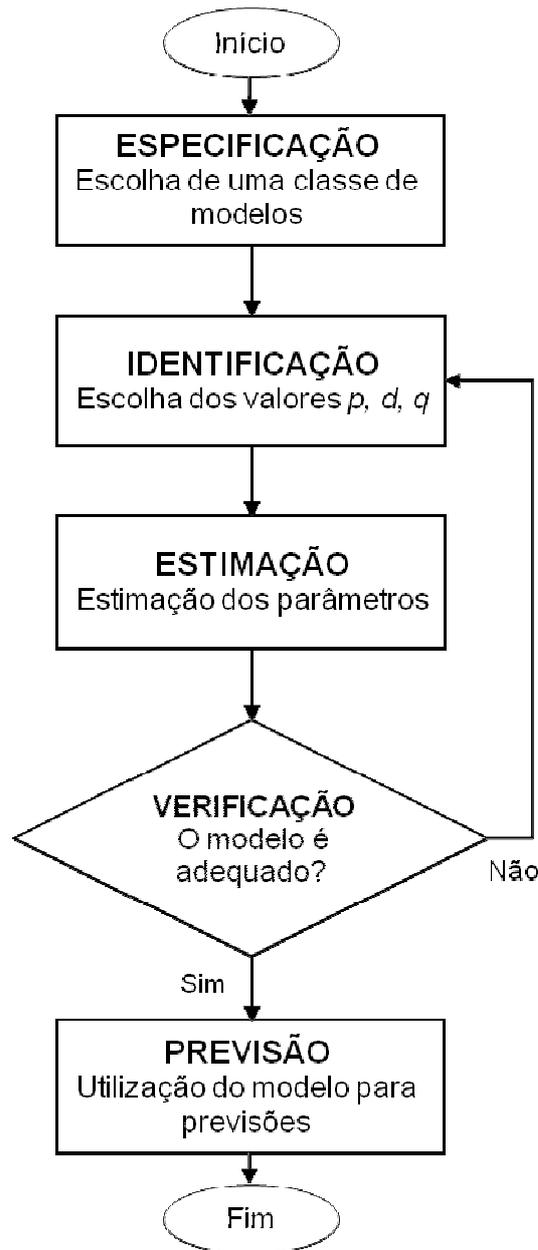


Figura 3.1 – Ciclo iterativo de um modelo ARIMA.

### 3.2.1 Identificação dos Modelos

O objetivo da identificação é determinar os valores de  $p$ ,  $d$  e  $q$  do modelo  $ARIMA(p,d,q)$ , além de estimativas preliminares dos parâmetros a serem usadas no estágio de estimação (MORETTIN e TOLOI, 2006).

A identificação dos modelos ARIMA consiste teoricamente de dois estágios que se podem dizer distintos (SOUZA e CAMARGO, 2004):

1ª Fase: Escolha adequada do grau de diferenciação  $d$ . Deve-se tomar diferenças da série tantas vezes quantas necessárias para se obter uma série estacionária. Na maioria dos casos usuais uma ou duas diferenciações ( $d=1$  e  $d=2$ ) são suficientes para tornar estacionárias séries com não estacionariedade homogênea (MORETTIN e TOLOI, 2006).

2ª Fase: Escolha adequada dos graus dos polinômios  $p$  e  $q$  do modelo ARMA, através da análise das autocorrelações e autocorrelações parciais estimadas, cujos comportamentos devem imitar os comportamentos das respectivas quantidades teóricas.

Em ambas as fases, a identificação da ordem do modelo mencionada ( $p, d, q$ ) se faz através das funções autocorrelação e autocorrelação parcial. Dada a forma complicada das funções autocorrelação e autocorrelação parcial de um modelo ARMA, estas funções não são úteis para identificar tais modelos. O que se recomenda, neste caso, é ajustar alguns modelos de baixa ordem e utilizar critérios que permitam escolher o modelo mais adequado (MORETTIN e TOLOI, 2006).

A etapa de identificação é a fase crítica da metodologia Box-Jenkins. É possível que vários pesquisadores identifiquem modelos diferentes para a mesma série temporal (MORETTIN e TOLOI, 2006).

Vários métodos alternativos de identificação têm sido sugeridos na literatura, entre eles pode-se citar os Métodos baseados em uma função penalizada: critério de informação de Akaike, critério de informação Bayesiano, critério de Hannan e Quinn e critério CAT (*Criterion Autoregressive Transfer Function*). Para determinar as ordens de um modelo ARMA( $p,q$ ) tem-se o procedimento de Anderson para AR( $p$ ), procedimento de Gray, Kelley e McIntire, procedimento de Nerlove, Grether e Carvalho, função de autocorrelação inversa e função de autocorrelação estendida (MORETTIN e TOLOI, 2006).

### 3.2.2 Estimação dos Parâmetros

Após a etapa de identificação da ordem do modelo que melhor representa a série observada, torna-se necessária a obtenção de estimativas para os parâmetros do modelo identificado.

O processo de estimação do modelo geral  $ARIMA(p,d,q)$ , da equação (3.16), significa a obtenção de  $p$  estimativas para os parâmetros  $\phi_1, \phi_2, \dots, \phi_p$  e  $q$  estimativas para os parâmetros  $\theta_1, \theta_2, \dots, \theta_q$ , além da variância do ruído,  $\sigma_a^2$ .

Um método que pode ser utilizado na estimação dos parâmetros do modelo é o método de máxima verossimilhança.

O princípio de verossimilhança sustenta que, se o modelo está corretamente identificado, todas as informações que os dados observados fornecem sobre os parâmetros estão contidas na função de verossimilhança, e os demais aspectos informativos das observações são irrelevantes.

A idéia básica do método de máxima verossimilhança é achar grandezas populacionais que geram os valores que mais se assemelham aos da amostra observada, ou seja, o método busca estabelecer os valores populacionais hipotéticos que maximizam a verossimilhança da amostra observada. O método consiste em selecionar aqueles estimadores que maximizam a probabilidade de obter-se a amostra realmente observada (SOUZA e CAMARGO, 2004).

No caso de mínimos quadrados, o processo de estimação de um modelo  $ARIMA(p,d,q)$ , requer a minimização de (3.25).

$$S(\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q) = \sum_{t=1}^n \hat{\varepsilon}_t^2, \quad (3.25)$$

onde  $\hat{\varepsilon}_t$  é descrito pela equação (3.26).

$$\hat{\varepsilon}_t = \theta^{-1}(B)\phi(B)\omega_t. \quad (3.26)$$

Neste contexto,  $S(\hat{\phi}, \hat{\theta})$  depende de valores  $\omega_t$  e de  $\varepsilon_t$ , ou seja, valores anteriores ao período amostrado. Para a estimação da máxima verossimilhança, é necessário admitir, inicialmente, que os ruídos brancos  $\varepsilon_t$  têm distribuição normal.

Como consequência,  $\omega_t$  também terá distribuição normal e a distribuição conjunta de  $\omega=(\omega_1, \dots, \omega_n)$  é:  $\omega \sim N(0; \sigma_\varepsilon^2 \Omega)$ .

MORETTIN e TOLOI (2006) destacam dois procedimentos para a obtenção dos ruídos brancos. O primeiro chamado de condicional, onde, partindo de suposições razoáveis do ponto de vista teórico, são atribuídos valores não observáveis; o segundo, incondicional, em que os valores iniciais são estimados utilizando um procedimento denominado *backforecasting*, onde os valores não observáveis são estimados a partir da amostra de dados.

### 3.2.3 Verificação

Depois de identificado e estimado um determinado modelo, este deve ser verificado quanto à adequação.

Um teste que pode ser utilizado é o teste da sobrefixação ou do superajustamento. Este teste consiste, basicamente, na elaboração de um modelo com um número de parâmetros superior ao do modelo fixado, que cubra as supostas direções de discrepâncias. Este modelo mais elaborado é submetido à análise que indicará a necessidade ou não de parâmetros adicionais. Com este teste tenta-se corrigir algum erro que porventura tenha sido cometido quando da identificação do modelo.

Diferentes testes de diagnóstico de um modelo ajustado a uma série podem ser aplicados. Estes testes são baseados, em geral, nas autocorrelações estimadas dos resíduos.

O teste de comparação das autocorrelações verifica a validade do modelo identificado pelos correlogramas através da comparação entre as autocorrelações estimadas a partir da série e as correspondentes autocorrelações obtidas pela função geradora das autocorrelações usando os parâmetros estimados.

A verificação, também, pode ser realizada pela análise de resíduos. Na análise de resíduos, seus coeficientes de autocorrelação devem ser estatisticamente, iguais a zero, isto é, comportar-se como ruído branco.

O teste proposto por Box-Jenkins (1970), conhecido como *Portmanteau Test*, assume que, se o modelo fixado é adequado, então é validada a estatística (3.27),

modificada por Ljung e Box em 1978, conforme equação (3.28). Esta estatística pode ser aproximada a uma distribuição qui-quadrado,  $\chi^2(M)$ , onde  $M = k - (p+q)$ .

$$Q = n \sum_{i=1}^k \hat{\rho}_i^2(\hat{a}), \quad (3.27)$$

$$Q = n(n+2) \sum_{i=1}^k \frac{\hat{\rho}_i^2(\hat{a})}{(n-k)}, \quad (3.28)$$

Assim, referindo-se ao valor calculado de  $Q$  na tabela de distribuição qui-quadrado com  $M$  graus de liberdade, pode-se testar a validade do modelo dentro de um determinado nível de significância.

O teste do Periodograma Acumulado consiste na comparação dos periodogramas acumulados da série dos resíduos estimados  $\hat{a}_t$  e de um ruído branco  $a_t$  para verificar-se a existência de componentes periódicas na série dos resíduos estimados  $\hat{a}_t$ .

Neste caso, a análise da série temporal é baseada na suposição de que ela é constituída de ondas senoidais e cosenoidais com diferentes freqüências. O periodograma da série temporal  $a_t$ ,  $t = 1, \dots, n$  é definido pela equação (3.29).

$$I_a(f_i) = \frac{2}{n} \left[ \left( \sum_{t=1}^n a_t \cos(2\pi f_i t) \right)^2 + \left( \sum_{t=1}^n a_t \sin(2\pi f_i t) \right)^2 \right], \quad (3.29)$$

onde a freqüência  $f_i = \frac{i}{n}$ ,  $i = 1, 2, \dots, q$  com  $q = \frac{n-1}{2}$  e  $I(f_i)$  é chamada intensidade da

freqüência  $f_i$ . Então, o periodograma possui  $q = \frac{n-1}{2}$  valores se  $n$  é ímpar e quando

$n$  é par tem-se  $i = 1, 2, \dots, q-1$  e  $I(f_i) = N a_q^2$ . Um pico na freqüência  $f_i = \frac{i}{n}$  indica uma

periodicidade de período  $\frac{i}{f_i}$ . A função espectral  $p(f)$  para o ruído branco tem um

valor constante igual a  $2\sigma_a^2$  no intervalo  $[0, 0.5]$ . Conseqüentemente, a função espectral acumulada para o ruído branco é dada pela equação (3.30).

$$P_a(f) = \int_0^f p(g) dg. \quad (3.30)$$

No caso do ruído branco  $a_t \sim N(0, \sigma_a^2)$  tem-se a equação (3.31).

$$P_a(f) = \int_0^f 2\sigma_a^2 dg = \begin{cases} 0, & f < 0 \\ 2\sigma_a^2 f & 0 \leq f < 1/2 \\ \sigma_a^2 & f \geq 1/2, \end{cases} \quad (3.31)$$

ou seja, para o ruído branco  $a_t$  o espectro acumulado de potência varia linearmente com a frequência, conforme ilustrado na figura 3.2.

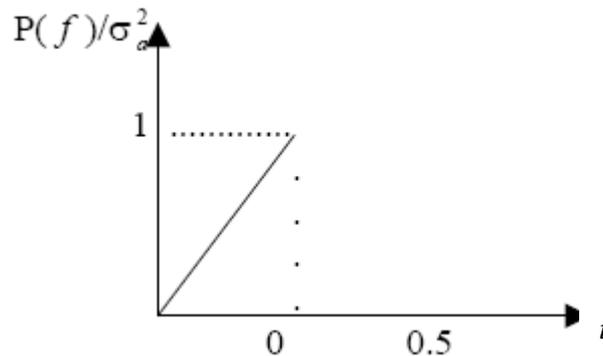


Figura 3.2 - Espectro acumulado do ruído branco

Assim, para a série dos resíduos  $\hat{a}_t$ , calcula-se:

- O periodograma da série  $I(f_i)$  com  $i = 1, 2, \dots, n/2$ ;
- O periodograma normalizado acumulado:

$$c(f_i) = \frac{1}{n\hat{\sigma}_a^2} \sum_{i=1}^j I(f_i) \quad (\text{estimador de } P(f) / \hat{\sigma}_a^2). \quad (3.32)$$

Então, quando a série  $\hat{a}_t$  se aproxima do ruído branco, a função  $c(f_i) \times f_i$  tem comportamento linear entre os pontos  $(0;0)$  e  $(0,5;1)$  e aceita-se o modelo fixado como válido. Ainda, como o periodograma acumulado está para o espectro teórico assim como a distribuição de frequências acumuladas observadas está para a experimental, tem-se que pode ser usado o teste de Kolmogorov-Smirnov para traçar linhas limites próximas da linha teórica e isto serve para indicar, com uma probabilidade determinada, a aproximação de  $\hat{a}_t$  para o ruído branco.

### 3.2.4 Previsão

Após o término do processo iterativo de identificação, estimação e checagem do diagnóstico, o qual proporciona uma estimativa do modelo gerador da série que se ajusta aos dados do problema em questão, pode-se utilizá-lo para prever os valores futuros da variável.

Os processos de previsão usando modelos de séries temporais são procedimentos que visam estender a valores futuros o modelo descrito e ajustado aos valores passados e ao valor presente da variável (SOUZA e CAMARGO, 2004).

Qualquer modelo ARIMA pode ser escrito nas seguintes formas para previsão: forma de equações a diferenças, forma de choques aleatórios e forma invertida. O mais conveniente modo de se fazer previsões como modelos ARIMA é escrever o modelo na forma de equações a diferenças.

A previsão de origem  $t$  e horizonte  $h$  é denotada por  $\hat{Z}_t(h)$ . Admitindo que  $Z_t$  seja estacionária e que  $\mu$  seja a média do processo uma forma para esta previsão é dada pela expressão (3.33).

$$\hat{Z}_t(h) = \phi_1 Z_{t+h-1} + \phi_2 Z_{t+h-2} + \dots + \phi_p Z_{t+h-p} + a_t - \theta_1 a_{t+h-1} - \dots - \theta_q a_{t+h-q} + \theta_0, \quad (3.33)$$

onde  $\theta_0$  é um termo constante, dado por (3.34).

$$\theta_0 = \mu (1 - \phi_1 - \phi_2 - \dots - \phi_p). \quad (3.34)$$

São considerados nulos os  $a_t$ 's para instantes a frente da origem  $t$ , o que é lógico pois  $E(a_t) = 0$  (ruído branco) e se referem a períodos futuros dos quais não é possível saber seu valor.

Se a série original não for estacionária a relação anterior serve para uma de suas diferenças  $\omega_t$  ( $\nabla Z_t$  ou  $\nabla^2 Z_t$ ) e a previsão é realizada com base nas equações a diferenças.

## 3.3 REDES NEURAIS ARTIFICIAIS

A pesquisa relacionada a RNAs tem sido motivada desde os seus primórdios pelo reconhecimento de que o cérebro humano processa informações de uma forma

diferente do computador digital convencional. O cérebro pode ser visto por um sistema de processamento de informação complexo, não-linear e paralelo. Ele tem a capacidade de organizar seus neurônios de forma a realizar certos processamentos mais rapidamente que o mais rápido computador digital hoje existente (HAYKIN, 2001).

O sistema nervoso humano pode ser visto como um sistema de três estágios, conforme descrito no diagrama mostrado na Figura 3.3. O centro do sistema é o cérebro, representado pela rede neural, o qual recebe continuamente informações, as compreende e toma decisões. As setas apontando da esquerda à direita indicam transmissão progressiva de sinais de informação externos e as setas apontando da direita para a esquerda indicam a presença de realimentação no sistema. Os receptores convertem o estímulo vindo do corpo humano ou do ambiente externo em impulsos elétricos que conduzem informação para a rede neural, ou seja, o cérebro. Os atuadores convertem os impulsos elétricos gerados pela rede neural em respostas discerníveis como saídas do sistema.

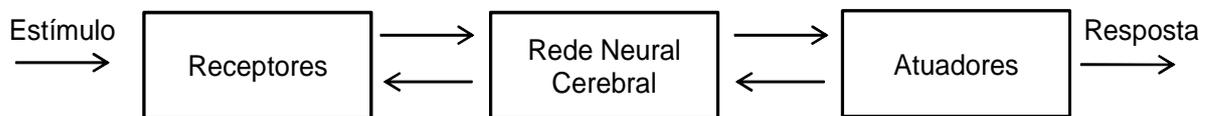


Figura 3.3 – Representação do sistema nervoso

Da mesma maneira que no cérebro, as RNAs são organizadas na forma de um número de elementos individuais simples, denominados neurônios ou nós, que se interconectam uns aos outros, formando redes capazes de armazenar e transmitir informação provinda do meio externo. Tais unidades são dispostas em uma ou mais camadas, interligadas por conexões. Na maioria dos modelos, estas conexões estão associadas a pesos, os quais armazenam o conhecimento representado no modelo e servem para ponderar a entrada recebida por cada neurônio da rede (GUERRA, 2006).

A RNA pode ser descrita como um grafo no qual cada nó  $i$  executa uma função de transferência  $f_i$  na forma da equação (3.35).

$$y_i = f_i \left( \sum_{j=1}^n w_{ij} x_j - \theta_i \right), \quad (3.35)$$

onde  $y_i$  é a saída do nó  $i$ ,  $x_j$  é a  $j$ -ésima entrada para o nó, e  $w_{ij}$  é o peso da conexão entre os nós  $i$  e  $j$ , e  $\theta_i$  é a polarização (*bias*) do nó. Usualmente,  $f_i$  é uma função não-linear, tais como função degrau, sigmóide ou gaussiana.

As RNAs são métodos computacionais que possuem a capacidade de auto-adaptação e aprendizado de uma certa tarefa ou comportamento, a partir de um conjunto de exemplos dados (SOUZA, 2008).

Uma capacidade importante das RNAs é a auto-organização ou plasticidade, ou seja, através de um processo de aprendizagem é possível alterar os padrões de interconexão entre seus elementos. Por este motivo, as RNAs são um tipo de sistema conexionista, no qual as propriedades computacionais são resultado dos padrões de interconexão da rede, como acontece também no sistema nervoso biológico (ROCHA, 2006).

Uma das mais importantes características das RNAs é sua capacidade de aprendizado. Os algoritmos responsáveis por modificar os pesos sinápticos e os parâmetros adaptativos das funções representativas do corpo dos neurônios, nos quais vão se codificar o aprendizado são denominados por algoritmos de aprendizagem. Entre os algoritmos existentes, o mais comum é o Retropropagação do Erro (*error backpropagation*). Neste algoritmo, existe a figura do “professor” ou supervisor, que informa qual deve ser a saída para uma determinada entrada. Assim, encontra-se o erro entre a saída desejada e a saída da rede e ajustam-se os pesos sinápticos e os parâmetros das funções para que esse erro seja o menor possível.

O uso de RNAs oferece as seguintes propriedades úteis e capacidades (HAYKIN, 2001):

- Não-linearidade: um neurônio artificial pode ser linear ou não-linear. Uma RNA constituída por conexões de neurônios não-lineares é ela mesma não-linear. É importante observar que a não-linearidade de uma RNA é distribuída por toda a rede. A não-linearidade é uma propriedade importante, particularmente se o mecanismo físico subjacente responsável pela geração do sinal de entrada for inerentemente não-linear.
- Mapeamento de entrada-saída: apresenta-se para a RNA um exemplo escolhido ao acaso do conjunto e os pesos sinápticos (parâmetros livres) da rede são modificados para minimizar a diferença entre a resposta desejada e

a resposta real da rede, produzida pelo sinal de entrada, de acordo com algum critério estatístico apropriado. O treinamento da RNA é repetido para exemplos do conjunto de treino até que a rede atinja um estado onde não haja mais mudanças significativas nos pesos sinápticos ou mesmo um limite do número de épocas de treinamento. Os mesmos exemplos do conjunto de treino podem ser reaplicados durante o processo de treinamento da rede, desde que em outra ordem de apresentação. Assim, a RNA aprende dos exemplos ao construir um mapeamento de entrada-saída para o problema considerado.

- **Adaptabilidade:** as RNAs têm uma capacidade inata de adaptar seus pesos sinápticos a modificações do meio ambiente em que está inserida. Uma RNA treinada para operar em um ambiente específico pode ser facilmente retreinada para tratar com pequenas modificações nas condições operacionais do ambiente. Quando operando em um ambiente não-estacionário, uma RNA pode ser projetada para modificar seus pesos sinápticos em tempo real.
- **Resposta a evidências:** no contexto de classificação de padrões, uma RNA pode ser projetada para fornecer informação não somente sobre qual padrão particular selecionar, mas também sobre a confiança ou crença na decisão tomada.
- **Informação contextual:** o conhecimento é representado pela própria estrutura e estado de ativação de uma RNA. Cada neurônio é potencialmente afetado pela atividade de todos os outros neurônios na rede. Conseqüentemente, a informação contextual é tratada naturalmente pela RNA.
- **Tolerância a falhas:** característica que permite à rede continuar a apresentar resultados aceitáveis no caso de falha de alguns neurônios. Se um neurônio ou suas conexões são danificados, por exemplo, a recuperação de um padrão armazenado é prejudicada em qualidade. Contudo, devido à natureza distribuída da informação armazenada na rede, o dano deve ser extenso para que a resposta global da rede seja degradada seriamente. Assim, uma RNA exhibe uma degradação suave do desempenho em vez de apresentar uma falha catastrófica.

- Implementação em VLSI (*Very-Large-Scale-Integrated*): esta característica permite considerar elevado grau de paralelismo no projeto da rede. A natureza fortemente paralela das RNAs as tornam potencialmente rápidas para computar determinadas tarefas. Esta mesma característica possibilita que sejam implementadas usando VLSI.
- Uniformidade de análise e projeto: as RNAs desfrutam de universalidade como processadores de informação, pois a mesma notação é utilizada em todos os domínios envolvendo aplicações de RNAs.
- Analogia neurobiológica: o projeto de uma RNA é motivado pela analogia com o cérebro, que é uma prova viva de que o processamento paralelo tolerante a falhas é não somente possível fisicamente, mas também rápido e poderoso.

As RNAs representam a procura constante do ser humano pela compreensão de sua existência e pela representação do comportamento humano através de modelos. Desde a mitologia grega até os dias de hoje, esta busca pela reprodução do comportamento e da mente humana têm sido uma obsessão do ser humano.

### 3.3.1 Modelo de um Neurônio

Um neurônio é uma unidade de processamento de informação que é fundamental para a operação de uma rede neural. Os neurônios possuem propensão natural para armazenar conhecimento experimental e torná-lo disponível para o uso. O diagrama da figura 3.4 apresenta o modelo básico de um neurônio, que forma a base de um projeto de RNA.

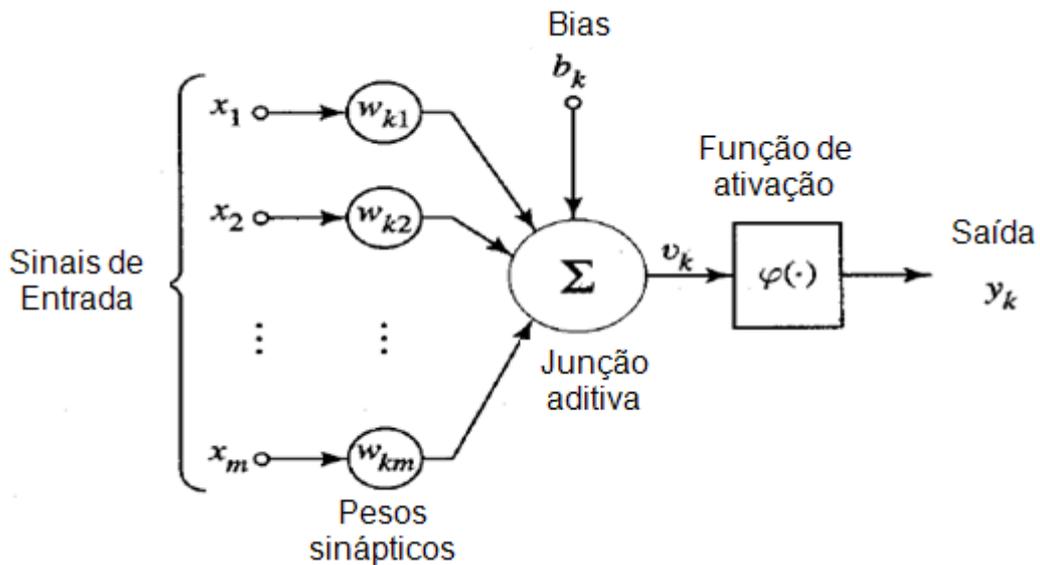


Figura 3.4 – Modelo não-linear de um neurônio.

O modelo neural consiste de três elementos básicos (HAYKIN, 2001):

- Um conjunto de sinapses ou elos de conexão, cada uma caracterizada por um peso ou força própria. Especificamente, um sinal  $x_j$  na entrada da sinapse  $j$  conectada ao neurônio  $k$  é multiplicado pelo peso sináptico  $w_{kj}$ . Ao contrário de uma sinapse do cérebro, o peso sináptico de um neurônio artificial pode estar em um intervalo que inclui valores negativos bem como positivos.
- Um somador (combinador linear) para somar os sinais de entrada, ponderados pelas respectivas sinapses do neurônio.
- Uma função de ativação para restringir a amplitude da saída de um neurônio. A função de ativação limita a faixa de amplitude permitida do sinal de saída a um valor finito. Tipicamente, o intervalo normalizado da amplitude da saída de um neurônio é escrito como o intervalo fechado  $[0,1]$  ou alternativamente  $[-1,1]$ .

O modelo neural da figura 3.4 inclui também um *bias* (polarização) aplicado externamente, denotado por  $b_k$ . O bias  $b_k$  tem o efeito de aumentar ou diminuir o argumento da função de ativação, caso seja positivo ou negativo, respectivamente.

Em termos matemáticos, um neurônio  $k$  pode ser descrito pelas equações (3.36) e (3.37).

$$u_k = \sum_{j=1}^m \omega_{kj} x_j \quad (3.36)$$

e

$$y_k = \varphi(v_k + b_k), \quad (3.37)$$

onde  $x_1, x_2, \dots, x_m$  são os sinais de entrada;  $w_{k1}, w_{k2}, \dots, w_{km}$  são os pesos sinápticos do neurônio  $k$ ;  $v_k$  é a saída do combinador linear devido aos sinais de entrada;  $b_k$  é o bias;  $\varphi(\cdot)$  é a função de ativação e  $y_k$  é o sinal de saída do neurônio.

O uso de *bias* tem o efeito de aplicar uma transformação à saída  $u_k$  do combinador linear, tal que:

$$v_k = u_k + b_k. \quad (3.38)$$

A função de ativação  $\varphi(v)$  define a saída de um neurônio em termos do potencial de ativação  $v$ . Identificam-se três tipos básicos de funções de ativação dentre outros: função de limiar, função linear por partes e função sigmóide (HAYKIN, 2001).

### 3.3.2 Arquiteturas de Redes Neurais

O projeto de uma RNA, ou seja, a maneira pela qual os neurônios da rede são estruturados está intimamente relacionada ao algoritmo de aprendizagem usado para treinar a rede. Em geral, pode-se identificar três diferentes classes fundamentais de arquiteturas de RNA (HAYKIN, 2001), mencionadas a seguir.

#### **Redes Alimentadas Adiante com Camada Única (*Perceptron* de camada única)**

Em uma RNA em camadas, os neurônios estão organizados na forma de camadas. Na forma mais simples de uma rede em camadas, tem-se uma camada de entrada de neurônios fonte que se projeta sobre uma camada de saída de neurônios, conforme mostrado na figura 3.5. Esta RNA é estritamente do tipo alimentada adiante ou acíclica. Na figura 3.5 tem-se uma RNA de quatro neurônios tanto na camada de entrada como na de saída. Esta RNA é chamada de rede de camada única, sendo que a designação camada única se refere à camada de saída de neurônios computacionais.

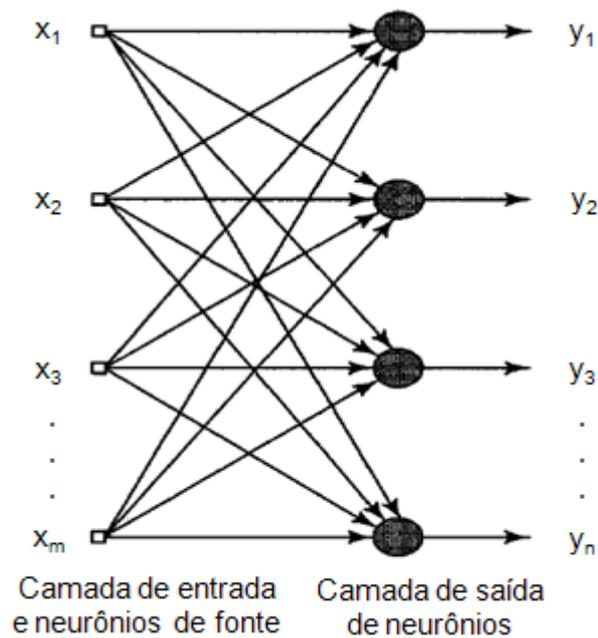


Figura 3.5 – Rede alimentada adiante com uma única camada de neurônios.

### **Redes Alimentadas Diretamente com Múltiplas Camadas (Rede Neural *Perceptron* Multicamadas)**

A segunda classe de uma RNA alimentada adiante tem por característica possuir uma ou mais camadas escondidas, cujos nós computacionais ou neurônios são correspondentemente denominados neurônios ocultos ou unidades ocultas. A função dos neurônios na camada oculta é intervir entre as camadas de entrada e a de saída da rede de uma maneira útil. Adicionando uma ou mais camadas ocultas, a rede pode extrair estatísticas de ordem superior. Pode-se dizer que a rede adquire uma perspectiva global apesar de sua conectividade local, devido ao conjunto extra de conexões sinápticas e à dimensão extra de interações neurais. A habilidade dos neurônios ocultos extraírem estatísticas de ordem superior é particularmente valiosa quando o tamanho da camada de entrada é grande.

Os neurônios fonte da camada de entrada da rede fornecem os vetores de entrada que constituem os sinais de entrada aplicados aos neurônios na segunda camada (primeira camada oculta). Os sinais de saída da segunda camada são utilizados como entradas para a terceira camada, e assim por diante para o resto da rede. Tipicamente, os neurônios em cada camada da rede têm como suas entradas apenas os sinais de saída da camada precedente. O conjunto de sinais de saída dos

neurônios da camada de saída da rede constitui a resposta global da rede para o padrão de ativação fornecido pelos nós de fonte da camada de entrada.

A figura 3.6 ilustra uma rede neural de múltiplas camadas, para o caso de uma única camada oculta, em que cada neurônio de cada camada da rede é conectado a cada outro neurônio da camada adjacente seguinte. Neste caso, a rede é dita completamente conectada. Entretanto, se algumas das conexões sinápticas estiverem faltando, a rede é dita parcialmente conectada.

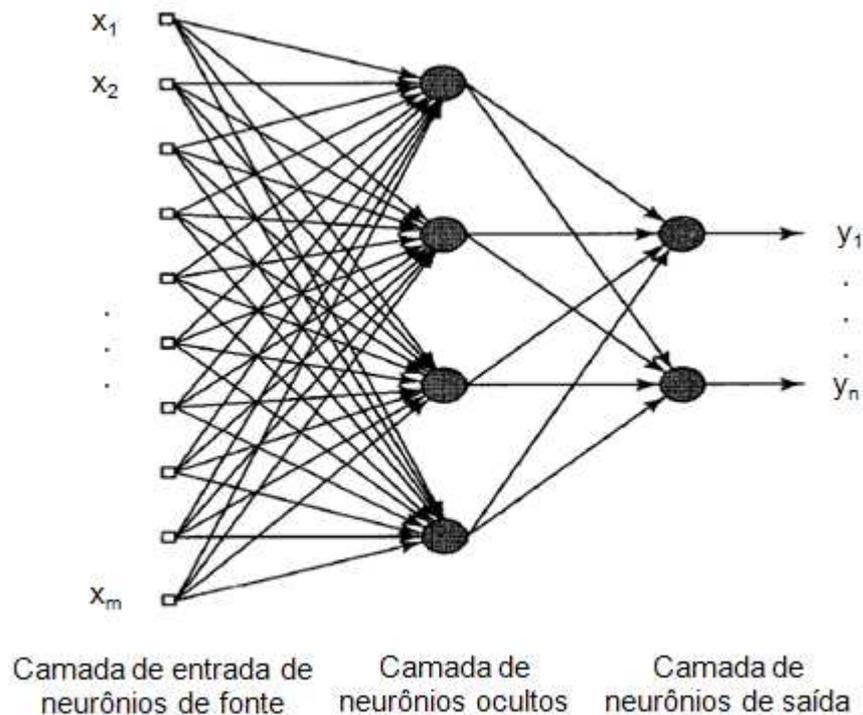


Figura 3.6 – Rede alimentada adiante totalmente conectada com uma camada oculta e uma camada de saída.

Uma RNA perceptron multicamadas apresenta três características distintas:

- O modelo de cada neurônio da rede inclui uma função de ativação não-linear. É importante salientar que esta não-linearidade é suave, ou seja, a função é diferenciável em qualquer ponto. Uma forma normalmente utilizada de não-linearidade que satisfaz este requisito é a não-linearidade sigmoideal definida pela função logística (3.39).

$$y_j = \frac{1}{1 + e^{-v_j}} \quad (3.39)$$

onde  $v_j$  é o potencial de ativação do neurônio  $j$  e  $y_j$  é a saída do neurônio.

- A rede contém uma ou mais camadas de neurônios ocultos, que não são parte da entrada ou da saída da rede. Estes neurônios ocultos possibilitam que a rede aprenda tarefas complexas, extraindo progressivamente mais características significativas dos padrões de entrada.

- A rede exibe um alto grau de conectividade, determinado pelas sinapses da rede. Uma mudança na conectividade da rede requer uma mudança no conjunto das conexões sinápticas ou de seus pesos.

É através da combinação destas características, juntamente com a habilidade de aprender da experiência através de treinamento, que o *perceptron* multicamadas deriva seu poder computacional.

### Redes Recorrentes

Uma rede neural recorrente se distingue de uma rede neural alimentada adiante por ter pelo menos um laço de realimentação.

Por exemplo, uma rede recorrente pode consistir de uma única camada de neurônios, em que cada neurônio alimenta seu sinal de saída de volta para as entradas de todos os outros neurônios, conforme ilustra a figura 3.7.

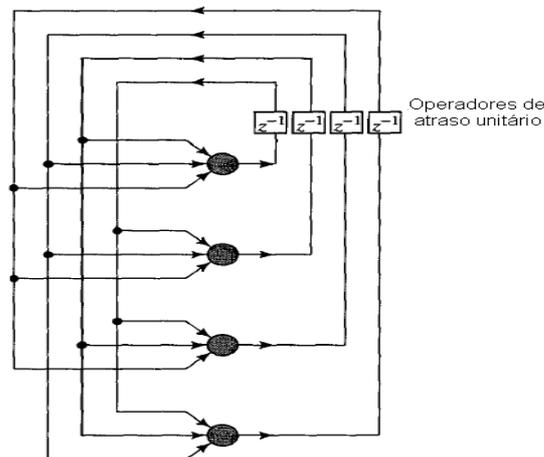


Figura 3.7 – Rede recorrente.

### 3.4 SISTEMAS NEBULOSOS

Um sistema nebuloso, também conhecido como sistema *fuzzy*, pode ser definido como uma abordagem com objetivo de modelar o modo aproximado do raciocínio, imitando a habilidade humana de tomar decisões em um ambiente de

incerteza e imprecisão. Esta lógica permite que os sistemas inteligentes de controle e suporte a decisão tratem informações imprecisas ou nebulosas, ou seja, permite capturar informações imprecisas, descritas em linguagem natural e convertê-las para um formato numérico (CAMPOS, JESUS, MENDES, 2007).

A teoria dos sistemas nebulosos nasceu da constatação de que quando a complexidade de um sistema aumenta a habilidade humana para concluir fatos e tomar decisões que sejam ao mesmo tempo precisos e significativos, tende a diminuir até um limite a partir do qual precisão e relevância passam a ser características quase excludentes.

O conceito de conjuntos nebulosos e sistemas nebulosos permitem representar o conhecimento e a experiência de seres humanos através de uma abordagem lingüística. Esta abordagem é um mecanismo matemático natural para emular certos atributos lingüísticos e de percepção associado ao pensamento e ao raciocínio presentes no processo cognitivo humano. Os seres humanos estão acostumados a tomar decisões, onde a resposta para determinado caso não seja totalmente verdadeira. Os sistemas nebulosos derivam desta capacidade do ser humano, de inferir conclusões e gerar respostas através de informações vagas, ambíguas e qualitativamente incompletas e imprecisas.

Inseridos em ambientes computacionais, os sistemas nebulosos podem ser entendidos como programas de computador que tentam modelar tarefas desempenhadas pelo ser humano que tendem a empregar as capacidades típicas do raciocínio. Estes sistemas devem estabelecer metas em função do reconhecimento da situação e da utilização de um conhecimento adquirido. A realização destas tarefas se dá através da tentativa de imitar sua habilidade de adquirir e aplicar o conhecimento, perceber e manipular objetos, aprender através de exemplos e a mais fundamental e inata capacidade do ser humano que é sua habilidade de raciocinar e oferecer respostas apropriadas para solução de problema de forma que elas não sejam pré-programadas e que sejam geradas para situações novas e inesperadas.

A teoria nebulosa surgiu com a publicação do artigo intitulado "*Fuzzy Sets*" por Zadeh (1965). Neste trabalho, ele introduziu o conceito de conjunto nebuloso, cujas fronteiras não são bem demarcadas ou precisas, em contraste com os conjuntos tradicionais, nitidamente definidos e que possuem fronteiras precisas (ALLEMÃO, 2004). A teoria dos conjuntos *fuzzy* possui como principal característica

o uso de regras do tipo SE-ENTÃO nas quais o conhecimento pode ser representado. Tradicionalmente, uma proposição lógica tem dois extremos: ou “completamente verdadeiro” ou “completamente falso”. Entretanto, na lógica nebulosa, uma premissa varia em grau de verdade de 0 a 1, o que leva a ser parcialmente verdadeira ou parcialmente falsa.

Um conjunto nebuloso é caracterizado por uma função de pertinência que mapeia os elementos do domínio  $X$  no intervalo  $[0,1]$ . Isto é,

$$A : X \rightarrow [0,1]. \quad (3.40)$$

Utilizando uma conversão algébrica, o conjunto nebuloso  $A$  em  $X$  pode ser representado como um conjunto de pares ordenados de um elemento genérico  $x \in X$ , dado por (3.41).

$$A = \{(x, \mu_A(x)) \mid x \in X\}, \quad (3.41)$$

onde  $\mu_A(x)$  representa o grau de  $x$  em  $A$ , ou seja, é uma função de pertinência que determina com que grau um objeto  $x$  pertence a um conjunto  $A$  e  $X$  é chamado de universo.

As funções de pertinência representam os aspectos fundamentais de todas as ações teóricas e práticas de sistemas *fuzzy*. Uma função de pertinência é uma função numérica, gráfica ou tabulada que atribuem valores de pertinência *fuzzy* para valores discretos de uma variável em seu universo de discurso (FERREIRA, 2008).

Uma função de pertinência pode ser expressa por diversas equações. Nas formas mais comuns são equações triangulares, trapezoidais, gaussianas, entre outras. Neste estudo serão aplicadas as funções de pertinências gaussianas, forma de sino e triangular.

- Função triangular

A função triangular é definida da seguinte forma:

$$\mu_A(x) = \begin{cases} 0, & \text{se } x \leq a \\ \frac{x-a}{m-a}, & \text{se } x \in [a,m] \\ \frac{b-x}{b-m}, & \text{se } x \in [m,b] \\ 0, & \text{se } x \geq b, \end{cases} \quad (3.42)$$

onde  $m$  é um valor modal e  $a$  e  $b$  são os limites inferior e superior, respectivamente.

A figura 3.8 ilustra um exemplo de função triangular.

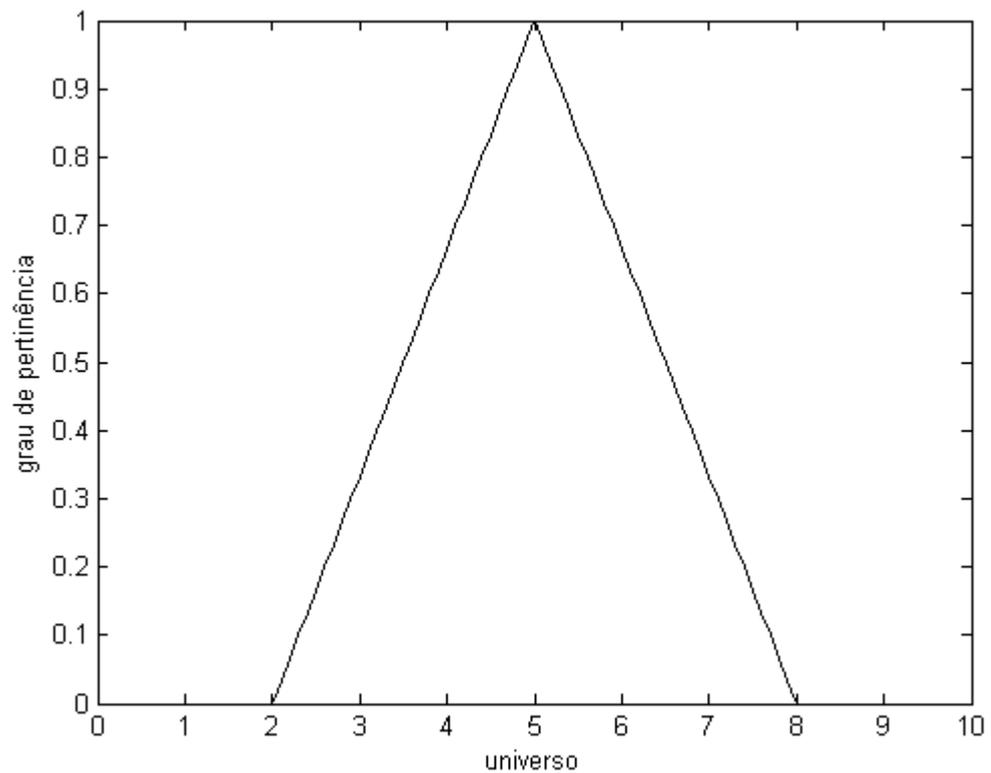


Figura 3.8 – Exemplo de Função de pertinência triangular

- Função Gaussiana

A função Gaussiana é definida por (3.43).

$$\mu_A(x) = \exp^{-\sigma_k(x-m)^2}, \text{ com } \sigma_k > 0. \quad (3.43)$$

A figura 3.9 ilustra a função Gaussiana, especificando os parâmetros associados.

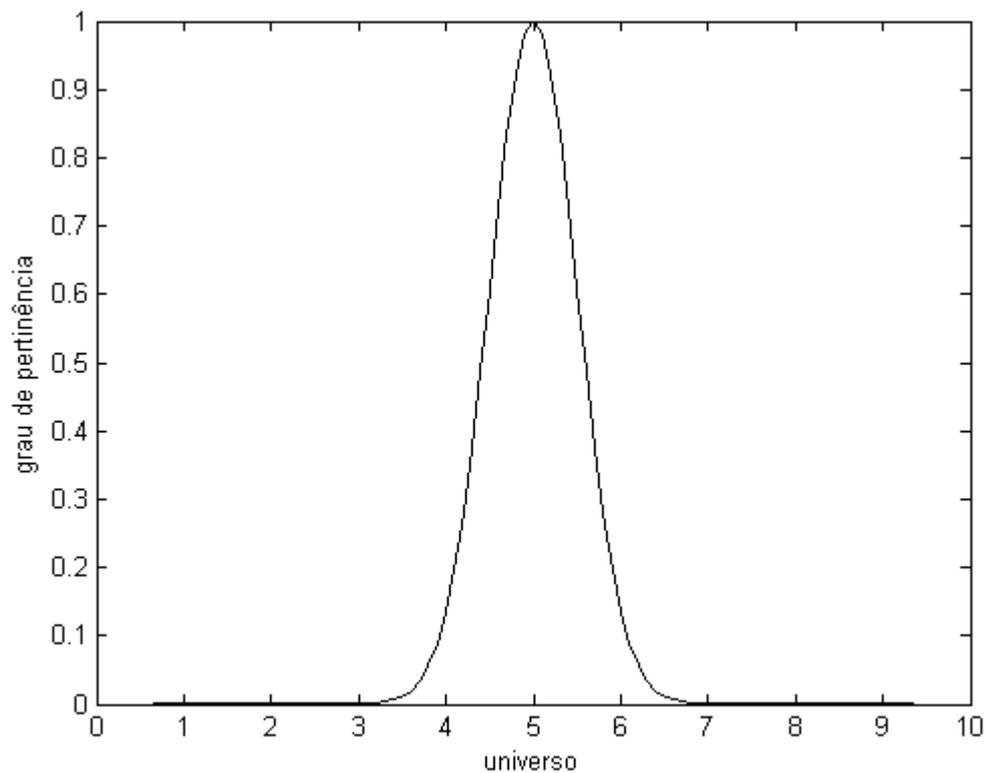


Figura 3.9 – Exemplo de Função de pertinência Gaussiana

- Função em forma de Sino (*Bell shaped*)

A função sino é definida pela equação (3.44).

$$\mu_A(x) = \frac{1}{1 + \left| \frac{x-c}{a} \right|^{2b}}, \quad (3.44)$$

e é especificada por três parâmetros  $\{a,b,c\}$ , sendo  $c$  o centro da curva e  $b$  usualmente um parâmetro com valor positivo. Na figura 3.10 observa-se um exemplo de função sino.

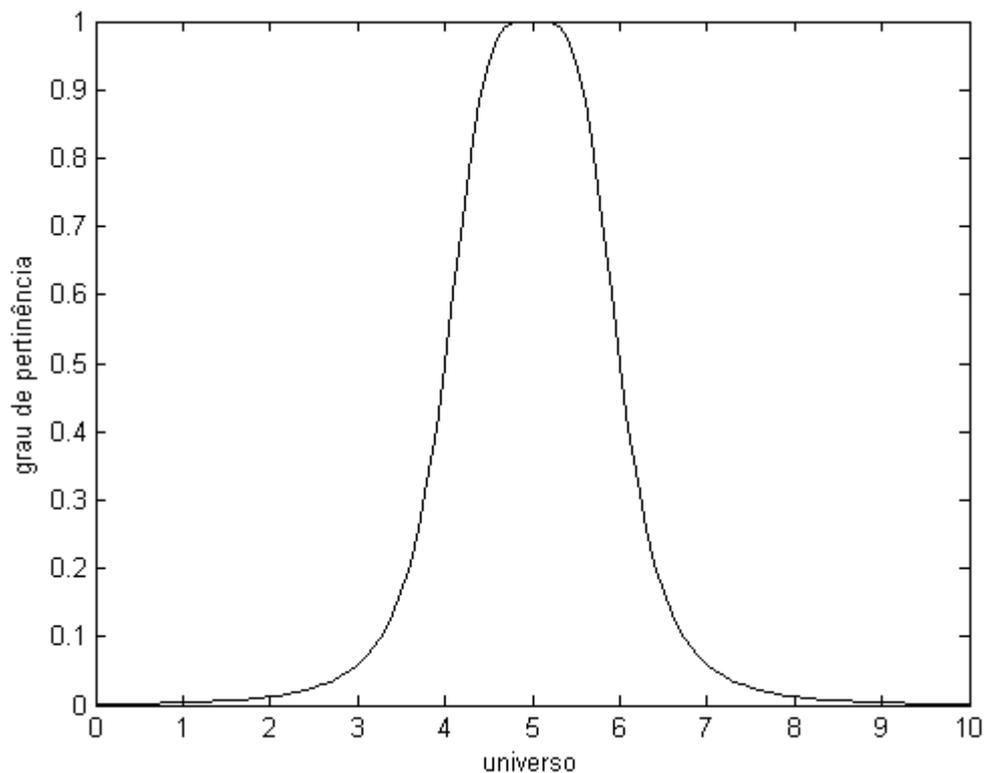


Figura 3.10 – Exemplo de Função de pertinência sino

### 3.4.1 Sistema de Inferência Nebuloso

Um sistema de Inferência nebuloso, em inglês *Fuzzy Inference Systems* (FIS), é uma abordagem que trabalha com conjuntos *fuzzy* e permite a implementação de aplicações com relações de conjuntos e regras *fuzzy* SE-ENTÃO.

A estrutura básica de um sistema nebuloso possui três componentes conceituais: uma base de regras que contém o conjunto de regras nebulosas; uma base de dados que define as funções de pertinências usadas nas regras nebulosas; e um mecanismo de raciocínio que realiza um procedimento de inferência, raciocínio nebuloso, para obter a saída ou conclusão baseado nas regras e fatos conhecidos (FERREIRA, 2008).

Na figura 3.11 está ilustrada a estrutura básica de um sistema de inferência nebuloso constituído por um módulo de nebulização que converte os valores de entrada do sistema para termos lingüísticos representados por conjuntos nebulosos; uma base de conhecimento onde todo o conhecimento sobre o domínio do problema em questão é armazenado; uma unidade tomadora de decisão que é responsável pelo desenvolvimento do raciocínio nebuloso baseado no conhecimento

representado na base de conhecimento; e por um módulo de desnebulização que realiza uma transformação da resposta do sistema nebuloso a qual está representada por um conjunto nebuloso, em uma resposta não nebulosa. A base de conhecimento é formada por uma base de regras que contém o conjunto de regras nebulosas e por uma base de dados que define as funções de pertinência usadas nas regras nebulosas.

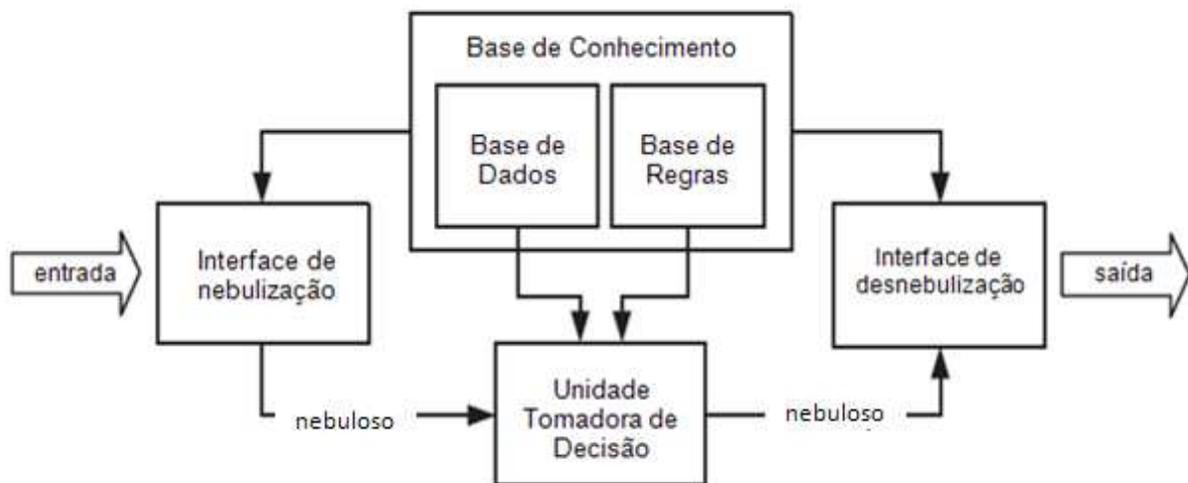


Figura 3.11 – Sistema de inferência nebuloso

A base de regras contém um conjunto de regras nebulosas onde as variáveis dos antecedentes e dos conseqüentes são variáveis lingüísticas e os possíveis valores de uma variável lingüística são representados por conjuntos nebulosos. Uma proposição nebulosa expressa relações entre variáveis lingüísticas e conjuntos nebulosos. As regras podem ser fornecidas por especialistas, em forma de sentenças lingüísticas e se constituem em um aspecto fundamental no desempenho de um sistema de inferência nebuloso. Na base de dados definem-se as funções de pertinência do conjunto nebuloso nas regras nebulosas.

A unidade tomadora de decisão realiza operações de inferência, para obter, a partir da avaliação dos níveis de compatibilidade das entradas com as condições impostas pela base de regras, uma ação a ser realizada pelo sistema.

No sistema de inferência nebuloso consideram-se entradas não-nebulosas resultantes de medições ou observações, assim é necessário efetuar um mapeamento dos dados para os conjuntos nebulosos (de entrada) relevantes, o que é realizado no estágio de nebulização. Neste estágio ocorre também a ativação das

regras relevantes para uma dada situação. O processo de nebulização utilizando as funções de pertinência pré-estabelecidas mapeia cada variável de entrada do sistema em graus de pertinência de algum conjunto nebuloso que representa a variável em questão.

Uma vez obtido o conjunto nebuloso de saída, no estágio de desnebulização é efetuada uma interpretação dessa informação. A desnebulização transforma os resultados nebulosos da inferência em valores de saída. Calcula a saída com base na inferência obtida na unidade tomadora de decisão, com as funções de pertinência das variáveis lingüísticas da parte conseqüente das regras para obter uma saída não nebulosa.

Existem basicamente três implementações de modelos FIS: o modelo Mamdani, o Takagi-Sugeno e o Tsukamoto. Basicamente, estes modelos diferem entre si na forma de operar os parâmetros conseqüentes e na forma de realizar o processo de desnebulização. Neste estudo é aplicado o modelo nebuloso Takagi-Sugeno também conhecido por Sugeno. Este modelo utiliza uma função de mapeamento para cada saída de regra nebulosa SE-ENTÃO e mapeia a entrada e a saída da regra por uma combinação linear das entradas. O processo de desnebulização é realizado através da média ponderada da pertinência de saída. A figura 3.12 ilustra o processo do modelo de Takagi-Sugeno (SANDMANN, 2006).

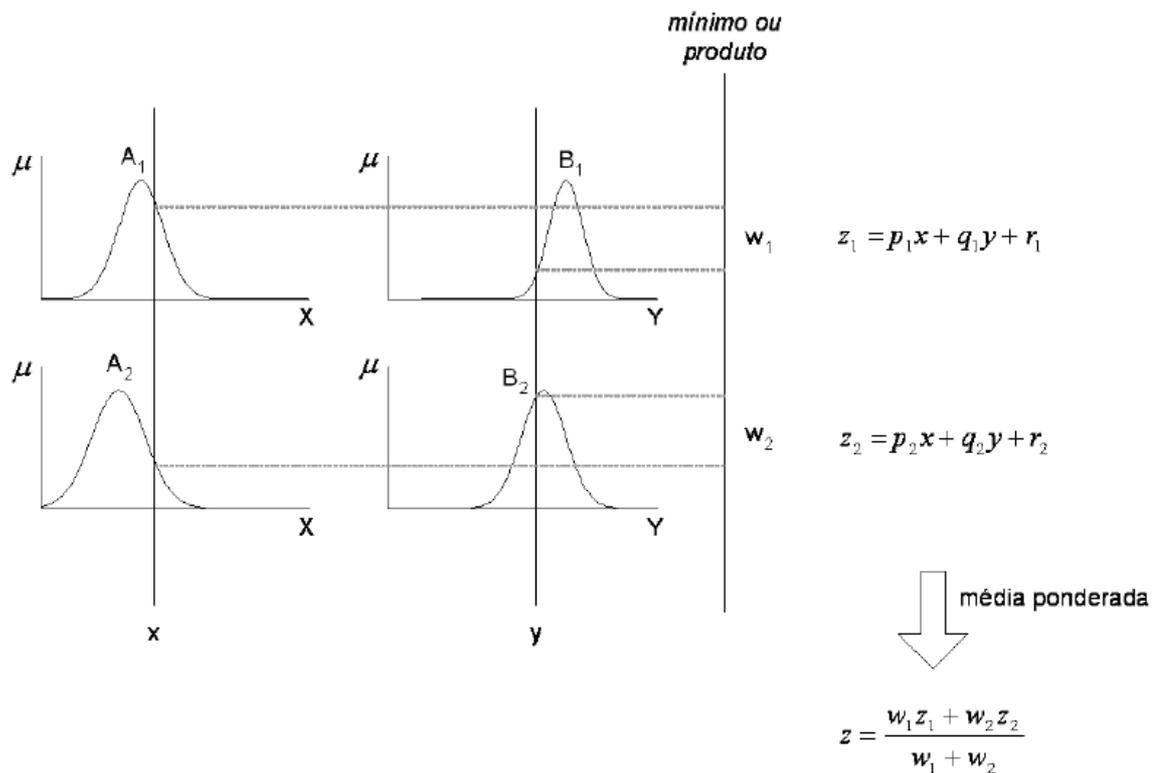


Figura 3.12 – Modelo nebuloso de Takagi-Sugeno

### 3.5 SISTEMA NEURO-NEBULOSO DO TIPO ANFIS

Cerca de trinta anos após a introdução da teoria dos conjuntos nebulosos, o pesquisador Jyn-Shing Roger Jang publicou um artigo no qual os parâmetros nebulosos são calculados através da técnica de retropropagação do erro (*Backpropagation error*), vastamente utilizada para o ajuste dos pesos sinápticos nas RNAs. Esta técnica de associar sistema nebuloso com redes neurais artificiais ficou conhecida como neuro-nebuloso e o modelo implementado foi denominado *Adaptive-Network-Based Fuzzy Inference System*, ou ANFIS (RODRIGUES, ARAÚJO, MAITELLI, 2006).

Um sistema neuro-nebuloso é uma ferramenta que está sendo cada vez mais utilizada para a solução de problemas. Sendo este a fusão de duas ferramentas já conhecidas: redes neurais artificiais e a lógica nebulosa, no qual agregam-se as características de transparência de raciocínio da lógica nebulosa juntamente com a capacidade de aprendizado e generalização das redes neurais artificiais.

Uma rede neuro-nebulosa pode ser definida como um sistema nebuloso treinado por um algoritmo provido de uma RNA. A união da RNA com sistemas

nebulosos tem o intuito de amenizar a deficiência de cada um destes sistemas fazendo com que tenha-se um sistema mais eficiente, robusto e de fácil entendimento (TEIXEIRA, CORRÊA, ASSIS, 2007).

Existem três formas de combinar arquiteturas de RNAs e lógica nebulosa: arquiteturas neuro-nebulosas concorrentes, arquiteturas neuro-nebulosas cooperativas e arquitetura neuro-nebulosa híbrida (RAPOSO, 2001).

1. Arquiteturas neuro-nebulosas concorrentes: a rede neural serve como pré ou pós-processador para a lógica nebulosa. A RNA não altera nenhum parâmetro da lógica nebulosa. Esta é a forma mais fraca de acoplamento.
2. Arquiteturas neuro-nebulosas cooperativas: a RNA altera parâmetros da lógica nebulosa (regras, conjuntos nebulosos, etc.). Após o aprendizado, a RNA não é mais utilizada.
3. Arquitetura neuro-nebulosa híbrida: os modelos de RNA e lógica nebulosa são misturados de forma a se extrair o que se tem de melhor em cada uma das arquiteturas.

### 3.5.1 Arquitetura ANFIS

O ANFIS tem a função de implementar um FIS, ou seja, sua função consiste em processar variáveis lingüísticas como o processamento encontrado nos sistemas de engenharia *fuzzy*, sob o domínio das redes neurais. Assim como um FIS, o ANFIS possui modelos que implementam formas diferentes dos dados de entrada. Neste trabalho, o ANFIS é referenciado no modelo FIS de Takagi-Sugeno.

Uma arquitetura ANFIS típica possui cinco camadas de neurônios. Cada camada realiza um processamento específico sobre os dados da camada anterior a ela. A figura 3.13 apresenta um exemplo de arquitetura ANFIS baseada no modelo de Takagi-Sugeno.

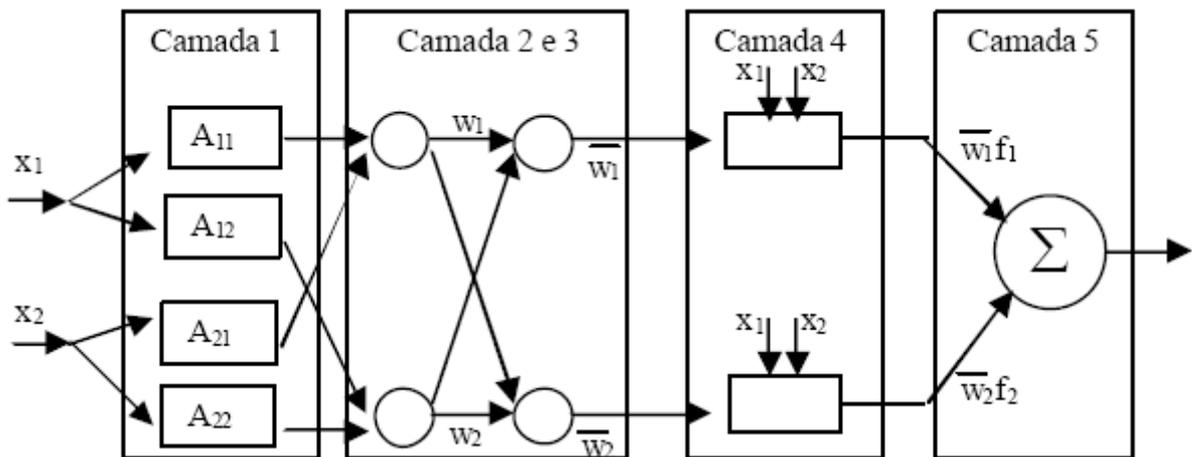


Figura 3.13 – Arquitetura ANFIS do tipo Takagi-Sugeno

O ANFIS recebe as variáveis no formato dos conjuntos clássicos e através de iterações em suas camadas, os dados de entrada são transformados em variáveis linguísticas e ao final retorna à variável em formato dos conjuntos clássicos.

Assim como as redes neurais, cada neurônio do ANFIS possui uma função de ativação. Essa função de ativação é específica para a atuação do neurônio sobre a funcionalidade da camada ao qual este pertence.

Considerando-se, por exemplo, um sistema de inferência de duas variáveis de entrada  $x$  e  $y$  e uma saída  $f$ , para um modelo nebuloso do tipo Takagi-Sugeno de primeira ordem, um conjunto de regras possíveis seria dado por:

Regra 1: Se  $x$  é  $A_1$  e  $y$  é  $B_1$ , então  $f_1 = p_1x + q_1y + r_1$ .

Regra 2: Se  $x$  é  $A_2$  e  $y$  é  $B_2$ , então  $f_2 = p_2x + q_2y + r_2$ .

A figura 3.14 apresenta a base da estrutura ANFIS, onde os parâmetros  $p$ ,  $q$  e  $r$  são identificados de acordo com os seguintes passos (camadas):

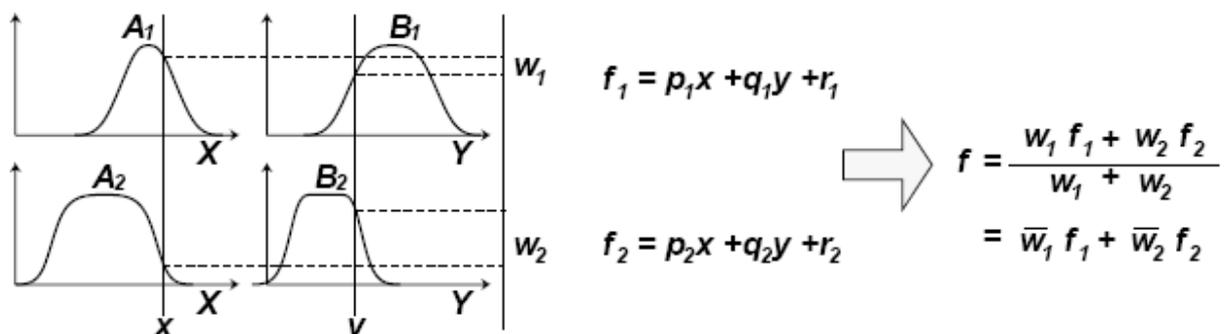


Figura 3.14 – Modelo nebuloso do tipo Sugeno de primeira ordem

**Camada 1:** Esta camada é responsável por transformar cada variável de entrada em variáveis do universo nebuloso, assim é responsável pelo processo de nebulização. Nesta camada cada uma das unidades representa uma função de pertinência associada a uma das entradas e tem como característica a adaptabilidade. Todo nó dessa camada é um nó quadrado, possui parâmetros que podem ser ajustados, com uma função nó dada por (3.45).

$$O_{1,i} = \mu_{A_i}(x), \text{ para } i = 1, 2 \text{ ou} \quad (3.45)$$

$$O_{1,i} = \mu_{B_{i-2}}(y), \text{ para } i = 3, 4$$

onde  $x$  ou  $y$  é a entrada do nó  $i$ , e  $A_i$  ou  $B_{i-2}$  é o valor lingüístico associado a essa função nó, ou seja,  $O_{1,i}$  é a função de pertinência de  $A_i$  e especifica o grau pelo qual um dado valor  $x$  pertence ao conjunto  $A_i$ . Geralmente, é escolhida uma função de pertinência do tipo formato de sino. Funções de pertinência triangulares, trapezoidais ou Gaussianas também são candidatas a serem funções de nó dessa camada. Os parâmetros dessa camada são denominados antecedentes.

**Camada 2:** Com base nos parâmetros de premissas obtidos na primeira camada, na segunda camada são realizadas as operações da base de conhecimento FIS, logo é nesta camada que se encontram as regras de conhecimento do especialista. Todo nó dessa camada é circular, sem parâmetros ajustáveis, e tem como saída o produto dos sinais de entrada. Cada nó de saída representa uma força de ativação de uma regra:

$$O_{2,i} = w_i = \mu_{A_i}(x)\mu_{B_i}(y), \quad i = 1, 2. \quad (3.46)$$

**Camada 3:** Nesta camada é realizada a normalização, onde cada nó calcula a razão entre o nível de disparo da regra  $i$  pela soma dos níveis de disparo de todas as regras. As saídas dessa camada são denominadas de intensidades de disparo normalizadas. O  $i$ -ésimo nó calcula a razão da intensidade de disparo da  $i$ -ésima regra pela soma das intensidades de disparo de todas as regras, conforme (3.47).

$$\bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2. \quad (3.47)$$

**Camada 4:** As saídas do neurônio são calculadas pelo produto entre os níveis de disparo normalizados e o valor das regras do conseqüente da regra em si. Cada nó neste nível é um nó adaptativo com função de ativação dada por (3.48).

$$o_{4,i} = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i). \quad (3.48)$$

onde  $\bar{w}_i$  a força de atuação da camada 3, e  $p_i, q_i, r_i$  são os parâmetros definidos neste nó. Os parâmetros nesta camada são chamados de parâmetros conseqüentes.

**Camada 5:** Esta camada finaliza o processo, transforma as variáveis nebulosas de conseqüência obtidas na camada 4 em um único valor. Em resumo, esta camada realiza o processo de desnebulização de um FIS. Unidade de processamento não adaptativa que calcula a saída final como a soma de todos os sinais de entrada:

$$o_{5,i} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (3.49)$$

## 4 OBTENÇÃO E ANÁLISE DOS RESULTADOS

Neste estudo os modelos para previsão de séries temporais Holt-Winters, ARIMA e ANFIS foram implementados e comparados a fim de se analisar a qualidade de previsão obtida por cada modelo. As análises comparativas foram realizadas utilizando os indicadores de desempenho  $R^2$  (Coeficiente de determinação múltipla), Erro Médio Quadrático (MSE), Erro Percentual Médio Absoluto (MAPE) e Raiz Quadrada do Erro Médio Quadrático (RMSE). Para implementação dos modelos foi utilizado o *software* MATLAB na versão 7.5.

Os modelos para previsão de séries temporais implementados foram aplicados a duas séries de consumo mensal de dois diferentes insumos utilizados no processo produtivo da empresa Beta em estudo. Recorrendo-se a base de dados da empresa pesquisada foram obtidas duas séries considerando o período de novembro de 2003 até maio de 2009, totalizando, em cada série, 67 observações. Para facilitar as análises, as séries estudadas foram denominadas de Insumo X e Insumo Y. Nas figuras 4.1 e 4.2 são apresentadas as duas séries consideradas para o presente estudo.

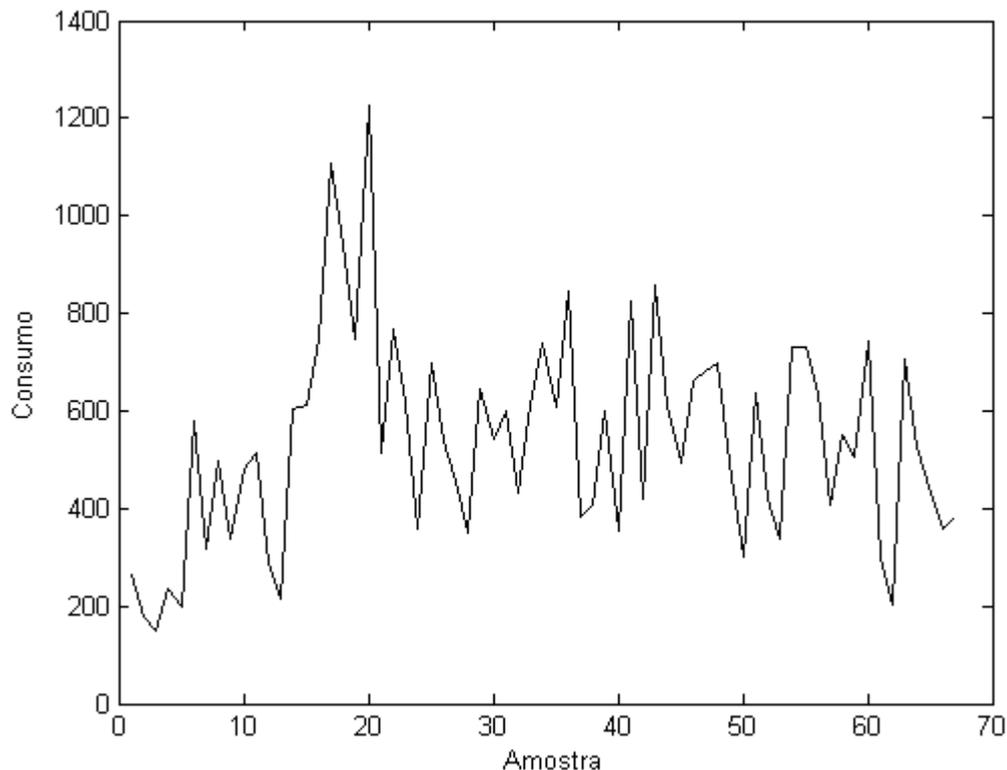


Figura 4.1 – Dados do Insumo X

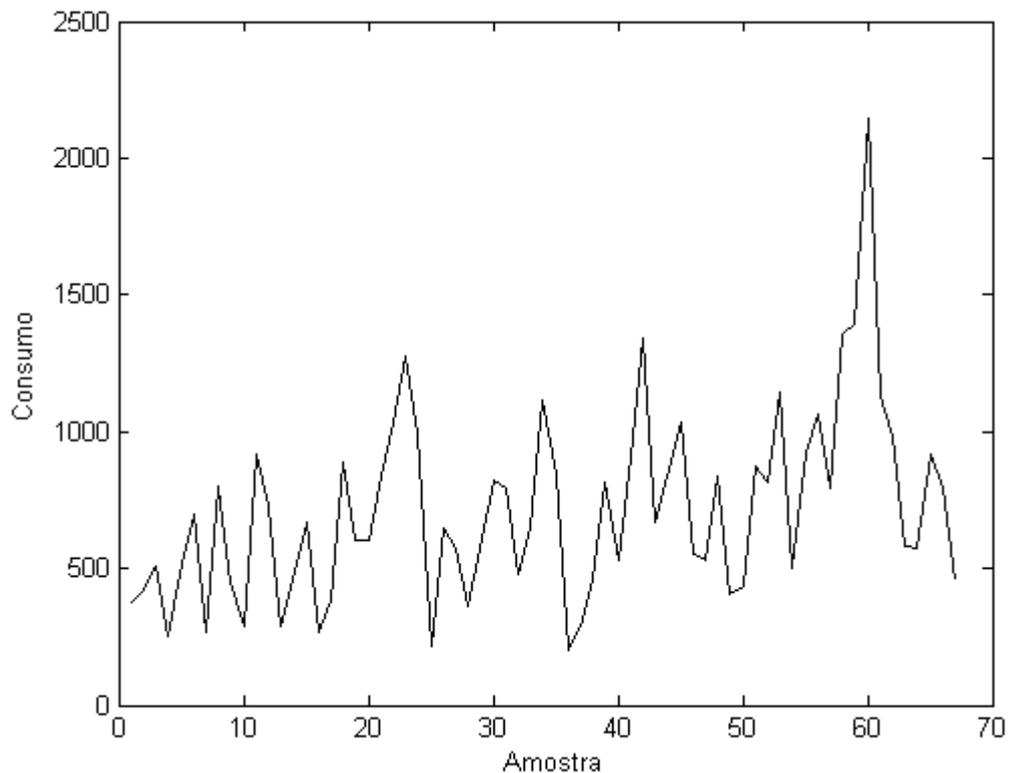


Figura 4.2 – Dados do Insumo Y

O modelo ARIMA implementado utiliza a aproximação de Box-Jenkins com loops para variação das ordens dos parâmetros de 1 a 4 e a função *bj* do *software* MATLAB para calcular a estimativa do erro de previsão do modelo. Na implementação são consideradas duas fases: a fase de estimação onde são utilizadas 50 observações e a fase de validação com 67 observações. Após a finalização dos testes é apresentada a melhor estrutura ARIMA BJ, considerando os melhores valores para as ordens  $nb$ ,  $nc$ ,  $nd$  e  $nf$  e para o atraso  $nk$  do modelo Box-Jenkins obtidos pela minimização dos erros. Através do modelo podem-se minimizar as funções  $RMSE_{estimação}$  e  $RMSE_{total}$  ( $RMSE_{estimação} + RMSE_{validação}$ ).

A forma multiplicativa do modelo Holt-Winters implementada estima os parâmetros Alfa, Beta e Gama com loops para variação de 0 a 1 com um passo de 0,1 e fator sazonalidade estimado com variação de 1 até 20 a um passo de 1. Através do modelo podem-se minimizar as funções  $RMSE_{total}$  ( $RMSE_{estimação} + RMSE_{validação}$ ),  $RMSE_{validação}$  e realizar previsões  $n$  passos à frente e um passo a frente. Após a finalização dos testes são apresentados os melhores parâmetros que minimizam a função objetivo obtida pela minimização dos erros.

O método ANFIS implementado utiliza o Sistema de Inferência *Fuzzy* do tipo Takagi-Sugeno também conhecido por Sugeno. Uma combinação de mínimo quadrado e método *Backpropagation Error* é utilizada no treinamento dos parâmetros das funções FIS para modelar um dado conjunto de entradas e saídas. No modelo são utilizadas as funções de pertinências gaussianas, forma de sino e triangular e o número de funções de pertinência a serem utilizadas podem ser variadas de 2 até 6. Das 67 observações disponíveis para cada série são utilizadas 50 observações na fase de estimação e 67 na fase de validação.

A seguir são apresentadas as simulações realizadas com os três modelos de previsão de séries temporais considerados e os resultados obtidos para cada série em estudo.

#### 4.1 SIMULAÇÕES PARA A SÉRIE INSUMO X

##### 4.1.1 Simulações usando Modelo ARIMA

Para a previsão utilizando o modelo ARIMA foram realizadas duas simulações: a primeira considerando o índice de desempenho a ser minimizado  $RMSE_{est}$  e a segunda considerando o índice de desempenho  $RMSE_{total}$ .

Para cada simulação foram encontradas as melhores estruturas para as séries, sendo  $nb$ ,  $nc$ ,  $nd$  e  $nf$  as ordens e  $nk$  os atrasos do modelo Box-Jenkins. Nas duas simulações realizadas a melhor estrutura encontrada possui os seguintes valores para as ordens e o atraso  $nb = 4$ ,  $nc = 4$ ,  $nd = 4$ ,  $nf = 1$  e  $nk = 3$  com função objetivo igual a 129,285264 para primeira simulação e função objetivo igual a 313,743218 para a segunda simulação.

As figuras 4.3 e 4.4 apresentam a previsão da série Insumo X usando o modelo ARIMA para as duas simulações realizadas.

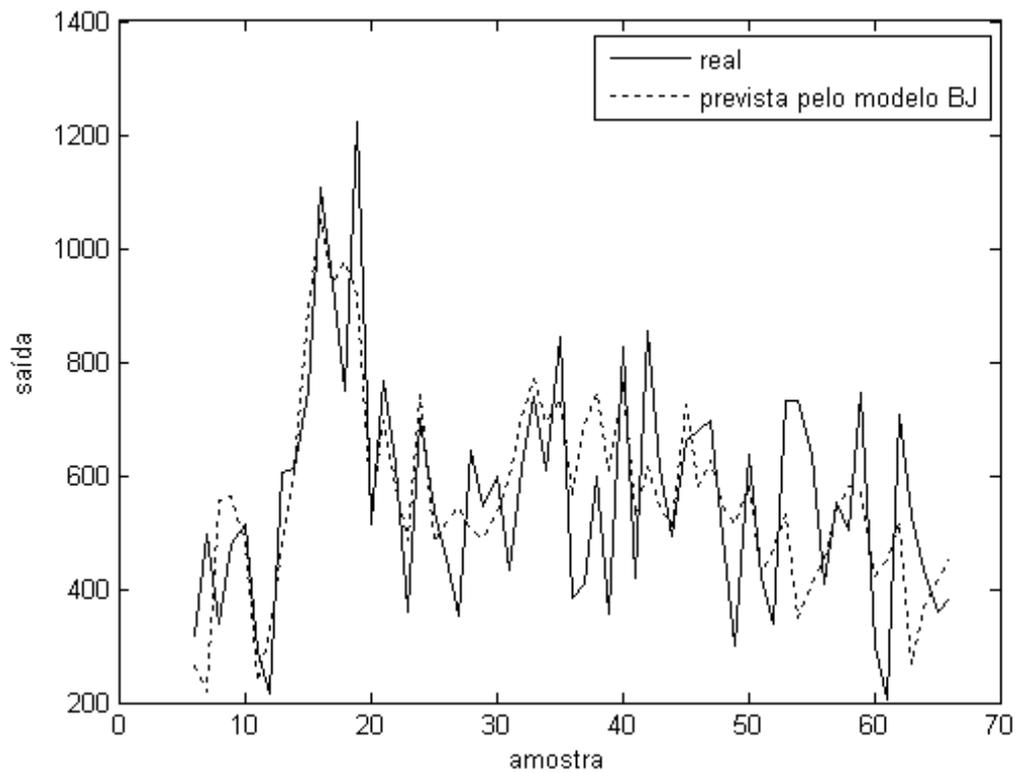


Figura 4.3 – Dados do Insumo X versus previsão do modelo ARIMA (Simulação 1)

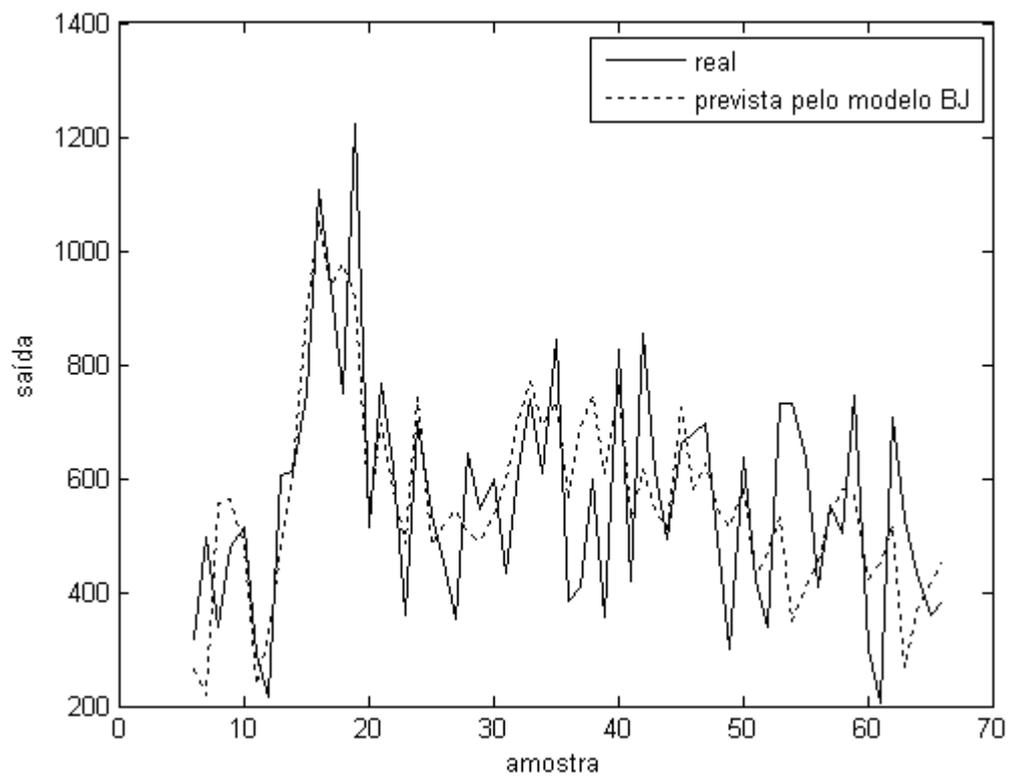


Figura 4.4 – Dados do Insumo X versus previsão do modelo ARIMA (Simulação 2)

Os erros obtidos pelo modelo em cada simulação são apresentados nas figuras 4.5 e 4.6.

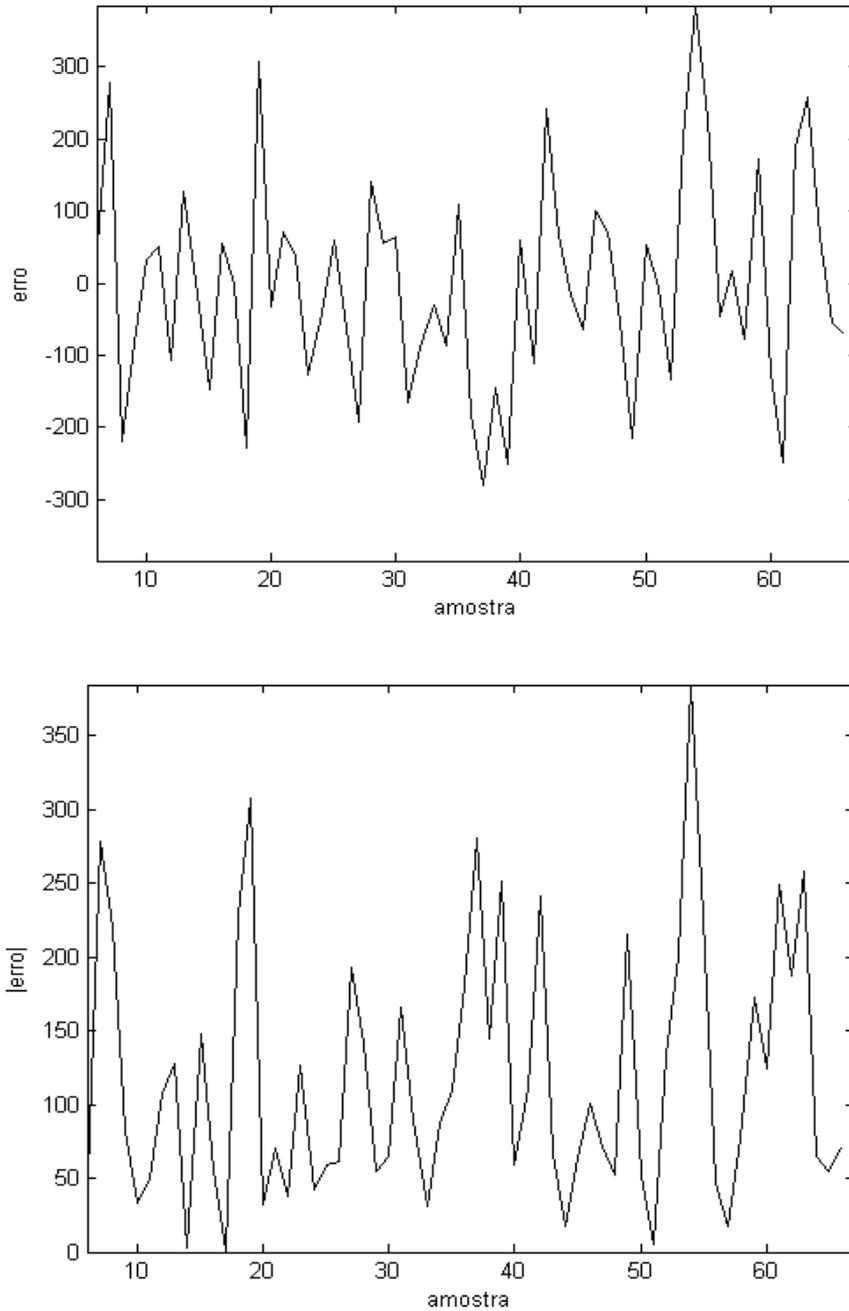


Figura 4.5 – Erros do modelo ARIMA Insumo X (Simulação 1)

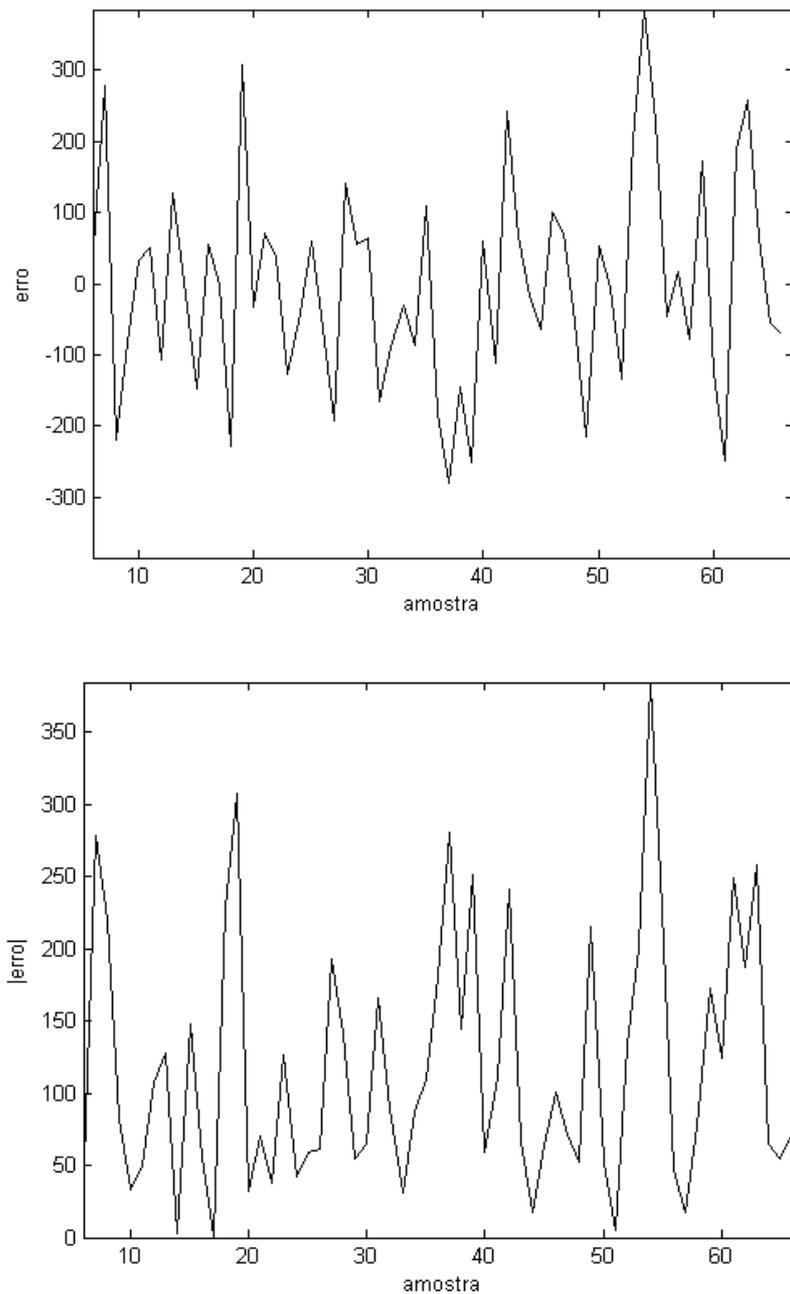


Figura 4.6 – Erros do modelo ARIMA Insumo X (Simulação 2)

Para cada simulação foram encontrados os índices de desempenho RMSE, MSE, MAPE e  $R^2$ . Na tabela 4.1 são descritos os índices encontrados nas fases de estimação, validação e total.

Fase	RMSE	MSE	MAPE	$R^2$
<b>Estimação</b>	$1,29285 \times 10^2$	$1,61021 \times 10^4$	$2,02129 \times 10^1$	$6,70385 \times 10^{-1}$
<b>Validação</b>	$1,84457 \times 10^2$	$2,56722 \times 10^4$	$3,32841 \times 10^1$	0
<b>Total</b>	$1,46064 \times 10^2$	$3,23578 \times 10^2$	$2,28413 \times 10^1$	$5,60694 \times 10^{-1}$

Tabela 4.1- Índices de Desempenho do modelo ARIMA para o Insumo X

#### 4.1.2 Simulações usando Modelo Holt-Winters

Para o modelo Holt-Winters considerando o Insumo X foram realizadas 4 simulações considerando a minimização das funções RMSE\_total (RMSE\_estimacão + RMSE\_validacão), RMSE validacão e horizonte de previsão um passo a frente ou  $n$  passos a frente.

Na primeira simulacão realizada foi considerada a minimizacão da funcão RMSE\_total e na segunda simulacão a minimizacão da funcão RMSE\_validacão com horizonte de previsão  $n$  passos à frente. Para as duas simulacões a melhor sazonalidade encontrada foi 12 e os valores 0.1, 0.4 e 0.4 foram estimados para os parâmetros alfa, beta e gama com funcão objetivo igual a 140,808554.

A figura 4.7 apresenta a previsão da série Insumo X obtida pelo modelo Holt-Winters e a figura 4.8 os erros obtidos para as simulacões.

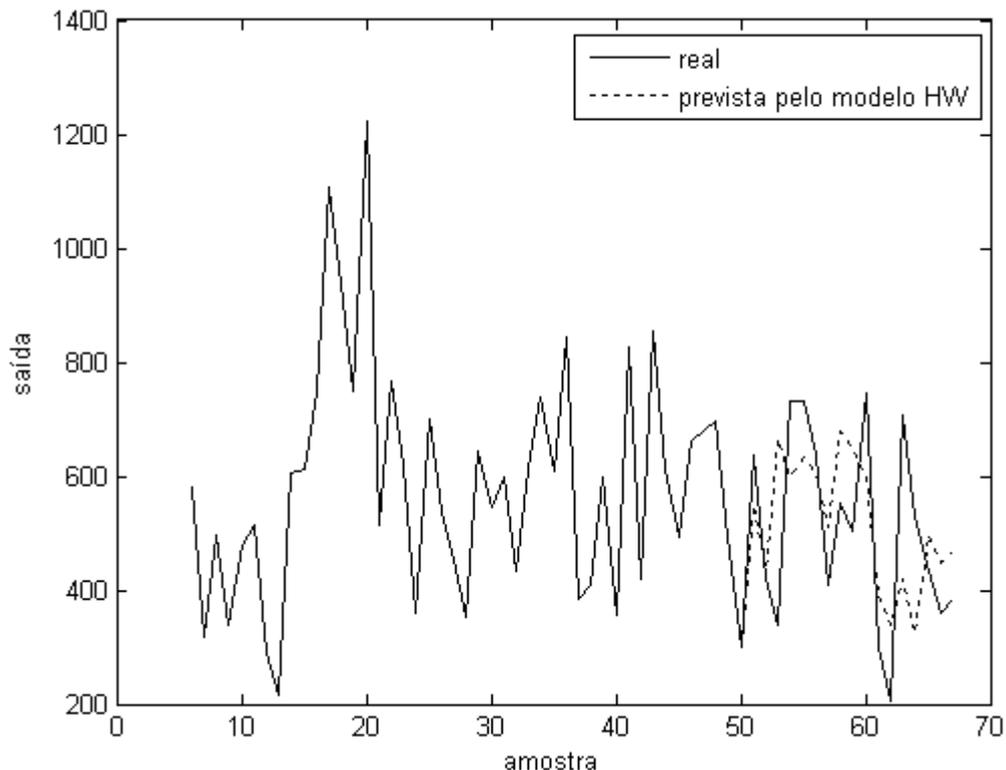


Figura 4.7 – Insumo X versus previsão do modelo HW

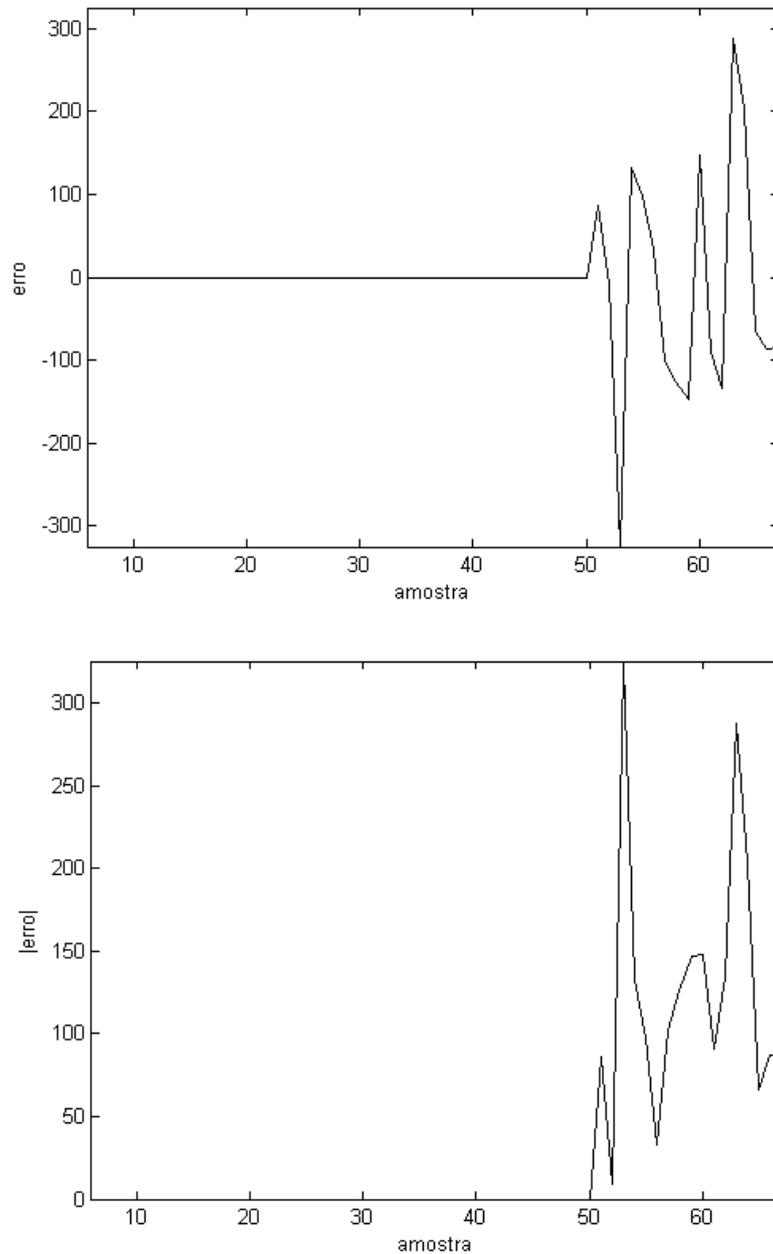


Figura 4.8 – Erros do modelo HW Insumo X

Na tabela 4.2 são descritos os índices de desempenho RMSE, MSE, MAPE e  $R^2$  encontrados nas fases de estimação, validação e total para as duas primeiras simulações. Observa-se que o modelo estimado possui uma adequação aos dados da série na fase de estimação onde os índices RMSE, MSE e MAPE são iguais zero e o índice  $R^2$  igual a 1.

Fase	RMSE	MSE	MAPE	R <sup>2</sup>
<b>Estimação</b>	0	0	0	1
<b>Validação</b>	$1,40808 \times 10^2$	$1,60829 \times 10^4$	$2,84919 \times 10^1$	$3,60561 \times 10^{-1}$
<b>Total</b>	$7,840415 \times 10^1$	$5,16227 \times 10^2$	7,20429	$8,765204 \times 10^{-1}$

Tabela 4.2- Índices de desempenho do modelo HW para o Insumo X

Na terceira simulação foi considerada a minimização da função RMSE\_total e para quarta simulação foi considerada a minimização da função RMSE\_validação com horizonte de previsão um passo a frente. Após a realização dos testes, a melhor sazonalidade encontrada foi 12 e os valores encontrados para os parâmetros alfa, beta e gama foram, respectivamente, 0.1, 0.5 e 0.2 com função objetivo igual a 143,543739.

Na figura 4.9 apresenta a previsão da série Insumo X obtida pelo modelo Holt-Winters.

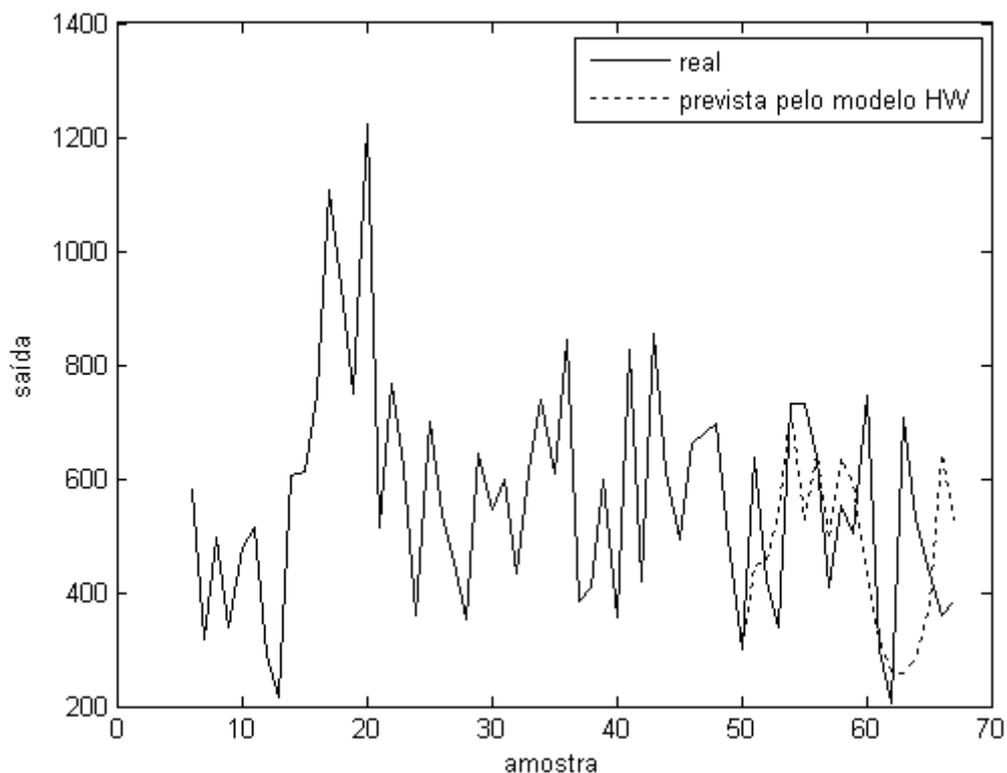


Figura 4.9 – Insumo X versus previsão do modelo HW (Simulações 3 e 4)

A figura 4.10 apresenta os erros obtidos nas simulações e a tabela 4.3 os índices de desempenho calculados em cada fase.

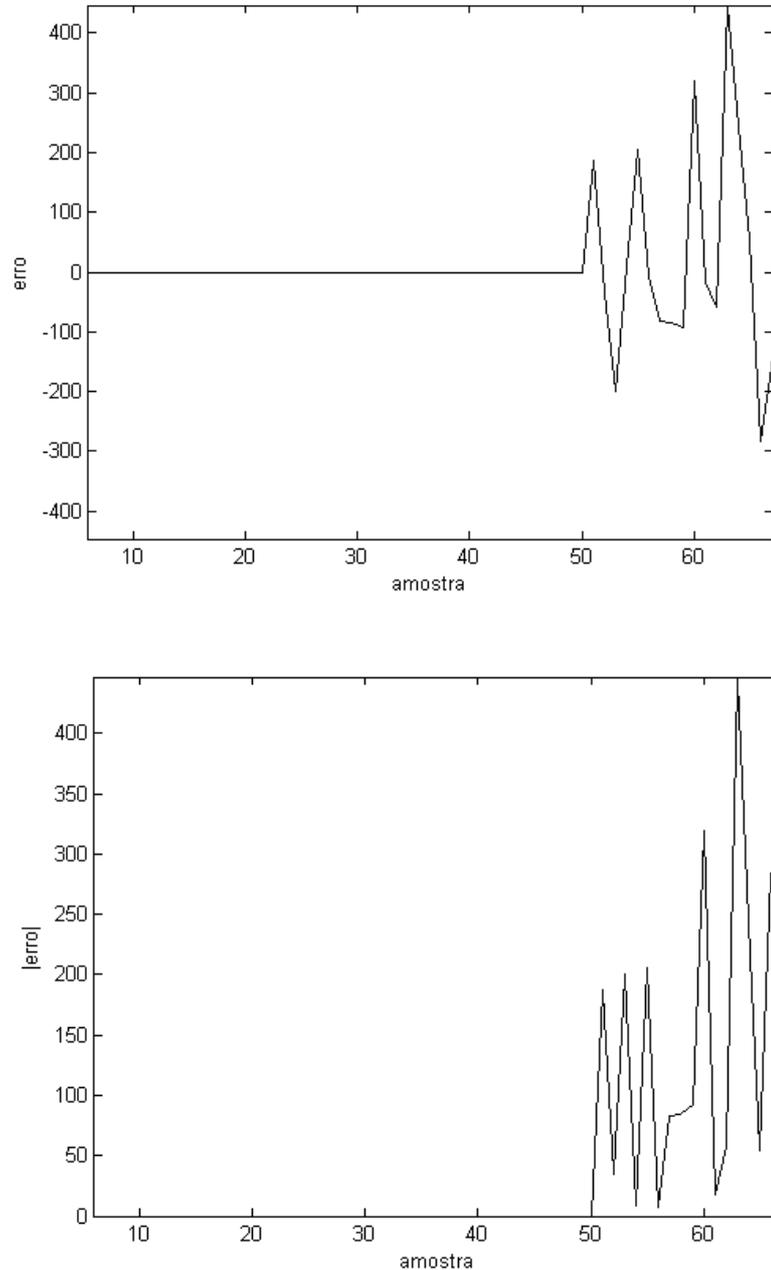


Figura 4.10 – Erros do modelo HW Insumo X (Simulações 3 e 4)

Fase	RMSE	MSE	MAPE	R <sup>2</sup>
Estimação	0	0	0	1
Validação	$1,43543 \times 10^2$	$1,70976 \times 10^3$	$2,15406 \times 10^1$	$3,35478 \times 10^{-1}$
Total	$9,91647 \times 10^1$	$3,22281 \times 10^3$	7,39885	$8,024706 \times 10^{-1}$

Tabela 4.3- Índices de desempenho do modelo HW para o Insumo X (Simulações 3 e 4)

Comparando os índices de desempenho obtidos nas quatro simulações realizadas com o modelo Holt-Winters, para o Insumo X, tem-se que as quatro simulações se mostraram adequadas na fase de estimação, já na fase de validação as duas últimas simulações se adequaram mais a série e na fase total as duas primeiras simulações são mais adequadas comparando-se os índices de desempenho obtidos em cada simulação.

#### 4.1.3 Simulações usando Modelo ANFIS

Foram realizadas três simulações do Insumo X para o modelo ANFIS, variando o número de funções de pertinências e a função de pertinência utilizada. Na primeira simulação foram consideradas 2 funções de pertinência e a função de pertinência gaussiana, na segunda simulação 4 funções de pertinências e a função forma de sino, já na terceira simulação foram utilizadas 6 funções de pertinência e a função triangular.

Após as simulações os seguintes resultados foram obtidos com o modelo ANFIS.

#### **Simulação 1**

Os parâmetros gerados pela estrutura ANFIS, após os testes realizados podem ser visualizados na tabela 4.4.

<b>Parâmetros</b>	<b>Valores</b>
Número de nós	21
Número de parâmetros lineares	12
Número de parâmetros não-lineares	8
Total de parâmetros	20
Número de dados de treinamento	50
Número de dados de validação	17
Número de regras nebulosas	4

Tabela 4.4 - Resultados estrutura ANFIS para a Simulação 1 usando dados do Insumo X

Na figura 4.11 observa-se a série Insumo X e a saída obtida pelo treinamento ANFIS para esta simulação 1.

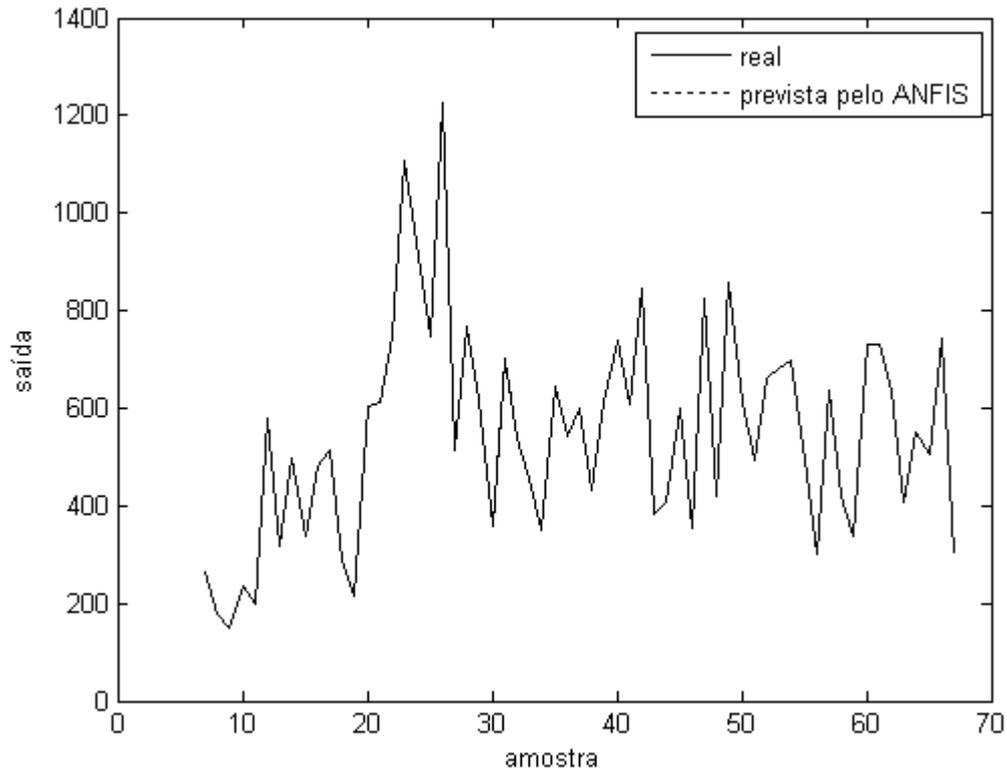


Figura 4.11 – Insumo X versus previsão do modelo ANFIS (Simulação 1)

Os valores obtidos para os índices de desempenho RMSE, MSE, MAPE e  $R^2$  são apresentados na tabela 4.5.

Fase	RMSE	MSE	MAPE	$R^2$
<b>Estimação</b>	$1,54261 \times 10^{-4}$	$1,19992 \times 10^{-7}$	$2,98135 \times 10^{-5}$	$9,99999 \times 10^{-1}$
<b>Validação</b>	$1,20649 \times 10^{-4}$	$4,135207 \times 10^{-10}$	$1,99773 \times 10^{-5}$	$9,99999 \times 10^{-1}$
<b>Total</b>	$1,48762 \times 10^{-4}$	$9,30123 \times 10^{-8}$	$2,80397 \times 10^{-5}$	$9,99999 \times 10^{-1}$

Tabela 4.5 – Índices de desempenho do modelo ANFIS para o Insumo X (Simulação 1)

## Simulação 2

Os parâmetros gerados pela estrutura ANFIS para esta simulação estão apresentados na tabela 4.6.

Parâmetros	Valores
Número de nós	53
Número de parâmetros lineares	48
Número de parâmetros não-lineares	24
Total de parâmetros	72
Número de dados de treinamento	50
Número de dados de validação	17
Número de regras nebulosas	16

Tabela 4.6 – Resultados da estrutura ANFIS para a simulação 2 usando dados do Insumo X

Na figura 4.12 é apresentada a série Insumo X e a saída obtida pelo treinamento ANFIS para esta simulação e na tabela 4.7 apresenta-se os valores obtidos para os índices de desempenho considerados.

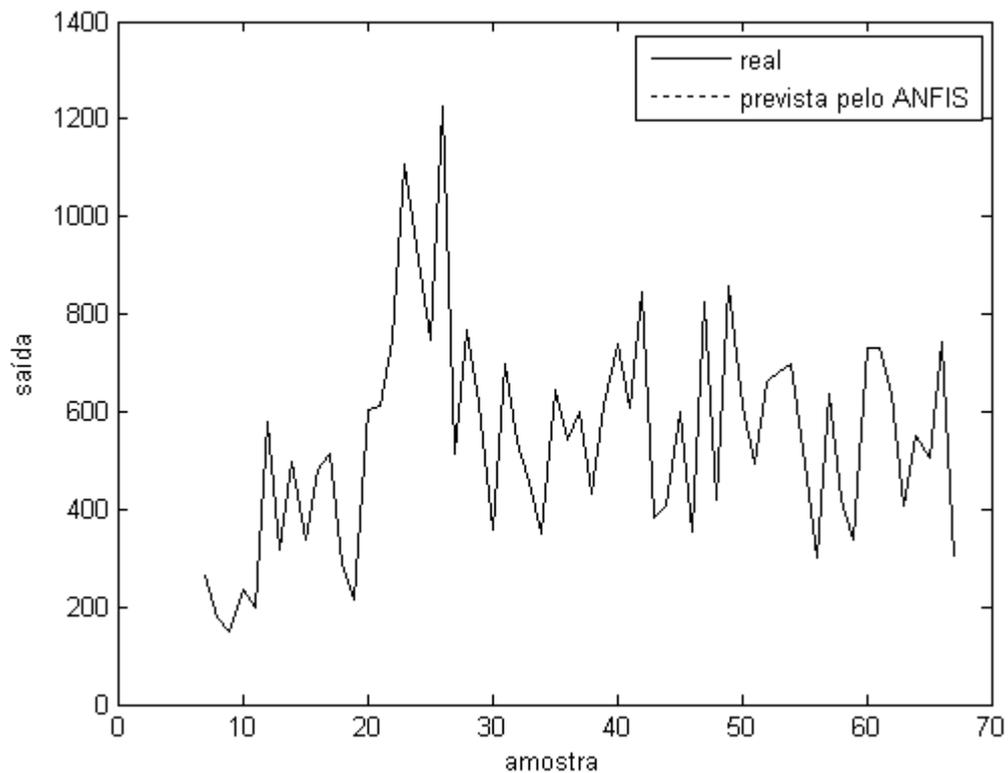


Figura 4.12 – Insumo X versus previsão do modelo ANFIS (Simulação 2)

Fase	RMSE	MSE	MAPE	R <sup>2</sup>
<b>Estimação</b>	$6,83894 \times 10^{-3}$	$8,90703 \times 10^{-5}$	$3,08235 \times 10^{-4}$	$9,9999 \times 10^{-1}$
<b>Validação</b>	$1,667408 \times 10^{-2}$	$2,26877 \times 10^{-4}$	$2,26877 \times 10^{-4}$	$9,9999 \times 10^{-1}$
<b>Total</b>	$9,40599 \times 10^{-3}$	$2,23226 \times 10^{-4}$	$4,33114 \times 10^{-4}$	$9,9999 \times 10^{-1}$

Tabela 4.7 – Índices de Desempenho do modelo ANFIS para o Insumo X (Simulação 2)

### Simulação 3

Os parâmetros gerados pela estrutura ANFIS para esta simulação estão apresentados na tabela 4.8.

Parâmetros	Valores
Número de nós	101
Número de parâmetros lineares	108
Número de parâmetros não-lineares	36
Total de parâmetros	144
Número de dados de treinamento	50
Número de dados de validação	17
Número de regras nebulosas	36

Tabela 4.8 – Resultados da estrutura ANFIS para a simulação 3 usando dados do Insumo X

Na figura 4.13 é apresentada a série Insumo X versus a saída obtida pelo treinamento ANFIS para esta simulação e na tabela 4.9 apresenta-se os valores obtidos nesta simulação para os índices de desempenho considerados.

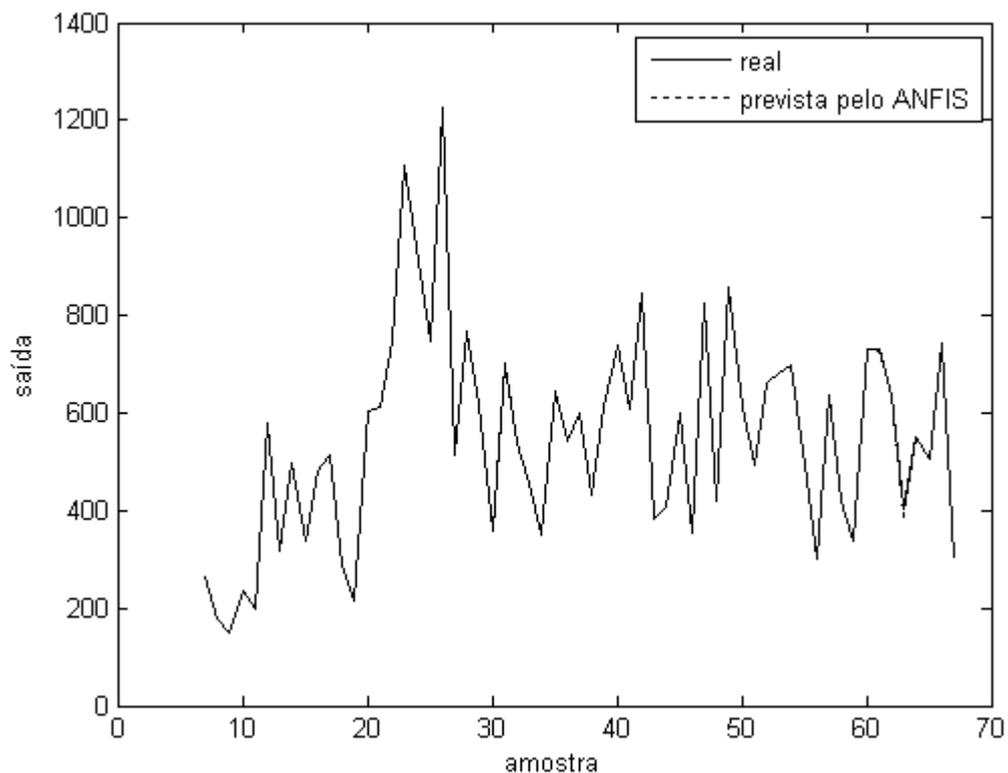


Figura 4.13 – Insumo X versus previsão do modelo ANFIS (Simulação 3)

<b>Fase</b>	<b>RMSE</b>	<b>MSE</b>	<b>MAPE</b>	<b>R<sup>2</sup></b>
<b>Estimação</b>	$2,31107 \times 10^{-2}$	$5,296902 \times 10^{-5}$	$2,18938 \times 10^{-3}$	$9,99999 \times 10^{-1}$
<b>Validação</b>	7,06932	$6,69902 \times 10^1$	$5,78607 \times 10^{-1}$	$9,97926 \times 10^{-1}$
<b>Total</b>	3,00206	$1,20344 \times 10^1$	$1,06133 \times 10^{-1}$	$9,99808 \times 10^{-1}$

Tabela 4.9 – Índices de desempenho do modelo ANFIS para o Insumo X (simulação 3)

Comparando os índices de desempenho obtidos em todas as três simulações realizadas com o modelo ANFIS para o Insumo X tem-se que a simulação 1 considerando 2 funções de pertinência e a função de pertinência gaussiana se aproximou melhor dos dados da série. Comparando os valores obtidos para os índices de desempenho nas três fases consideradas em cada simulação tem-se uma valor de  $R^2$  próximo a 1 e os índices RMSE, MSE e MAPE próximos a zero indicando uma melhor adequação do modelo aos dados da série considerada.

#### 4.1.4 Análise comparativa dos modelos para o Insumo X

Depois de realizadas todas as simulações com os modelos ARIMA, Holt-Winters e ANFIS obteve-se os índices de desempenho de cada modelo estudado, os valores encontrados são apresentados na tabela 4.10.

Índice de desempenho	ARIMA	Holt-Winters	ANFIS
<b>Estimação</b>			
RMSE	$1,29285 \times 10^2$	0	$1,54261 \times 10^{-4}$
MSE	$1,61021 \times 10^4$	0	$1,19992 \times 10^{-7}$
MAPE	$2,02129 \times 10^1$	0	$2,98135 \times 10^{-5}$
$R^2$	$6,70385 \times 10^{-1}$	1	$9,99999 \times 10^{-1}$
<b>Validação</b>			
RMSE	$1,84457 \times 10^2$	$1,40808 \times 10^2$	$1,20649 \times 10^{-4}$
MSE	$2,56722 \times 10^4$	$1,60829 \times 10^4$	$4,135207 \times 10^{-10}$
MAPE	$3,32841 \times 10^1$	$2,84919 \times 10^1$	$1,99773 \times 10^{-5}$
$R^2$	0	$3,60561 \times 10^{-1}$	$9,99999 \times 10^{-1}$
<b>Total (Estimação + Validação)</b>			
RMSE	$1,46064 \times 10^2$	$7,840415 \times 10^1$	$1,48762 \times 10^{-4}$
MSE	$3,23578 \times 10^2$	$5,16227 \times 10^2$	$9,30123 \times 10^{-8}$
MAPE	$2,28413 \times 10^1$	7,20429	$2,80397 \times 10^{-5}$
$R^2$	$5,60694 \times 10^{-1}$	$8,765204 \times 10^{-1}$	$9,99999 \times 10^{-1}$

Tabela 4.10 – Comparativo de resultados para o Insumo X

Para a série Insumo X o modelo ANFIS obteve o melhor desempenho em todos os índices, para as três fases de previsão obtendo uma melhor aproximação da série Insumo X. O modelo ANFIS considerado corresponde a simulação 1 utilizando duas funções de pertinência e a função de pertinência gaussiana. Na fase de estimação o Modelo Holt-Winters obteve os melhores resultados para os índices de desempenho demonstrando uma melhor aproximação da série nesta fase.

## 4.2 SIMULAÇÕES PARA A SÉRIE INSUMO Y

### 4.2.1 Simulações usando Modelo ARIMA

Para previsão da série Insumo Y utilizando o modelo ARIMA foram realizadas duas simulações: a primeira considerando o índice de desempenho (função objetivo) a ser minimizado RMSE\_est e a segunda com o índice de desempenho RMSE\_total.

Na primeira simulação, a melhor estrutura encontrada possui os seguintes valores para as ordens e o atraso do modelo ARIMA  $nb = 1$ ,  $nc = 4$ ,  $nd = 4$ ,  $nf = 4$  e  $nk = 0$  com função objetivo igual a 174,797885. Já na segunda simulação foram

obtidos os seguintes valores para a estrutura do modelo  $nb = 4$ ,  $nc = 4$ ,  $nd = 2$ ,  $nf = 3$  e  $nk = 1$  com função objetivo igual a 636,228756.

As figuras 4.14 e 4.15 apresentam a previsão da série Insumo Y usando o modelo ARIMA para as duas simulações realizadas.

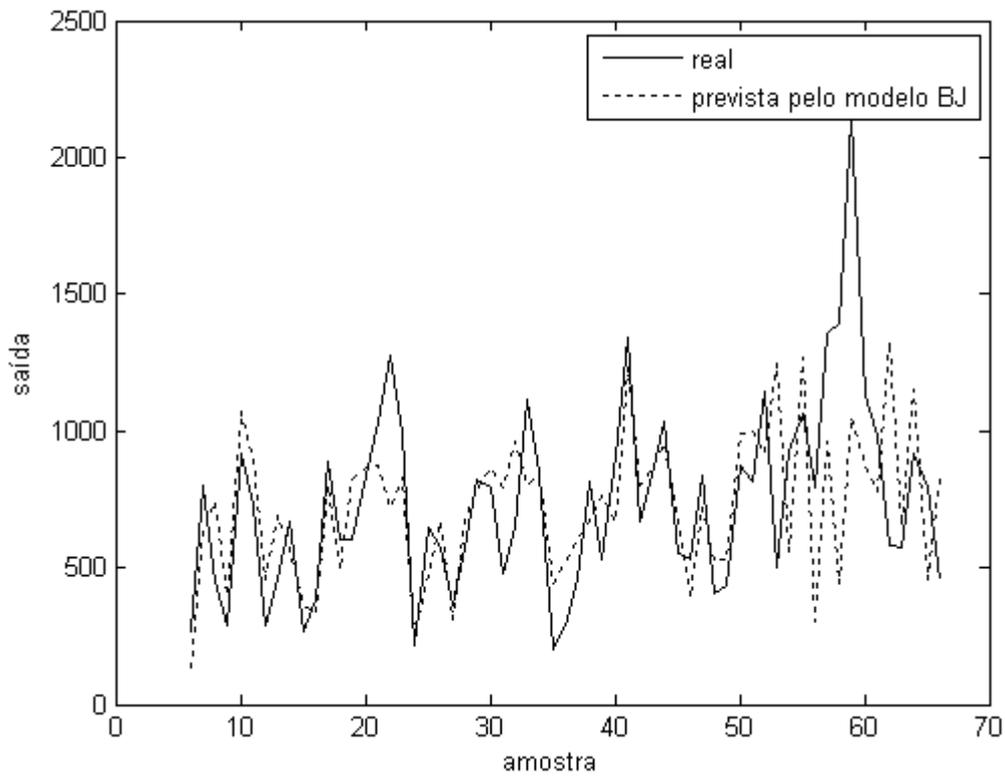


Figura 4.14 – Insumo Y versus previsão do modelo ARIMA (simulação 1)

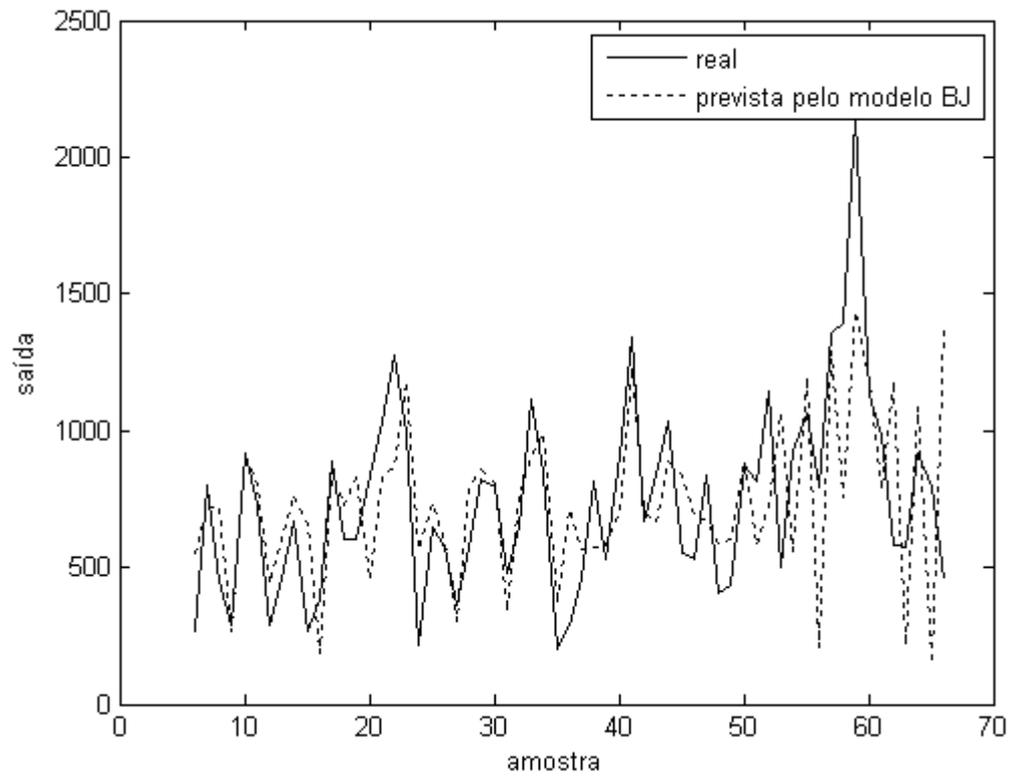
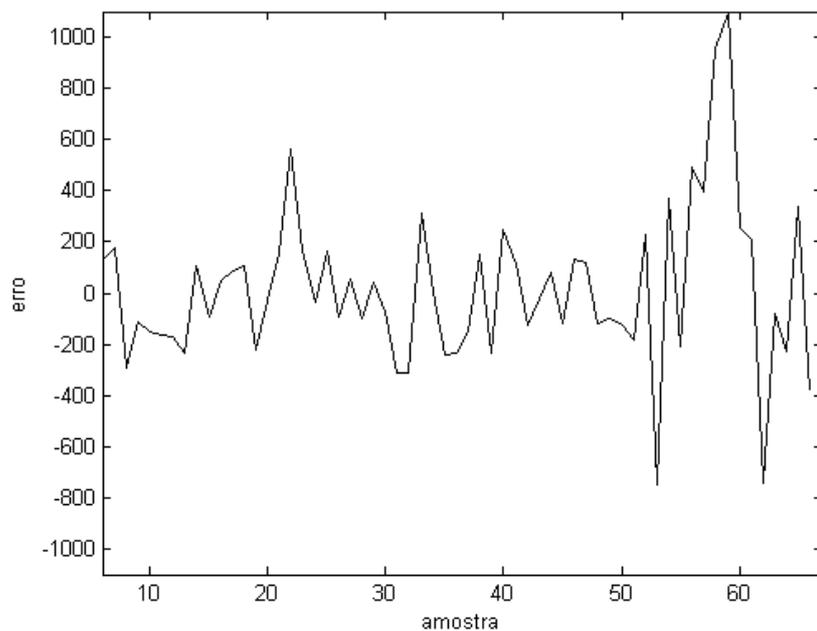


Figura 4.15 – Insumo Y versus previsão do modelo ARIMA (simulação 2)

Os erros obtidos pelo modelo em cada simulação são apresentados nas figuras 4.16 e 4.17.



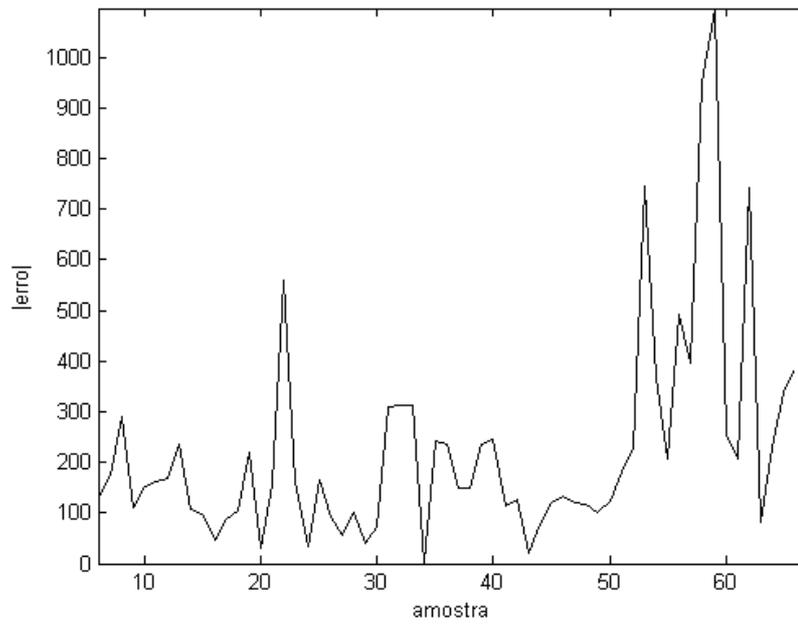
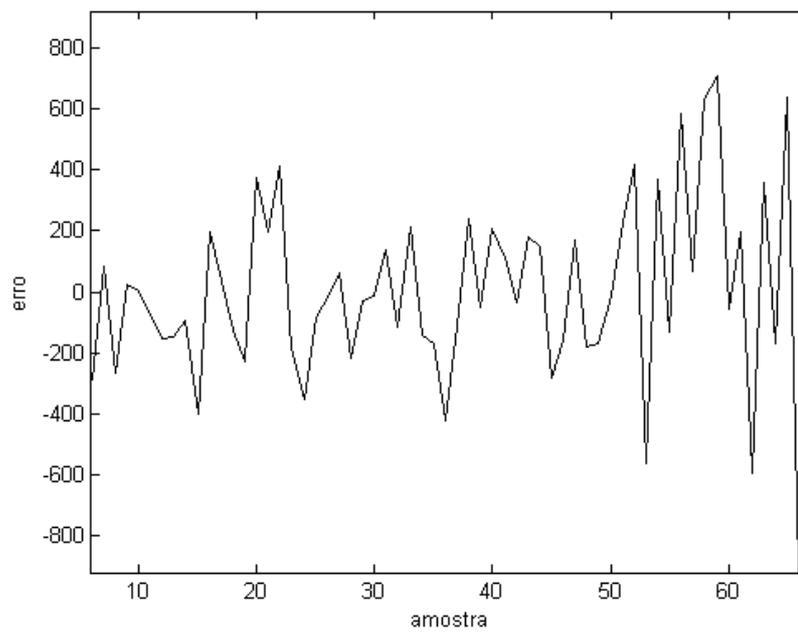


Figura 4.16 – Erros do modelo ARIMA para o Insumo Y (Simulação 1)



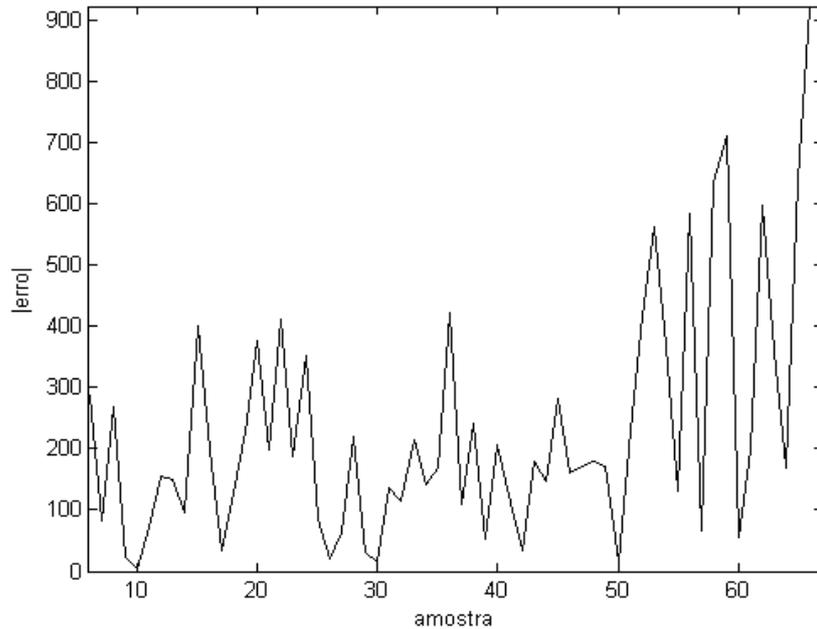


Figura 4.17 – Erros do modelo ARIMA para o Insumo Y (Simulação 2)

Para cada simulação foram encontrados os índices de desempenho RMSE, MSE, MAPE e  $R^2$ . Nas tabelas 4.11 e 4.12 são descritos os índices encontrados nas fases de estimação, validação e total.

Fase	RMSE	MSE	MAPE	$R^2$
<b>Estimação</b>	$1,74797 \times 10^2$	$1,25242 \times 10^4$	$2,64115 \times 10^1$	$5,96536 \times 10^{-1}$
<b>Validação</b>	$5,75954 \times 10^2$	$3,73011 \times 10^5$	$5,25424 \times 10^1$	0
<b>Total</b>	$3,06577 \times 10^2$	$1,41908 \times 10^4$	$3,20693 \times 10^1$	$2,53136 \times 10^{-1}$

Tabela 4.11 - Índices de Desempenho do modelo ARIMA para o Insumo Y (Simulação 1)

Fase	RMSE	MSE	MAPE	$R^2$
<b>Estimação</b>	$1,93121 \times 10^2$	$7,88948 \times 10^4$	$3,48196 \times 10^1$	$5,07514 \times 10^{-1}$
<b>Validação</b>	$4,43107 \times 10^2$	$2,90954 \times 10^5$	$4,28488 \times 10^1$	0
<b>Total</b>	$3,01651 \times 10^2$	$6,684608 \times 10^2$	$3,96926 \times 10^1$	$2,76942 \times 10^{-1}$

Tabela 4.12 - Índices de Desempenho do modelo ARIMA para o Insumo Y (Simulação 2)

Comparando os índices de desempenho obtidos em todas as simulações realizadas com o modelo ARIMA para a série Insumo Y, obteve-se na fase de estimação uma melhor adequação da estrutura do modelo encontrada na simulação

1. Já nas fases de validação e total a melhor estrutura foi a obtida na simulação 2 comparando-se os valores dos índices de desempenho obtidos em cada simulação.

#### 4.2.2 Simulações usando o Modelo Holt-Winters

Para o modelo Holt-Winters considerando o Insumo Y foram realizadas 4 simulações: utilizando a minimização dos índices de desempenhos (função objetivo) RMSE\_total e RMSE\_validação com horizonte de previsão um passo a frente ou  $n$  passos a frente.

Na primeira simulação realizada foi considerada a minimização da função RMSE total e na segunda simulação a minimização da função RMSE validação com horizonte de previsão  $n$  passos à frente. Para as duas simulações, a melhor sazonalidade encontrada foi 18 e os valores 0,1, 0,3 e 0,8 foram estimados para os parâmetros alfa, beta e gama com função objetivo igual a 282,737237.

A figura 4.18 apresenta a previsão da série Insumo Y obtida pelo modelo Holt-Winters e a figura 4.19, os erros obtidos para as simulações.

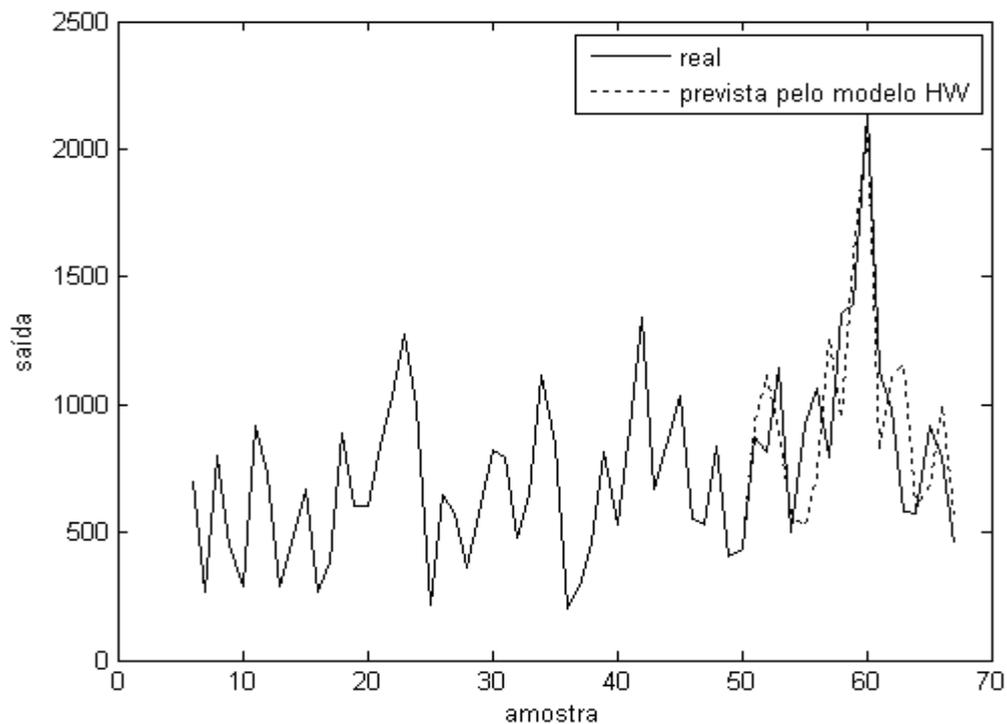


Figura 4.18 – Insumo Y versus previsão do modelo HW

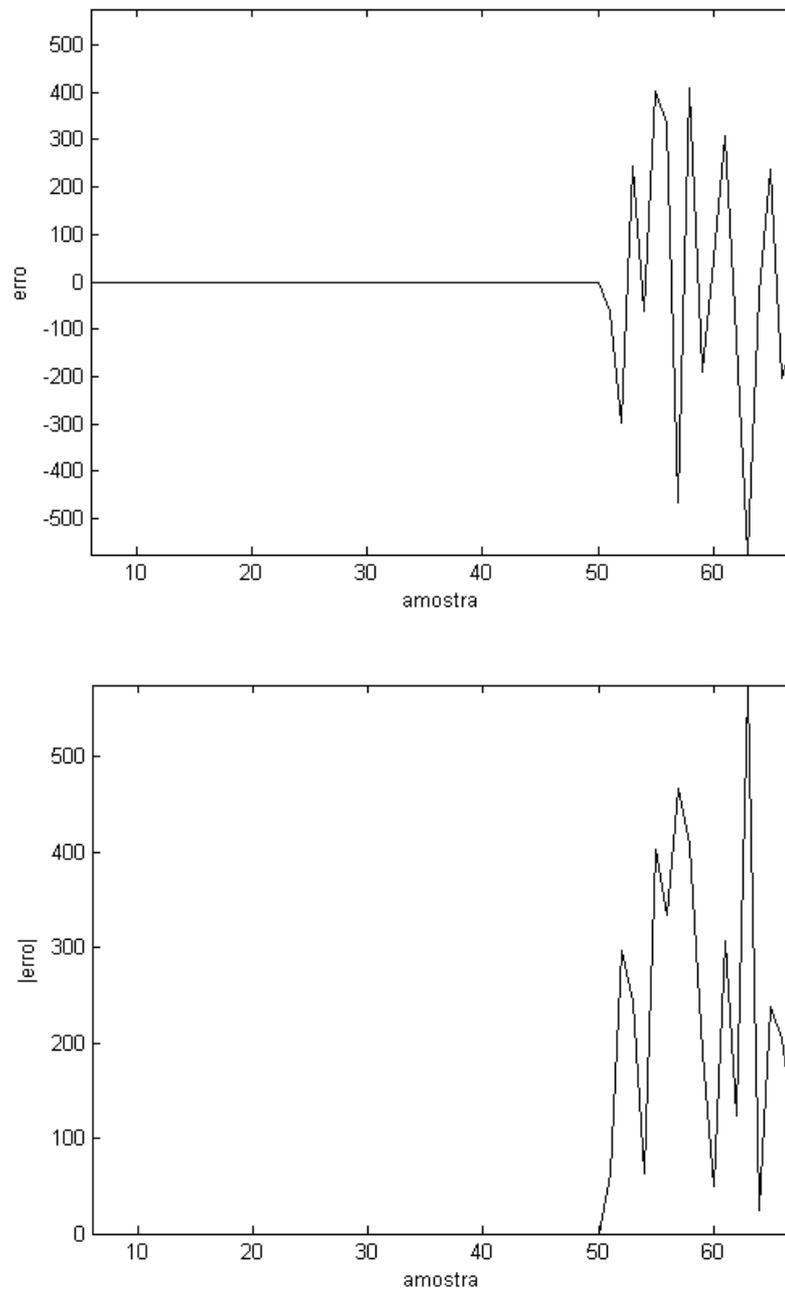


Figura 4.19 – Erros do modelo HW para o Insumo Y

Na tabela 4.13 são descritos os índices de desempenho RMSE, MSE, MAPE e  $R^2$  encontrados nas fases de estimação, validação e total para as duas primeiras simulações. Observa-se que o modelo estimado possui uma adequação aos dados da série na fase de estimação onde os índices RMSE, MSE e MAPE são iguais zero e o índice  $R^2$  igual a 1.

Fase	RMSE	MSE	MAPE	R <sup>2</sup>
<b>Estimação</b>	0	0	0	1
<b>Validação</b>	$2,82737 \times 10^2$	$2,46244 \times 10^4$	$2,47333 \times 10^1$	$4,8989 \times 10^{-1}$
<b>Total</b>	$1,50325 \times 10^2$	$2,61403 \times 10^2$	7,10271	$8,23043 \times 10^{-1}$

Tabela 4.13- Índices de desempenho do modelo HW para o Insumo Y

Na terceira simulação foi considerada a minimização da função objetivo RMSE\_total e para quarta simulação foi considerada a minimização da função RMSE\_validação com horizonte de previsão um passo a frente. Após a realização dos testes, a melhor sazonalidade encontrada foi 18 e os valores encontrados para os parâmetros alfa, beta e gama foram, respectivamente, 0,1, 0,1 e 0,8 com função objetivo igual a 307,134267.

Na figura 4.20 apresenta a previsão da série Insumo X obtida pelo modelo Holt-Winters.

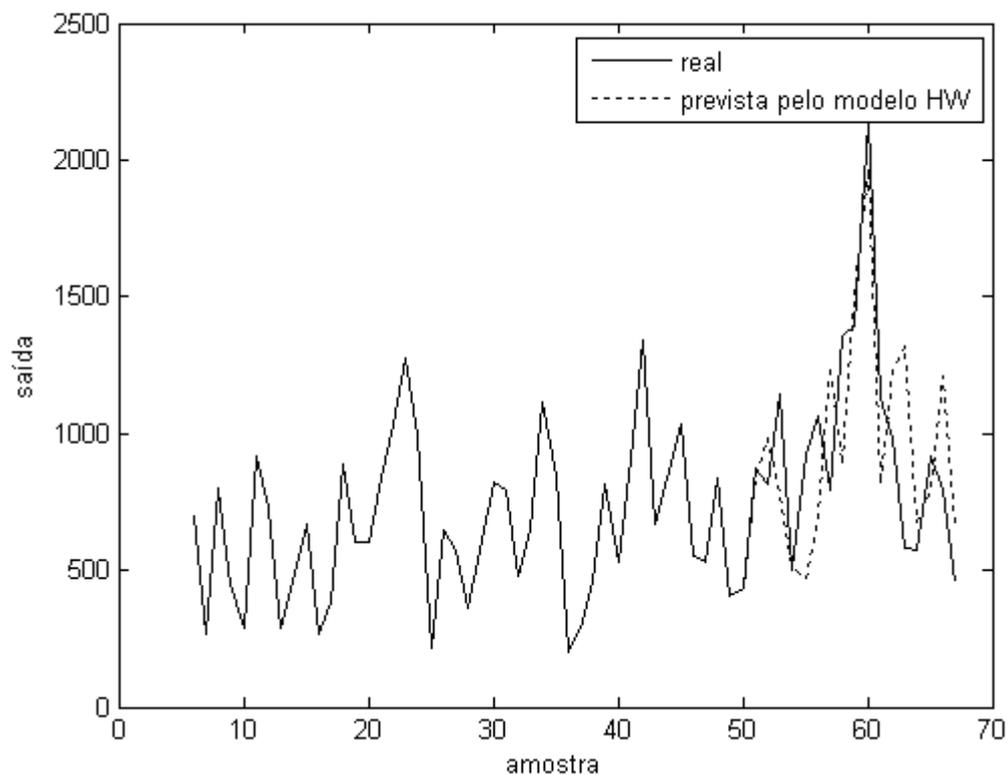


Figura 4.20 – Insumo Y versus previsão do modelo HW (Simulações 3 e 4)

A figura 4.21 apresenta os erros obtidos nas simulações e a tabela 4.14 os índices de desempenho calculados em cada fase.

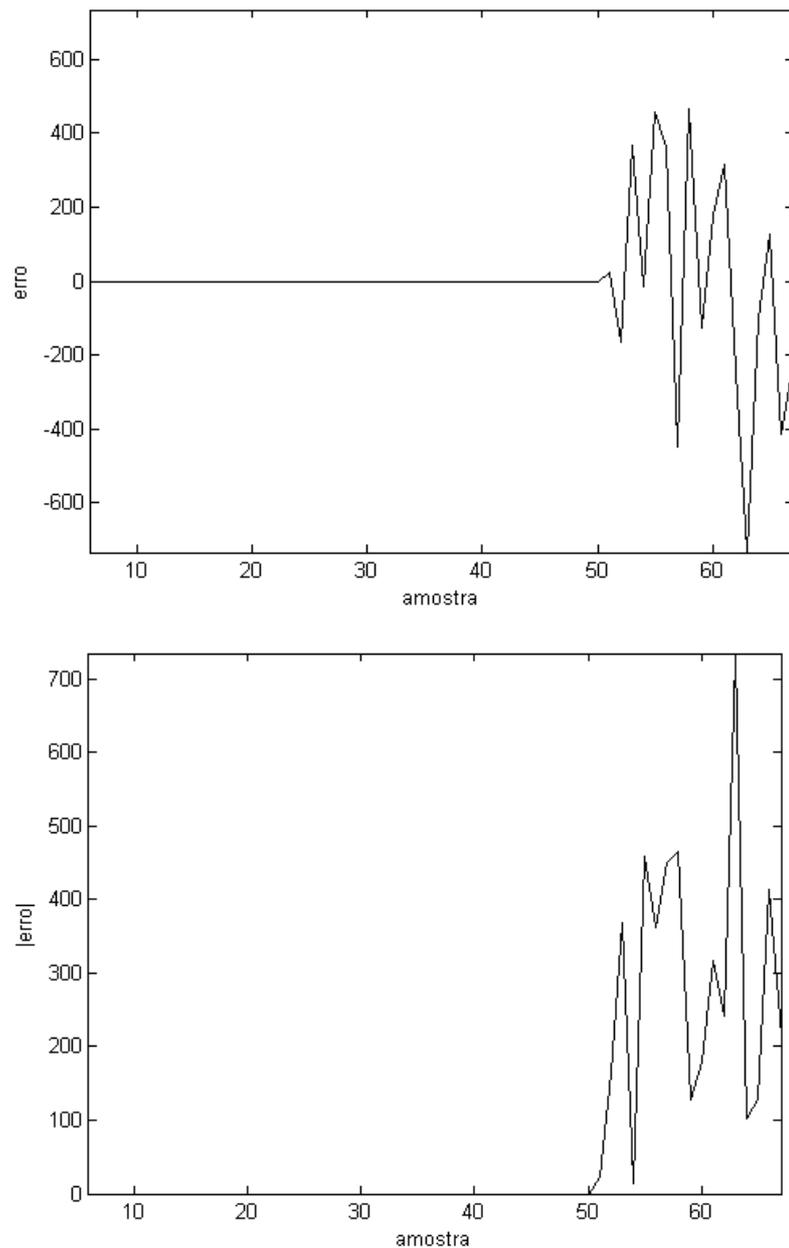


Figura 4.21 – Erros do modelo HW para o Insumo Y (Simulações 3 e 4)

Fase	RMSE	MSE	MAPE	R <sup>2</sup>
<b>Estimação</b>	0	0	0	1
<b>Validação</b>	$3,07134 \times 10^2$	$1,15356 \times 10^5$	$2,52301 \times 10^1$	$3,98059 \times 10^{-1}$
<b>Total</b>	$1,75853 \times 10^2$	$3,96573 \times 10^2$	8,36476	$7,57839 \times 10^{-1}$

Tabela 4.14- Índices de desempenho do modelo HW para o Insumo Y (simulações 3 e 4)

Comparando os índices de desempenho obtidos nas quatro simulações realizadas com o modelo Holt-Winters, para o Insumo Y, tem-se que as quatro simulações se apresentaram adequadas na fase de estimação. As duas primeiras

simulações realizadas se adequaram mais a série comparando-se os índices de desempenho obtidos em cada simulação.

#### 4.2.3 Simulações usando Modelo ANFIS

Foram realizadas três simulações do Insumo Y para o modelo ANFIS, variando o número de funções de pertinências e a função de pertinência utilizada. Na primeira simulação foram consideradas 2 funções de pertinência e a função de pertinência gaussiana; na segunda simulação 4 funções de pertinências e a função forma de sino; já na terceira simulação foram utilizadas 6 funções de pertinência e a função triangular.

Após as simulações os seguintes resultados foram obtidos com o modelo ANFIS.

#### Simulação 1

Os parâmetros gerados pela estrutura ANFIS para esta simulação podem ser visualizados na tabela 4.15.

<b>Parâmetros</b>	<b>Valores</b>
Número de nós	21
Número de parâmetros lineares	12
Número de parâmetros não-lineares	8
Total de parâmetros	20
Número de dados de treinamento	50
Número de dados de validação	17
Número de regras nebulosas	4

Tabela 4.15 - Resultados estrutura ANFIS para a simulação 1 usando dados do Insumo Y

Na figura 4.22 é apresentada a série Insumo Y e a saída obtida pelo treinamento ANFIS para esta simulação e na tabela 4.16 apresenta-se os valores obtidos para os índices de desempenho considerados.

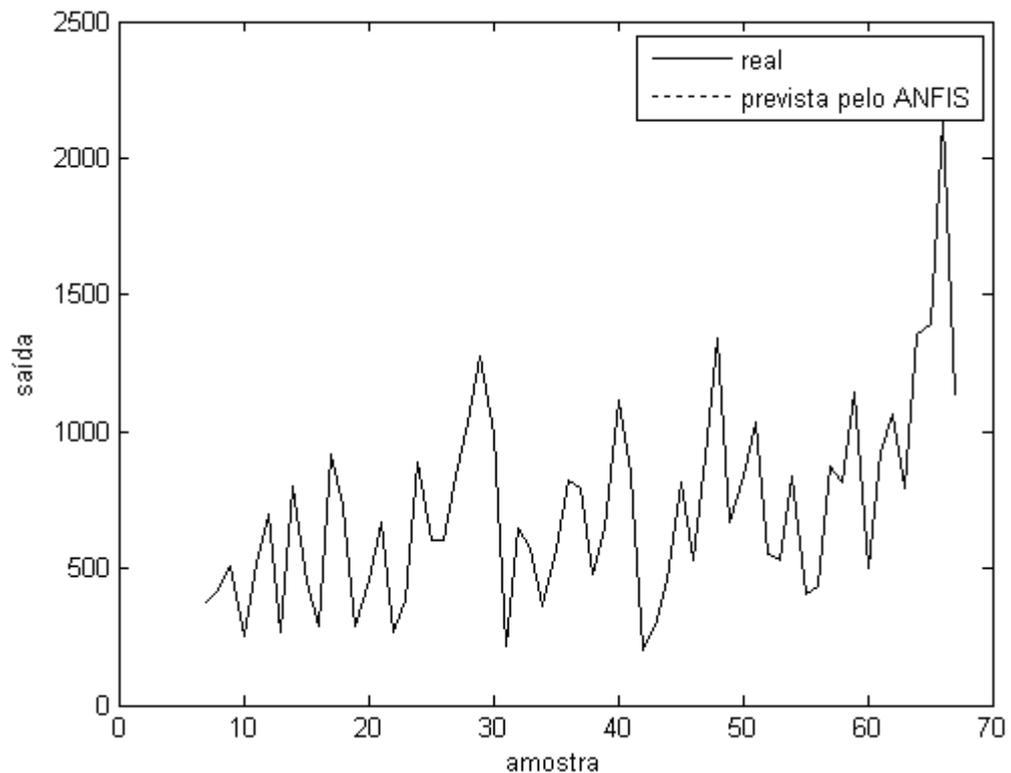


Figura 4.22 – Insumo Y versus previsão do modelo ANFIS (Simulação 1)

Fase	RMSE	MSE	MAPE	R <sup>2</sup>
<b>Estimação</b>	$4,844201 \times 10^{-5}$	$2,97397 \times 10^{-8}$	$8,24642 \times 10^{-6}$	$9,99999 \times 10^{-1}$
<b>Validação</b>	$7,53515 \times 10^{-5}$	$4,856515 \times 10^{-11}$	$3,491408 \times 10^{-6}$	$9,99999 \times 10^{-1}$
<b>Total</b>	$5,42894 \times 10^{-5}$	$2,53097 \times 10^{-8}$	$7,38896 \times 10^{-6}$	$9,99999 \times 10^{-1}$

Tabela 4.16 – Índices de desempenho do modelo ANFIS para o Insumo Y (simulação 1)

### Simulação 2

Após o treinamento é possível visualizar, através da tabela 4.17, os parâmetros gerados pela estrutura ANFIS.

Parâmetros	Valores
Número de nós	53
Número de parâmetros lineares	48
Número de parâmetros não-lineares	24
Total de parâmetros	72
Número de dados de treinamento	50
Número de dados de validação	17
Número de regras nebulosas	16

Tabela 4.17 – Resultados da estrutura ANFIS para a simulação 2 usando dados do Insumo Y

Na figura 4.23 é apresentada a série Insumo Y versus a saída obtida pelo treinamento ANFIS para esta simulação.

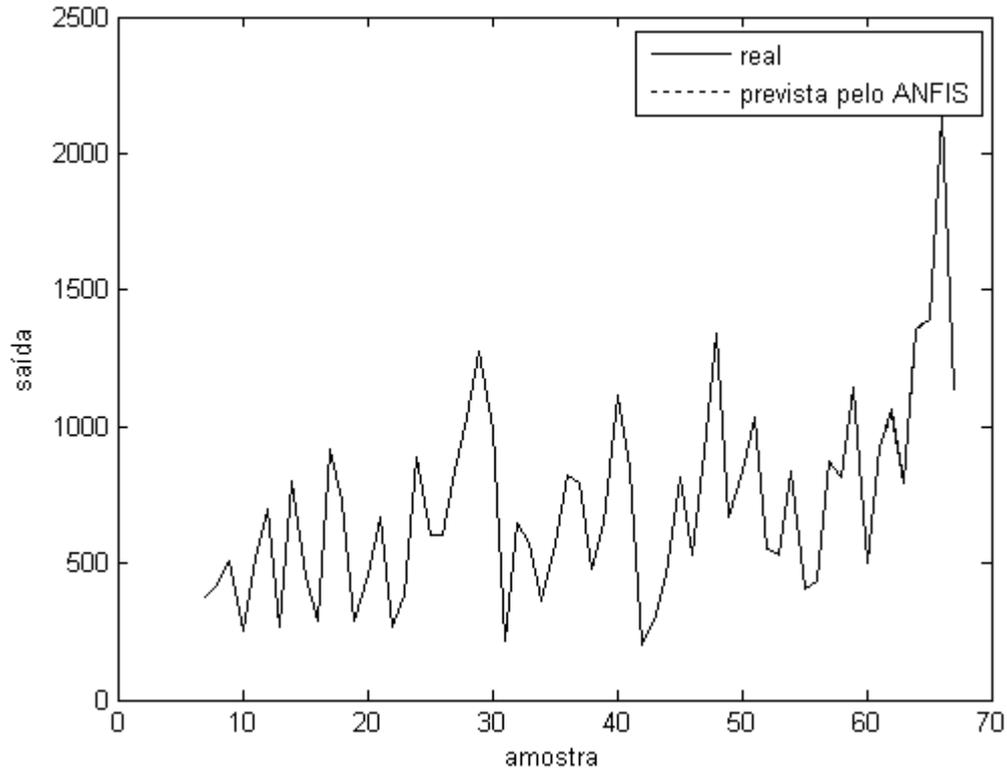


Figura 4.23 – Insumo Y versus previsão do modelo ANFIS (Simulação 2)

Na tabela 4.18 apresentam-se os valores obtidos nesta simulação para os índices de desempenho considerados.

Fase	RMSE	MSE	MAPE	R <sup>2</sup>
<b>Estimação</b>	$2,30816 \times 10^{-2}$	$1,07627 \times 10^{-4}$	$1,411817 \times 10^{-3}$	$9,99999 \times 10^{-1}$
<b>Validação</b>	5,705639	7,36222	$2,264279 \times 10^{-1}$	$9,99808 \times 10^{-1}$
<b>Total</b>	2,42299	1,30605	$4,19884 \times 10^{-2}$	$9,99953 \times 10^{-1}$

Tabela 4.18 – Índices de desempenho do modelo ANFIS para o Insumo Y (simulação 2)

### Simulação 3

Os parâmetros gerados pela estrutura ANFIS para esta simulação podem ser visualizados na tabela 4.19.

Parâmetros	Valores
Número de nós	101
Número de parâmetros lineares	108
Número de parâmetros não-lineares	36
Total de parâmetros	144
Número de dados de treinamento	50
Número de dados de validação	17
Número de regras nebulosas	36

Tabela 4.19 – Resultados da estrutura ANFIS para a simulação 3 usando dados do Insumo Y

Na figura 4.24 é apresentada a série Insumo Y e a saída obtida pelo treinamento ANFIS para esta simulação e na tabela 4.20 apresenta-se os valores obtidos para os índices de desempenho considerados.

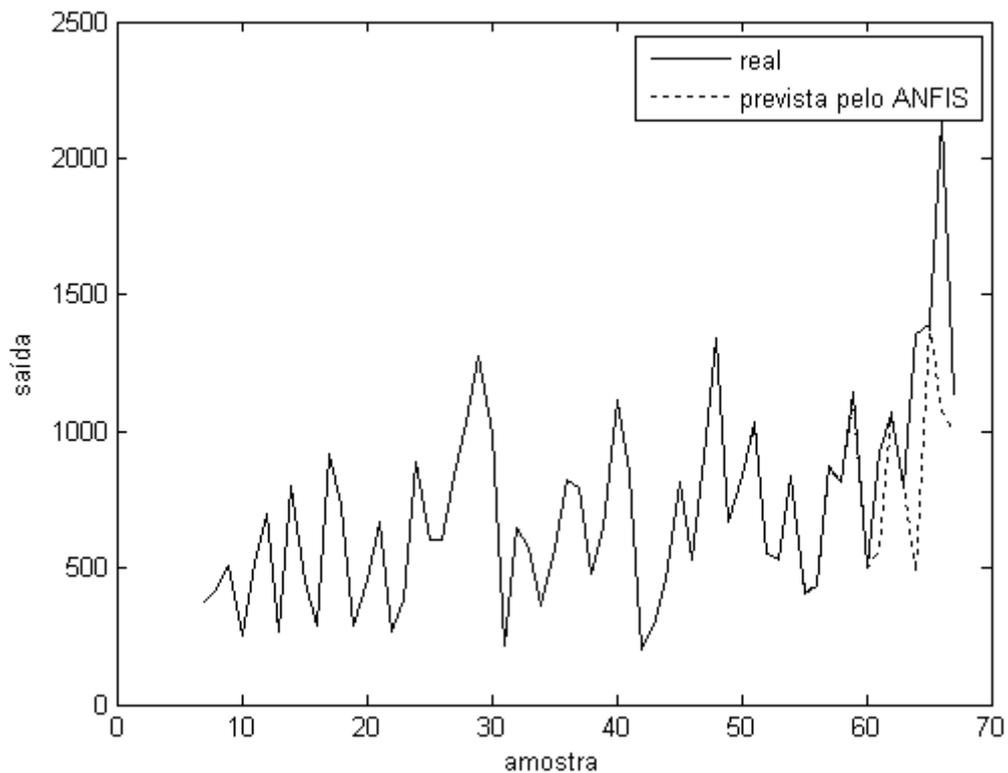


Figura 4.24 – Insumo Y versus previsão do modelo ANFIS (Simulação 3)

Fase	RMSE	MSE	MAPE	R <sup>2</sup>
<b>Estimação</b>	$3,68995 \times 10^{-1}$	$1,83008 \times 10^{-2}$	$1,49838 \times 10^{-2}$	$9,99998 \times 10^{-1}$
<b>Validação</b>	$4,31579 \times 10^2$	$5,423604 \times 10^5$	$1,54459 \times 10^1$	0
<b>Total</b>	$1,832704 \times 10^2$	$9,77261 \times 10^4$	2,797625	$7,337001 \times 10^{-1}$

Tabela 4.20 – Índices de desempenho do modelo ANFIS para o Insumo Y (simulação 3)

Comparando os índices de desempenho obtidos nas três simulações realizadas utilizando o modelo ANFIS para o Insumo Y tem-se que a simulação 1, considerando 2 funções de pertinência e a função de pertinência gaussiana se aproximou melhor dos dados da série. Comparando os valores obtidos para os índices de desempenho nas três fases consideradas em cada simulação tem-se um valor de  $R^2$  próximo a 1 e os índices RMSE, MSE e MAPE próximos a zero indicando uma melhor adequação do modelo aos dados da série considerada. Comparando-se a estrutura do modelo ANFIS para cada simulação, observa-se que a simulação 1 utilizou um número menor de parâmetros para estrutura do ANFIS.

#### 4.2.4 Análise comparativa dos modelos para o Insumo Y

Depois de realizadas as simulações com os modelos ARIMA, Holt-Winters e ANFIS, utilizando a série Insumo Y, obteve-se os índices de desempenho de cada modelo estudado, os valores encontrados são apresentados na tabela 4.21.

Índice de desempenho	ARIMA	Holt-Winters	ANFIS
<b>Estimação</b>			
RMSE	$1,93121 \times 10^2$	0	$4,844201 \times 10^{-5}$
MSE	$7,88948 \times 10^4$	0	$2,97397 \times 10^{-8}$
MAPE	$3,48196 \times 10^1$	0	$8,24642 \times 10^{-6}$
$R^2$	$5,07514 \times 10^{-1}$	1	$9,99999 \times 10^{-1}$
<b>Validação</b>			
RMSE	$4,43107 \times 10^2$	$2,82737 \times 10^2$	$7,53515 \times 10^{-5}$
MSE	$2,90954 \times 10^5$	$2,46244 \times 10^4$	$4,856515 \times 10^{-11}$
MAPE	$4,28488 \times 10^1$	$2,47333 \times 10^1$	$3,491408 \times 10^{-6}$
$R^2$	0	$4,8989 \times 10^{-1}$	$9,99999 \times 10^{-1}$
<b>Total (Estimação + Validação)</b>			
RMSE	$3,01651 \times 10^2$	$1,50325 \times 10^2$	$5,42894 \times 10^{-5}$
MSE	$6,684608 \times 10^2$	$2,61403 \times 10^2$	$2,53097 \times 10^{-8}$
MAPE	$3,96926 \times 10^1$	7,10271	$7,38896 \times 10^{-6}$
$R^2$	$2,76942 \times 10^{-1}$	$8,23043 \times 10^{-1}$	$9,99999 \times 10^{-1}$

Tabela 4.21 – Comparativo de resultados para o Insumo Y

Considerando a série Insumo Y, o modelo ANFIS obteve o melhor desempenho em todos os índices, para as três fases de previsão obtendo uma melhor aproximação da série. O modelo ANFIS considerado corresponde a simulação 1 utilizando duas funções de pertinência e a função de pertinência gaussiana. Na fase de estimação o modelo Holt-Winters obteve os melhores resultados para os índices de desempenho demonstrando uma melhor aproximação da série nesta fase.

## 5 CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS

A previsão de séries temporais tem recebido uma atenção especial 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. Antecipar o futuro e suas implicações é uma habilidade vital para as organizações num mundo dinâmico e complexo. Técnicas de previsão podem fornecer informações que permitem o controle das variáveis de negócio e alcancem os objetivos propostos pelas organizações.

Com a visão que a previsão de séries temporais é fundamental, este estudo teve como objetivo principal a comparação de modelos de previsão, tanto lineares quanto não-lineares, para duas séries temporais de consumo de dois insumos utilizados na área produtiva da empresa paranaense Beta. Foram comparados os modelos Holt-Winters, ARIMA e ANFIS utilizando os indicadores de desempenho  $R^2$ , MSE, MAPE e RMSE.

Os modelos ARIMA e Holt-Winters 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 diversas combinações foram testadas para encontrar o menor RMSE que indicou o melhor conjunto de parâmetros a serem utilizados.

Para as duas séries em estudo, o modelo ANFIS se demonstrou mais adequado obtendo os melhores índices de desempenho em todas as fases consideradas, assim o modelo ANFIS se demonstrou uma alternativa para previsões de séries temporais. O modelo ANFIS que obteve um desempenho considerável utilizou duas funções de pertinência e a função de pertinência Gaussiana.

Para trabalhos futuros, seria interessante comparar os modelos utilizados com outros modelos para previsão de séries temporais, como por exemplo, utilizar um sistema híbrido de RNAs combinado com o modelo ARIMA ou o modelo Holt-Winters. Quanto às séries de dados, um ponto a ser testado seria utilizar séries com um volume maior de observações.

## REFERÊNCIAS

ABELÉM, A. J. G. **Redes Neurais Artificiais na Previsão de Séries Temporais**. 1994. Dissertação de mestrado, Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ.

ALLEMÃO, M. A. F. **Redes Neurais Aplicadas à Previsão de Demanda de Numerário em Agências Bancárias**. 2004. Dissertação de mestrado, Mestrado em Ciência da Computação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ.

BRUSAFERRO, G. **Previsão do Mercado Automotivo Brasileiro usando Modelos Matemáticos e Inteligência Artificial**. 2006. Dissertação de mestrado, Pró-Reitoria de Pós-Graduação e Pesquisa do Instituto Tecnológico de Aeronáutica, Instituto Tecnológico de Aeronáutica, São José dos Campos, SP.

CAMPOS, R.J.; JESUS, T.A; MENDES E.M.A.M. Uma abordagem fuzzy para a previsão de curto-prazo do consumo de energia elétrica. **CNMAC - Congresso Nacional de Matemática Aplicada e Computacional**, 2007, Florianópolis - SC.

CASTRO, F.C.C; CASTRO, M.C.F. **Redes Neurais Artificiais**. Mestrado em Engenharia Elétrica, Pontifícia Universidade Católica do Rio Grande do Sul, RS.

DELGADO, M.R.B.S. **Projeto Automático de Sistemas Nebulosos: Uma abordagem Co-Evolutiva**. 2002. Tese de doutorado, Pós-Graduação da Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas, SP.

FERREIRA, B.B. **Aplicação de Ferramentas de Lógica Nebulosa à Predição de Séries Temporais**. 2008. Dissertação de mestrado, Curso de Mestrado em Engenharia Mecânica, Instituto Militar de Engenharia, Rio de Janeiro, RJ.

GOOIJER, J. G.; HYNDMAN, R.J. 25 years of time series forecasting. **International Journal of Forecasting**. v. 22. n. 3, 2006. p. 443 – 473.

GUERRA, F.A. **Análise de Métodos de Agrupamento para o treinamento de Redes Neurais de Base Radial Aplicadas à Identificação de sistemas**. 2006. Dissertação de mestrado, Programa de Pós-Graduação em Engenharia de Produção e Sistemas, Pontifícia Universidade Católica do Paraná, Curitiba, PR.

HAYKIN, S. **Redes Neurais: Princípios e Prática**. 2ª edição. Porto Alegre, RS: Bookman, 2001.

JANG, J.S.R.; SUN, C.T.; MIZUTANI, E. **Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence**. Upper Saddle River, N.J. Prentice Hall, 1997.

TRAUTWEIN, B. **Avaliação de Métodos para Previsão de Consumo de Água para Curtíssimo Prazo: um Estudo de Caso em Empresa de Saneamento**. 2004. Dissertação de mestrado, Programa de Pós-Graduação em Engenharia de Produção e Sistemas, Pontifícia Universidade Católica do Paraná, Curitiba, PR.

OLIVEIRA, W. **Estudo Comparativo entre Modelos Lineares e Redes Neurais Artificiais como Tecnologias Geradoras de Previsão de Valores Financeiros**. 2007. Dissertação de mestrado, Programa de Pós-Graduação *Stricto-Sensu* em Gestão do Conhecimento e da Tecnologia da Informação, Universidade Católica de Brasília, Brasília, DF.

KIRCHNER, R. M. **Modelagem e Previsão por meio de Metodologia Box & Jenkins: uma Ferramenta de Gestão**. 2006. Dissertação de mestrado, Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Maria, Santa Maria, RS.

LOPES, M.M. **Programação Genética para Otimização de Séries Temporais com Dados Faltantes**. 2007. Dissertação de mestrado, Programas de Pós-Graduação de Engenharia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ.

MARTÍNEZ, R. O.; ZAMPROGNO, B. Comparação de algumas técnicas de previsão em análise de séries temporais. **Revista Colombiana de Estatística**. v. 26, n. 2. 2003. p. 129–157.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. 2ª edição. São Paulo, SP: Edgard Blücher, 2006.

MORETTIN, P. A.; TOLOI, C. M. C. **Previsão de Séries Temporais**. 2ª edição. São Paulo, SP: Atual, 1987.

OLIVEIRA, G. A. **Sistema de Controle de Estoques utilizando a Metodologia Box & Jenkins de Séries Temporais**. 2002. Dissertação de mestrado, Curso de Pós-Graduação em Métodos Numéricos em Engenharia, Universidade Federal do Paraná, Curitiba, PR.

RAPOSO, R.C.T. **Investimentos em Empresas utilizando Sistemas Neuro-Nebulosos**. 2001. Dissertação de mestrado, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ.

ROCHA, F.L. **Identificação de Sistemas Não-lineares Multivariáveis usando Redes Neurais Perceptron Multicamadas e Função de Base Radial**. 2006. Dissertação de mestrado, Programa de Pós-Graduação em Engenharia de Produção e Sistemas, Pontifícia Universidade Católica do Paraná, Curitiba, PR.

RODRIGUES, M.C.; ARAÚJO, F.M.U.; MAITELLI, A.L. Controladores Neuro-Fuzzy para sistemas não-lineares. CBA - Anais do Congresso Brasileiro de Automática. 2006. Salvador, BA

SANDMANN, H.R. **Predição não-linear de séries temporais usando sistemas de arquitetura neuro-fuzzy**. 2006. Dissertação de mestrado, Escola Politécnica da Universidade de São Paulo, São Paulo, SP.

SOUTO, D. P.; BALDEÓN, R. A.; RUSSO, S. L. Estudo dos modelos exponenciais na previsão. **Industrial Data**. vol. 9. 2006. p. 97-103.

SOUZA, R. C.; CAMARGO, M. E. **Análise e Previsão de Séries Temporais: Os Modelos ARIMA**. 2ª edição. Rio de Janeiro, RJ: Gráfica e Editora Regional, 2004.

SOUZA, R.C.T. **Previsão de séries temporais utilizando Rede Neural treinada por Filtro de Kalman e Evolução Diferencial**. 2008. Dissertação de mestrado, Programa de Pós-Graduação em Engenharia de Produção e Sistemas, Pontifícia Universidade Católica do Paraná, Curitiba, PR.

TEIXEIRA, R.A; CORRÊA M.V; ASSIS, J.C.O. Redes Neuro Fuzzy: Algoritmo de Treinamento e Aplicações. **Principium Online: Iniciação Científica no Unileste-MG**. v.1. n. 3. p. 152-168. dez. 2007.

VALENZUELA, O.; ROJAS, I.; ROJAS, F.; POMARES, H.; HERRERA, L.J.; GUILLEN, A.; MARQUEZ, L.; PASADAS, M. Hybridization of Intelligent Techniques and ARIMA Models for Time Series Prediction. **Fuzzy Sets and Systems**. 159. 2008. p. 821 – 845.

ZANDONADE, E. **Aplicação da Metodologia de Redes Neurais em previsão de Séries Temporais**. 1993. Dissertação de mestrado, Mestrado em Ciências de Engenharia Elétrica: Teoria de Controle e Estatística, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ.

ZHANG, G. P. Time Series Forecasting using a Hybrid ARIMA and Neural Network model. **Neurocomputing**. vol 50. 2003. p. 159 – 175.