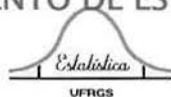




UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE MATEMÁTICA  
DEPARTAMENTO DE ESTATÍSTICA



# **Aplicação de Redes Neurais Artificiais à Previsão de Vendas de Máquinas Agrícolas – Um Estudo de Caso junto a AGCO do Brasil Ltda.**

**Autor: João Henrique Ferreira Flores  
Orientadora: Professora Dra. Liane Werner**

Porto Alegre, 13 de Dezembro de 2006.

Universidade Federal do Rio Grande do Sul  
Instituto de Matemática  
Departamento de Estatística

# Aplicação de Redes Neurais Artificiais à Previsão de Vendas de Máquinas Agrícolas – Um Estudo de Caso junto a AGCO do Brasil Ltda.

Autor: João Henrique Ferreira Flores

Monografia apresentada para obtenção  
do grau de Bacharel em Estatística.

Banca Examinadora:

Professora Dra. Liane Werner (orientadora)

Professor Dr. Flávio Augusto Ziegelmann (convidado do Departamento de  
Estatística da UFRGS)

Porto Alegre, 13 de Dezembro de 2006.

UFRGS  
SISTEMA DE BIBLIOTECAS  
BIBLIOTECA SETORIAL DE MATEMÁTICA

Dedico este trabalho à minha companheira de todas as horas, Patrice Z. Basso.

*“Tudo sempre acaba bem. Se ainda não está bem, é porque ainda não acabou.”*  
*Fernando Sabino.*

## **AGRADECIMENTOS**

Agradeço aos meus pais, pelo auxílio e compreensão nos momentos difíceis, à minha irmã e meu sobrinho pelos momentos de descontração, assim como aos meus amigos.

Agradeço também a minha orientadora, professora doutora Liane Werner e aos meus colegas pelo apoio técnico, científico e sabedoria que auxiliaram neste trabalho.

Acima de tudo, agradeço aos meus avós, dentre eles, em especial, a Dona Alani, pelo esforço dedicado a mim e aos meus familiares e ao meu amor, Patrice Zeidler Basso, que foi a companheira de todas as horas e uma motivadora sem igual. Sem estas pessoas este trabalho, provavelmente, não existiria.

## RESUMO

A previsão de vendas para implementos agrícolas é, reconhecidamente, complexa devido as mais diversas variáveis que podem afetar um modelo. Além disso, diferentes modelos de implementos requerem diferentes modelos de previsão. Sendo a empresa AGCO do Brasil Ltda. líder de vendas neste segmento, optou-se por realizar previsões para o modelo de trator MF 275, da marca Massey Ferguson, líder de vendas, além disto outro fator motivacional para esta escolha, foi a facilidade de obtenção de informações necessárias, fornecidas pela empresa. Para a análise da série de vendas, obtida junto a Associação Nacional dos Fabricantes de Veículos Automotores (ANFAVEA), foram desenvolvidos modelos de redes neurais artificiais (RNA). Dentre os modelos de RNA construídos, o melhor é o modelo que considera apenas as vendas do mês anterior e que possui 3 neurônios na camada oculta. Ao comparar este modelo final com o que a empresa utiliza é possível observar que o modelo proposto mostrou-se melhor quanto aos erros de previsão. O modelo final apresentado se destaca, também, porque não depende de variáveis externas. O modelo utiliza apenas a própria série para as previsões

Palavras-chave: previsão de vendas, seleção de modelos, redes neurais artificiais, modelos de previsão.

## ABSTRACT

The sales prediction to agricultural equipment is, admittedly, complex thanks to a variety of variables that can affect a model. Moreover, different models of agricultural equipment need different prediction models. The reason to choose the AGCO do Brasil Ltda. company and the MF 275 tractor, of the Massey Ferguson trademark, to predict is due to its leadership in sales and the easiness to obtain the information needed, supplied by the company. To the analysis of the sales series, that was obtained with the Associação Nacional dos Fabricantes de Veículos Automotores (ANFAVEA), some Artificial Neural Network (ANN) models were developed. Among the ANN models built, the best model is the model that considers only the sales of the previous month and had only 3 neurons on the hidden layer. By comparing this final model with the model that the company uses, it's possible to observe that the proposed model appeared better as for the prediction errors. The final model presented stands out, too, because it doesn't depend on external variables. The model only uses the sales series to forecast.

Keywords: sales forecast, model selection, artificial neural network, tractor sales, AGCO.

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>11</b>
1.1	COMENTÁRIOS INICIAIS .....	11
1.2	OBJETIVOS .....	12
1.3	JUSTIFICATIVAS .....	13
1.4	METODOLOGIA DE PESQUISA .....	14
1.5	ESTRUTURA .....	16
<b>2</b>	<b>REFERENCIAL TEÓRICO .....</b>	<b>17</b>
2.1	PREVISÃO DE DEMANDA .....	17
2.2	SÉRIE TEMPORAL .....	19
2.3	REDES NEURAIS ARTIFICIAIS .....	20
2.3.1	<i>Neurônio de McCulloch</i> .....	21
2.3.2	<i>Perceptron</i> .....	23
2.3.3	<i>Princípio de Aprendizado de Hebb</i> .....	26
2.3.4	<i>ADALINE e MADALINE</i> .....	30
2.3.5	<i>Retropropagação</i> .....	31
2.3.6	<i>Redes Neurais Artificiais na análise de séries temporais</i> .....	33
2.3.7	<i>Redes Neurais Artificiais e a Estatística</i> .....	35
<b>3</b>	<b>ESTUDO DE CASO .....</b>	<b>36</b>
3.1	APRESENTAÇÃO DA EMPRESA .....	36
3.2	MODELO DA PRODUÇÃO DA EMPRESA .....	38
3.3	MÉTODO DE PREVISÃO DE VENDAS ATUALMENTE UTILIZADO .....	39
3.4	APRESENTAÇÃO DO PRODUTO ESCOLHIDO .....	40
3.5	APRESENTAÇÃO DA SÉRIE DE VENDAS .....	41
3.6	ANÁLISE PRÉVIA DA SÉRIE .....	42
3.7	APLICAÇÃO DA TÉCNICA DE REDES NEURAIS ARTIFICIAIS .....	43
3.8	ESCOLHA DO MELHOR MODELO E AS PREVISÕES .....	49
	<b>CONCLUSÃO .....</b>	<b>54</b>
	<b>REFERÊNCIAS .....</b>	<b>56</b>



## LISTA DE FIGURAS

Figura 1 - Separação do espaço euclidiano .....	22
Figura 2 - Representação dos funções booleanas .....	23
Figura 3 - Rede de neural com dois neurônios na camada oculta .....	24
Figura 4 - Marcas que a AGCO representa .....	37
Figura 5 - Processo de produção da AGCO de forma simplificada .....	39
Figura 6 - Gráfico da série de vendas dos tratores MF 275 por atacado .....	42
Figura 7 - Saída do teste de Teraesvirta no R.....	43
Figura 8 - Modelo com 6 neurônios na camada oculta.....	47
Figura 9 - Modelo com 3 neurônios na camada oculta.....	47
Figura 10 - Modelo com 4 neurônios na camada oculta.....	48
Figura 11 - Modelo com 6 neurônios na camada oculta.....	48
Figura 12 - Gráfico com as previsões do modelo $X(t-1)$ , $X(t-12)$ e indicadoras.....	50
Figura 13 - Gráfico com as previsões do modelo $X(t-1)$ .....	50
Figura 14 - Gráfico com as previsões do modelo $X(t-1)$ , $X(t-12)$ e $X(t-13)$ .....	51
Figura 15 - Gráfico com as previsões do modelo $X(t-1)$ e $X(t-12)$ .....	51
Figura 16 - Gráfico do modelo $X(t-1)$ e $X(t-12)$ já com as previsões, com 2 passos a frente.....	53
Figura 17 - Gráfico do modelo $X(t-1)$ já com as previsões, com 2 passos a frente.....	53

## LISTA DE TABELAS

Tabela 1 - Especificações Técnicas do MF 275 .....	41
Tabela 2 – Unidades vendidas de tratores MF 275 por atacado.....	42
Tabela 3 - Critérios AIC dos modelos.....	45
Tabela 4 - Critérios BIC dos modelos .....	46
Tabela 5 - Previsões e erros dos modelos selecionados.....	49
Tabela 6 - Erros e previsões dos dois modelos escolhidos .....	52
Tabela 7 - Critérios de seleção dos modelos .....	52

# 1 INTRODUÇÃO

## 1.1 Comentários Iniciais

Desde os primeiros adivinhos que se busca uma boa previsão do futuro. É da natureza humana tentar prever o que lhe vai acontecer. Nas organizações isso não é diferente. Dado o nível de concorrência atual, em que empresas competem entre si em um mercado cada vez mais sem fronteiras, a previsão se tornou algo fundamental. Seja no campo da produção de produtos ou na produção de serviços, a capacidade de prever a demanda, de forma aceitável, auxilia os técnicos e gerentes a planejarem melhor o futuro da organização. E quanto melhor o planejamento, mais fácil fica o controle da produção e mais ágil se torna a organização, tornando-se, desta forma, mais competitiva.

Dentre os muitos usos da análise de séries temporais, destaca-se a previsão. Entretanto a possibilidade de previsão não é única à análise de séries temporais. Modelos de regressão também são usados para previsões. Previsão esta, que é utilizada para os mais diversos fins. Um destes é a de previsão da demanda nas organizações. A previsão da demanda é, mais comumente, feita com base em dados quantitativos, como vendas anteriores, preço do produto, entre outros. Mas também existem os modelos de previsão com base em dados qualitativos, como por exemplo o método Delphi.

Para os modelos quantitativos, por outro lado, existem diversos tipos de modelagens possíveis, desde a chamada modelagem clássica até métodos mais avançados como Redes Neurais e modelos Bayesianos. Cada qual possui suas características, vantagens e desvantagens específicas. Muitos destes métodos se desenvolveram graças aos avanços tecnológicos dos computadores, que hoje permitem que modelos tidos anteriormente como complexos tenham se tornado um tanto quanto rotineiros e razoavelmente mais ágeis.

E é, principalmente devido ao avanço tecnológico, que os modelos de previsão estão se tornando mais comuns ao uso não acadêmico. Eles têm sido muito utilizados em conjunto com outras ciências, como Economia e Tecnologia da Informação. Entretanto, no foco da produção, principalmente de produtos mais do que serviços, estes modelos ainda são razoavelmente desconhecidos. Muitas organizações preferem basear suas previsões muito mais na experiência de seus gerentes do que em modelos estatísticos.

No caso do ramo de máquinas e implementos agrícolas a previsão de demanda pode ser mais complexa, o que se pode notar pela falta de artigos e textos que tratem especificamente sobre este tema. Tal complexidade pode ser atribuída tanto ao fato de existirem as mais diversas variáveis que podem afetar as vendas deste tipo de produto, ou talvez, pela natureza do comportamento de algumas variáveis, como o clima, por exemplo. Aqui o papel do estatístico não se resume a lidar apenas com os dados de uma série temporal, como vendas, mas sim tentar identificar, junto a especialistas de outras áreas, informações capazes de auxiliar na construção de um modelo de previsão.

## 1.2 Objetivos

O objetivo principal deste trabalho é construir um modelo de previsão de rede neural para um produto específico da AGCO do Brasil Comércio e Indústria Ltda., empresa, esta, onde será realizada a aplicação prática. O modelo a ser encontrado deve minimizar os erros de previsão de vendas(demanda) em relação ao modelo atualmente utilizado na empresa e deve ser de uso prático, estando de acordo com a realidade da organização.

Além do objetivo geral, tem-se como objetivo específico, relatar o atual modelo utilizado pela organização, observando suas características e obter maiores informações a respeito do produto a ser previsto.

### 1.3 Justificativas

Segundo Slack et al. (2002, p. 86), “nenhuma organização pode planejar pormenorizadamente todos os aspectos de suas ações atuais e futuras, mas todas as organizações podem beneficiar-se de ter noção para onde estão dirigindo-se e de como podem chegar lá, o mesmo ocorrendo com a função produção”. A previsão de demanda é imperativa para que as organizações tenham vantagens competitivas, além é claro, da sua importância para o planejamento e controle da produção.

Mas, a importância da previsão de demanda vai além de auxiliar apenas o controle e planejamento da produção. Kotler (1998, p. 129) caracteriza melhor quando diz:

As previsões de vendas são usadas para o departamento de finanças levantar o dinheiro necessário aos investimentos e operações; para o departamento de manufatura definir os níveis de capacidade e produção de bens; para o departamento de compras adquirir o volume correto de suprimentos; e para o departamento de recursos humanos contratar o número de operários necessários.

Um dos mais importantes fatores, entretanto, é a capacidade das organizações, conhecendo as previsões, reduzirem o tempo de entrega do produto ou serviço ao cliente. A partir do momento que a empresa possui a previsão da demanda, ela pode se preparar antecipadamente, ou seja, antecipar os desejos de seus clientes. De acordo com Slack et al. (2002, p. 320-322) este modo de planejamento e controle é do tipo fazer-para-estoque (MTS, do original em inglês *Make To Stock*). Segundo alguns autores, o planejamento e controle da produção com previsão de demanda auxiliam também na redução de custos, pois contribui no controle de estoques (FERNANDES; SANTORO, 2005; PASQUAL; PEDROZO, 2005 e WANKE, 2005). Com uma previsão de demanda mais precisa, os estoques são mais bem controlados e com isso, se reduz também o tempo até a entrega do produto ou serviço. O prazo de entrega é importante para todo tipo de organização, tanto indústrias quanto serviços. Entretanto, para algumas organizações este prazo é fundamental. Como, por exemplo, nos serviços prestados por hospitais. No caso de uma cirurgia, o tempo é vital.

Um outro exemplo a respeito do prazo de entrega é o de empresas do ramo de agronegócios. Cada cultura tem seu momento de plantio e colheita bem definidos.

Um atraso na entrega de sementes ou implementos pode prejudicar, ou até inviabilizar, a plantação e colheita. Muitas vezes este atraso se deve a fatores externos, como condições climáticas. Mas muito também ocorre por falta de planejamento de ambas as partes, tanto do produtor agrícola como das empresas e indústrias.

O ramo de máquinas e implementos agrícolas foi escolhido devido a possível complexidade envolvida na construção de um modelo de previsão, tornando o trabalho mais interessante do ponto de vista científico. Segundo o especialista, responsável pelo planejamento e controle da produção na AGCO, variáveis de modelagem reconhecidamente complexa, como clima, são considerados fatores importantes para a construção de um modelo de previsão.

Quanto à organização, a AGCO foi escolhida tanto pela sua variedade de produtos ofertados como pela disponibilidade e acessibilidade dos dados e informações. E, também ainda, por pertencer ao ramo de mercado estudado. Além disso, segundo informações obtidas junto à empresa, os erros nas previsões da AGCO estão muito além do que se espera. De acordo com o especialista, a AGCO erra, em média, suas previsões em 50% (erro absoluto percentual) do valor real. Erros dessa magnitude geram um alto custo devido a ajustes que devem ser feitos no curto prazo. Além disso, pode causar sérios problemas frente aos fornecedores, que necessitam suprir matéria-prima em um prazo muito curto, prejudicando a qualidade.

A modelagem por Redes Neurais Artificiais (RNA) foi escolhida por ser uma técnica de modelagem muito utilizada atualmente e que possui uma vasta gama de usos, desde previsão até classificação, sendo este último, o uso mais antigo.

## **1.4 Metodologia de Pesquisa**

De acordo com a definição encontrada em Gil (2002, p.41-55), o foco principal deste trabalho trata-se de uma pesquisa exploratória, pois busca proporcionar uma maior familiaridade com o problema da previsão de vendas.

Quanto aos procedimentos técnicos a serem utilizados neste trabalho é tanto um estudo de caso como uma pesquisa *ex-post facto*. Isto se deve ao fato de que, mesmo que a amplitude e aprofundamento do trabalho o caracterizem como um estudo de caso, a busca de dados é feita de duas formas bem distintas. Inicialmente será realizada uma entrevista com um especialista da empresa, buscando identificar tanto o modelo atualmente utilizado pela empresa, assim como informações sobre variáveis externas a empresa que possam ser úteis na modelagem dos dados. Desta forma caracterizando o trabalho como um estudo de caso. Entretanto, a obtenção dos dados, inclusive das variáveis citadas pelo especialista, será obtida de dados que já ocorreram, uma série temporal discreta. Assim sendo, esta também é uma pesquisa *ex-post facto*. Resumidamente, quanto ao escopo do trabalho, este é um estudo de caso. Quanto à forma de obtenção dos dados é uma pesquisa *ex-post facto*. Dados estes, que são classificados como uma série temporal discreta.

Para a construção do modelo de previsão, será utilizada uma série temporal discreta das vendas, obtida junto a Associação Nacional dos Fabricantes de Veículos Automotores (ANFAVEA) e verificados junto a AGCO do Brasil, que possui dados sobre vendas de periodicidade mensal desde janeiro de 1999 até agosto de 2006. Quanto à coleta de dados das variáveis que possam influenciar o modelo, a busca será feita em bancos de dados governamentais, como o Instituto Brasileiro de Geografia e Estatística (IBGE), caso elas venham a ser utilizadas.

A análise dos dados e a previsão de demanda serão feitas utilizando métodos de análise de séries temporais por redes neurais artificiais. Esta técnica foi utilizada no pacote estatístico R de livre distribuição, com base no pacote *nnet* e com a função de mesmo nome. O modelo a ser recomendado será escolhido entre os modelos apresentados posteriormente. Sempre tendo em mente que a busca é por um modelo que seja de uso prático para a empresa, o mais preciso quanto as suas previsões e correto e válido do ponto de vista estatístico.

## 1.5 Estrutura

O primeiro capítulo desta monografia consiste de uma introdução ao assunto da previsão de demanda, dos objetivos principal e secundário e das justificativas a escolha deste tema. A metodologia a ser utilizada na busca dos objetivos também é apresentada. Além deste capítulo introdutório, esta monografia contém mais três capítulos.

O capítulo 2 apresenta o referencial teórico em três partes. A primeira parte apresenta uma breve introdução à previsão de demanda. O mesmo ocorre na segunda parte em relação a séries temporais. Na terceira parte, é abordada a técnica de redes neurais artificiais, com suas definições e usos tanto na análise de séries temporais como na Estatística de uma forma geral.

No capítulo 3 encontra-se o estudo de caso, onde serão apresentados: os dados e os passos necessários para realizar a modelagem, o passo a passo. Neste capítulo estão os cálculos e testes do modelo, além das previsões e da descrição teórica do problema a ser analisado.

Finalmente, no capítulo 4 encontram-se os comentários finais. Neste capítulo são apresentadas as vantagens e desvantagens do método utilizado, a conclusão e recomendações para estudos posteriores.



## 2 REFERENCIAL TEÓRICO

### 2.1 Previsão de Demanda

Kotler (1998, p. 132) define demanda de mercado para um produto como “o volume total que seria comprado por um grupo definido de consumidores em determinada área geográfica, em período de tempo definido, em ambiente de marketing definido, sob determinado programa de marketing”. Esta definição está mais ligada ao marketing envolvido na demanda de um produto. Em um foco mais voltado ao planejamento e controle da produção, Slack et al. (2002) caracteriza que a previsão de demanda tem uma função estratégica para a organização, auxiliando na programação da produção junto a fornecedores e funcionários e também para o planejamento e controle dos produtos.

De acordo com Kotler (1998, p. 129), “a demanda pode ser mensurada por seis *níveis de produto*, cinco *níveis de espaço* e três *níveis de tempo* diferentes” (grifo do autor). Os níveis de produto citados pelo autor são: venda total, venda do setor industrial, venda da empresa, linha de produtos, forma de produtos e item de produto. Os níveis de espaço são: mundo, país, região, território e consumidor. E como nível de tempo tem: curto, médio e longo prazo. Entretanto, segundo Slack et al. (2002, p. 315), a definição do espaço de tempo para o planejamento define, de antemão, o nível de produto. Diz o autor:

No longo prazo, os gerentes de produção fazem planos relativos ao que eles pretendem fazer, que recursos eles precisam e quais objetivos atingir. A ênfase está mais no planejamento do que no controle [...]. Eles vão usar previsão de demanda provável, descritas em termos agregados [nos níveis de Kotler, isso equivale à venda total]. O planejamento e controle de médio prazo estão preocupados com planejar em mais detalhes [...] para avaliar a demanda global que a operação deve atingir de forma mais desagregada. No planejamento e no controle de curto prazo, muitos dos recursos terão sido definidos [...]. Neste estágio, a demanda será avaliada de forma totalmente desagregada [neste caso, a equivalência seria na forma de produtos].

Além das definições acima mencionadas, existem ainda definições quanto à incerteza da demanda. Slack et al. (2002, p. 317-321) define dois tipos claros de demanda: a demanda independente e a demanda dependente. Estas demandas podem ser definidas também pela razão P:D (SHINGO, 1981 e MATHER, 1988). A razão P:D é a razão entre o tempo total do processo P e o tempo de demanda D. O tempo total P é o tempo que a operação leva para obter recursos, produzir e entregar o produto ou serviço. O tempo de demanda D depende do tipo de operação. A demanda dependente é quando P e D são muito próximos, ou seja, trata de processos de fácil previsão (tornando-a, muitas vezes, até desnecessária). Como exemplo, as operações sob encomenda, no qual a produção só se inicia depois de recebido o pedido. A demanda independente, por outro lado, é quando P é muito superior a D. Neste tipo de demanda, o processo se inicia baseado, geralmente, em previsões de pedidos, antecipando as necessidades do consumidor. Ainda segundo Slack et al. (2002, p. 322), as razões P:D indicam o grau de especulação da operação, ou seja, “quanto maior for P comparado com D, maior é a proporção de atividades especulativas na operação e, portanto, maior é o risco da operação”.

Segundo Frees e Miller (2004), a “previsão é parte integral do gerente de marketing”. Dizem os autores ainda que a “previsão de vendas é importante para entender o mercado e a competição, e alguns determinantes de vendas, incluindo promoções, tática de preços, propagandas e distribuição”. Tal definição é partilhada por alguns autores. Segundo Kuyven (2004), “a preparação da previsão de demanda é, normalmente, uma ação do departamento de marketing da empresa”.

Como se vê, a previsão de demanda é de suma importância para as organizações. De forma geral, uma boa previsão, ou seja, uma previsão mais acurada auxilia a organização como um todo, desde estratégias de vendas, até a parte da produção e da qualidade. E isso afeta não só a organizações, como também seus parceiros, sejam eles fornecedores ou clientes. Para os clientes por exigir um tempo menor de espera para entrega de produtos ou serviços. Com os fornecedores as vantagens vão desde prazos para entrega de matéria prima até o controle de estoques. Com isso a organização ganha em custos e em competitividade.

## 2.2 Série Temporal

Para Morettin e Toloí (2006), uma série temporal é qualquer conjunto de observações ordenadas no tempo. Ainda segundo os autores,

Um modelo que descreve uma série não conduz, necessariamente, a um procedimento (ou fórmula) de previsão. Será necessário especificar uma função-perda, além do modelo, para se chegar ao procedimento. [...] É importante salientar que a previsão não constitui um fim em si, mas apenas um meio de fornecer informações para uma conseqüente tomada de decisões, visando a determinados objetivos.

Vale ressaltar que “a palavra previsão sugere que se quer uma coisa antes que ela exista. Alguns autores preferem utilizar a palavra predição, para indicar algo que deverá existir no futuro. Ainda outros utilizam o termo projeção”. (MORETTIN; TOLOI, 2006).

Outros autores, entretanto, definem uma série temporal de maneira diferente. Segundo Shumway (1988, p. 1),

A análise de dados experimentais que tenham sido observados em diferentes pontos no tempo induziu a novos e únicos problemas para modelagem estatística e inferência. A óbvia correlação introduzida pela amostragem em pontos adjacentes no tempo pode restringir severamente a aplicabilidade da maioria dos métodos estatísticos convencionais [...]. A avaliação sistemática pela qual busca-se responder sobre as questões matemáticas e estatísticas apresentadas por estas correlações no tempo é comumente referida a *análise de séries temporais* (grifo do autor).

Ainda segundo o autor, a utilidade da análise de séries temporais pode ser mensurada pelos mais diversos campos de conhecimento que a utilizam, sendo seu uso mais comum nas ciências econômicas.

Anderson (1971, p. 1) segue a mesma linha de Shumway (1998) e Morettin e Toloí (2006), mas amplia o conceito. De acordo com o autor, “uma série temporal é uma seqüência de observações, usualmente ordenada no tempo, apesar de, em alguns casos, pode ser ordenada de acordo com outra dimensão”. O autor também destaca as correlações entre as observações e ainda uma característica especial da análise de séries temporais, que a diferencia de outras análises estatísticas: “o explícito reconhecimento da importância da ordem que as observações são feitas”. O autor ainda cita exemplos de fenômenos que se modificam ao longo do tempo, como aspectos climáticos, preços e a saúde de um indivíduo.

Enders (1995, p. 2) apresenta um breve histórico de séries temporais, quando diz que “o uso original de análise de séries temporais era, primeiramente, como ajuda para a previsão”.

A previsão, ou predição é um dos principais, senão o principal, interesse na análise de série temporais. Como dito anteriormente, a possibilidade de predizer o futuro com base no passado já é suficiente para chamar a atenção. Mas problemas práticos têm exigido cada vez mais dos modelos, tanto em termos da própria modelagem como para as previsões.

### 2.3 Redes Neurais Artificiais

O princípio de redes neurais, a neurocomputação, aparece pela primeira vez em 1943, num trabalho de McCulloch e Pitts. Os autores sugeriram a construção de uma máquina inspirada, no cérebro humano. Esta inspiração é considerada como a base dos primeiros conceitos para o estudo de redes neurais artificiais, de onde foi apresentando o modelo inicial de neurônio artificial denominado Psychon. Desde então muito foi feito no desenvolvimento de redes neurais artificiais. Rosenblatt, em 1959 foi quem idealizou o modelo básico do Perceptron que é utilizado até hoje.

A base de estudo das redes neurais artificiais (RNA) são as redes neurais naturais (RNN). A mais conhecida rede neural natural é a do cérebro humano que é composto por bilhões de neurônios que se conectam entre si através de sinapses. O sistema nervoso humano também é formado por um conjunto extremamente complexo de neurônios. Todas as funções e movimentos do organismo estão relacionados ao funcionamento dos neurônios. Desta forma, todos os impulsos externos, como: imagens, sons, e outros, são recebidos e tratados para oferecer uma resposta adequada. É baseado neste funcionamento que uma rede neural artificial funciona. Logicamente, não é apenas isso.

### 2.3.1 Neurônio de McCulloch

O neurônio de McCulloch é um dispositivo binário no qual a saída se classifica como pulso e não pulso e as várias entradas tem um ganho aleatório e podem ser excitativas ou inibidoras. Para determinar a saída do neurônio é calculada a soma ponderada das entradas com os respectivos ganhos como fatores de ponderação. Estes fatores são positivos no caso de ganhos excitativos e negativos quando os ganhos são inibitórios. Se o resultado for maior ou igual a um dado limite, então a saída do neurônio é pulso, caso contrário a saída é não pulso. Este neurônio pode ser modelado por um caso particular de discriminador linear cujas entradas são binárias. Genericamente, um discriminador linear de  $n$  entradas  $\{x_1, x_2, \dots, x_n\}$  e uma saída  $y$  é definido pela expressão (1):

$$\begin{aligned} y &= H\left(\sum_{i=1}^n w_i x_i - \Theta\right) = H(w'x - \Theta) \rightarrow y \in [0;1] \\ y &= \text{sgn}\left(\sum_{i=1}^n w_i x_i - \Theta\right) = \text{sgn}(w'x - \Theta) \rightarrow y \in [-1;1] \end{aligned} \quad (1)$$

onde os componentes do vetor  $w$ ,  $\{w_1, w_2, \dots, w_n\}$  são os ganhos associados às entradas  $x_i$ ,  $\Theta$  é o valor do limiar,  $H(v)$  é a função degrau unitário e  $\text{sgn}(v)$  o operador sinal.

A expressão (1) representa um hiperplano que divide o espaço euclidiano  $\mathfrak{R}^n$ , de dimensão  $n$ , em duas regiões, A e B. Assim um vetor  $x$  de componentes  $\{x_1, x_2, \dots, x_n\}$  estará em uma destas regiões na medida em que se verificar:

$$\begin{cases} w'x - \Theta > 0 \Rightarrow x \in A \\ w'x - \Theta < 0 \Rightarrow x \in B \end{cases} \quad (2)$$

enquanto que o valor da saída  $y$  será:

$$y = 1 \text{ se } x \in A \text{ e } y = 0 \text{ ou } -1 \text{ se } x \in B, \quad (3)$$

Esta situação é representada pelo diagrama da Figura 1, para o caso em que a dimensão do espaço é  $n = 2$ , ou seja, no plano euclidiano.

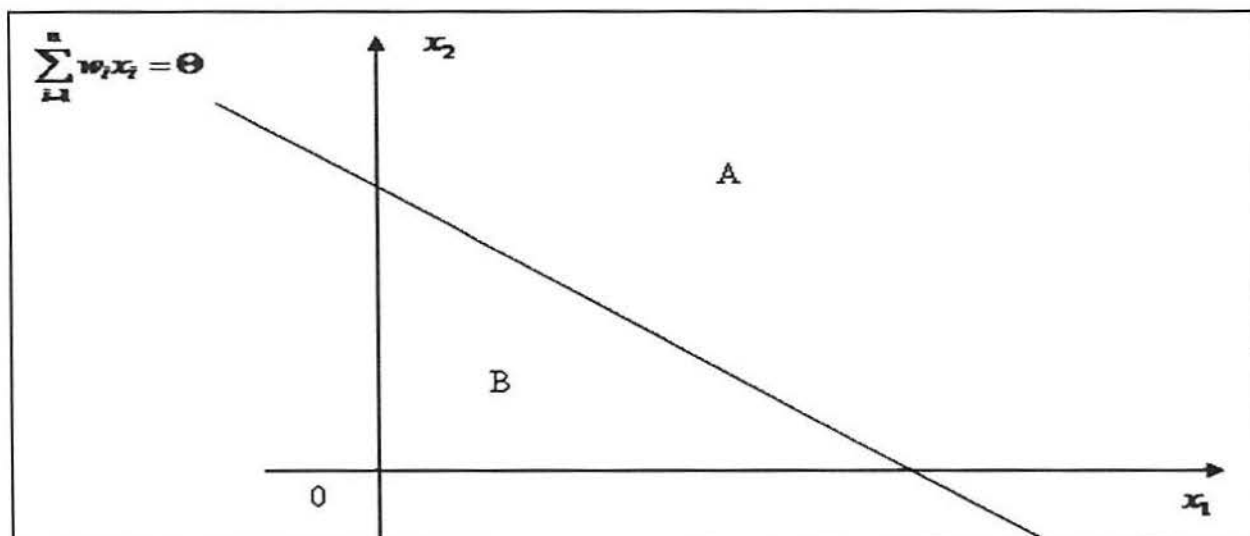


Figura 1 - Separação do espaço euclidiano  
 Fonte: KOVÁCS, 1995.

Em vista deste comportamento o discriminador linear pode ser utilizado, em certos casos, como um classificador de padrões ou separador de aglomerados de pontos. Como ilustração, sejam duas coleções  $\Theta_1 = \{u_1, u_2, \dots, u_k\}$  de  $k$  vetores  $n$ -dimensionais e  $\Theta_2 = \{z_1, z_2, \dots, z_m\}$  de  $m$  vetores  $n$ -dimensionais. Supondo que se pretende implementar um discriminador linear que separe estas duas coleções de vetores no sentido da expressão (3), isto é:

$$y = 1 \text{ se } x \in \Theta_1 \text{ e } y = 0 \text{ ou } -1 \text{ se } x \in \Theta_2. \quad (4)$$

Naturalmente isto só será possível se as coleções  $\Theta_1$  e  $\Theta_2$  formarem aglomerados no espaço  $\mathfrak{R}^n$ , tal que seja possível passar um hiperplano:

$$\sum_{i=1}^n w_i x_i = w^r x = \Theta, \quad (5)$$

que separe os dois aglomerados. Se as coleções  $\Theta_1$  e  $\Theta_2$  forem tais que isto é possível, então são ditas coleções linearmente separáveis. Caso contrário, não serão separáveis linearmente e não existirá nenhum discriminador linear capaz de executar esta função.

O neurônio de McCulloch é um discriminador linear com a simplificação adicional de que os componentes do vetor de entrada  $x$  também são binários, isto é,

$x_i \in \{0,1\}$ . Na Figura 2 estão representadas três funções booleanas: E, OU e XOR de duas variáveis.

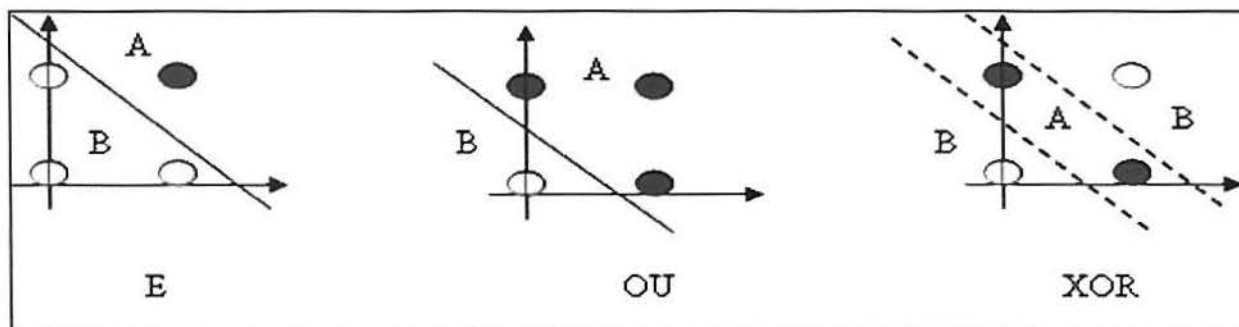


Figura 2 - Representação dos funções booleanas  
Fonte: KOVÁCS, 1995

Pela figura nota-se que as funções booleanas E e OU são linearmente separáveis e é possível encontrar retas que separem os vetores de entrada em regiões distintas, na qual a saída reflete corretamente o resultado destas funções. Portanto, estas funções podem ser usadas no neurônio de McCulloch. Fica claro, também, que a função XOR, o *ou exclusivo*, não é linearmente separável. Para separá-la são necessárias duas retas discriminatórias. Para as 16 funções booleanas de duas variáveis, somente duas (XOR e seu complemento) não são linearmente separáveis. Na medida em que a dimensão aumenta, o número de funções que não são linearmente separáveis também aumenta.

### 2.3.2 Perceptron

Embora um único neurônio não consiga implementar todas as funções booleanas, sempre existirá alguma rede de múltiplos neurônios que implementará qualquer função booleana. Isto é demonstrado observando-se que qualquer função booleana  $f(x_1, x_2, \dots, x_n)$  de  $n$  variáveis pode ser escrita como sendo:

$$f(x_1, x_2, \dots, x_n) = \bigcup_{k=1}^{2^n} \alpha_k \left( \prod_{i=1}^n (\beta_i x_i \cup (1 - \beta_i) x_i^c) \right), \text{ com } \alpha_k \in [0;1] \text{ e } \beta_k \in [0;1] \quad (6)$$

onde  $\cup$  representa uma soma booleana,  $\prod$  os produtos booleanos das  $2^n$  combinações possíveis dos  $x_i$  e seus complementos  $x_i^c$ .

A soma booleana e o produto booleano são duas das funções implementáveis com o neurônio de McCulloch, como visto anteriormente. Desta forma, o XOR de duas variáveis do exemplo anterior pode ser implementado por uma rede de três neurônios  $N_1$ ,  $N_2$  e  $N_3$ , com  $N_1$  e  $N_2$  implementando as funções  $f_1 = x_1 x_2^c$  e  $f_2 = x_1^c x_2$  e  $N_3$  implementando a função  $f_3 = f_1 \cup f_2$ , conforme Figura 3. Para esta arquitetura de rede é dado o nome de perceptron.

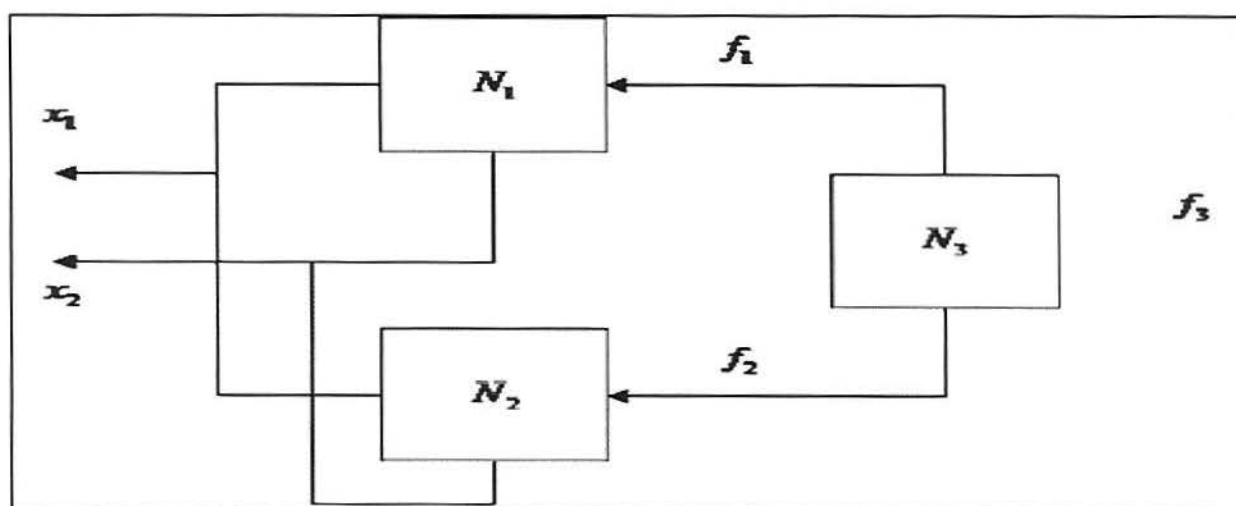


Figura 3 - Rede de neural com dois neurônios na camada oculta  
Fonte: KOVÁCS, 1995.

Foi no final da década de 1950 que Rosenblatt, na Universidade de Cornell, dando prosseguimento às idéias de McCulloch, criou uma rede de múltiplos neurônios do tipo discriminadores lineares (HAYKIN, 2001). Esta rede recebeu o nome de perceptron. Os neurônios que recebem diretamente as entradas da rede são a *camada de entrada*. Os neurônios que recebem como entrada as saídas dos neurônios da camada de entrada constituem a segunda camada e assim por diante, até a camada final que é a *camada de saída*. As camadas internas que não são nem a de entrada e nem a de saída são chamadas de *camadas ocultas*.

Uma rede neural multicamada de  $K$  camadas terá um vetor de entrada  $x_0$ , de dimensão  $J_0$  e de componentes  $x_{0,j}, j = 1, 2, \dots, J_0$ . Estas entradas são conectadas às entradas dos  $J_1$  neurônios da camada de entrada. As saídas  $x_{1,j}, j = 1, 2, \dots, J_1$



destes formam os componentes de um novo vetor  $x_1$  de dimensão  $J_1$ , que serão as entradas dos  $J_2$  neurônios da camada seguinte e assim sucessivamente até a última camada. Esta última camada, a camada de saída, consiste de  $J_K$  neurônios fornecendo como saída um vetor  $y = x_K$  de dimensão  $J_K$ . Desta forma,  $x_{k,j}$  denota a saída do  $j$ -ésimo neurônio na  $k$ -ésima camada, sendo que, para  $k = 0$ , representa a  $j$ -ésima entrada da rede.

O vetor de entrada  $x$  também é conhecido como a camada de entrada, que é constituída por neurônios de ligação direta, ou seja, que são simplesmente conectores de estímulos à rede. De qualquer forma, mantidas as referências explícitas quanto à dimensionalidade usada, é irrelevante a convenção utilizada quanto a quantidade de camadas na rede.

O problema que Rosenblatt propôs resolver é de que, em casos simples (como as funções booleanas E e OU de duas variáveis) é relativamente trivial escolher os ganhos sinápticos e o valor do limite de classificação. Porém, para que seja possível implementar uma função discriminatória arbitrária esta escolha não é tão simples. Dependendo do número de variáveis envolvidas, sem a existência de algum método, beira o impossível (KOVÁCS, 1996).

As redes neurais naturais (RNN) têm a propriedade de serem capazes de aprender uma função baseada nas respostas obtidas. Deveria existir, portanto, alguma maneira que tornasse uma rede neural artificial (RNA) capaz de aprender também, por meio de ensino ou treino. Treino é a palavra mais adequada para o conceito de redes neurais artificiais (HAYKIN, 2001). Uma RNA pode ser treinada de várias maneiras. Uma das mais intuitivas é com exemplos, chamado de *treinamento através de exemplos*. Neste tipo de treino, exemplos de comportamentos são apresentados à rede, de tal forma que, para um estímulo  $x_l^d$  a saída deva ser  $y_l^d$ , onde o índice  $d$  se refere ao desejado. Estes exemplos são repassados até que a RNA aprenda este comportamento, ou seja, até que encontre a função  $y_l^d = f(x_l^d)$  para todos os  $l = 1, 2, \dots, L$  exemplos.

O método de treinamento através de exemplos é definido por um conjunto  $\Psi$  com  $L$  exemplos de treinamento. Cada exemplo é um par do tipo  $(x_l^d, y_l^d)$ , onde os  $x_l^d$  são entradas que devem gerar saídas  $y_l^d$ . Estas saídas  $y_l^d$  serão +1 ou -1

conforme o vetor  $x_l^d$  pertença à classe  $\Theta_1$  ou  $\Theta_2$  que se pretende separar. Portanto o conjunto de treinamento é:

$$\Psi = \left\{ \left( x_l^d, y_l^d \right) \right\}_{l=1}^L. \quad (7)$$

O discriminador linear é inicializado com parâmetros aleatórios  $\{w_0, \Theta_0\}$  e os vetores  $x_l^d$  dos exemplos de treinamento são aplicados seqüencialmente à sua entrada. De posse de algum algoritmo para ajuste dos parâmetros  $\{w, \Theta\}$ , procura-se convergir para valores  $\{w^*, \Theta^*\}$  tais que as saídas do discriminador sejam idênticas no conjunto de treinamento para todos os exemplos  $\Psi$ . Ou seja:

$$y = \text{sgn}(w^{*t} x_{i,l}^d - \Theta^*) = y_l^d, \text{ para } l = 1, 2, \dots, L. \quad (8)$$

Quando o discriminador linear exhibe o comportamento expresso em (8), se diz que está treinado. Basta então determinar algum algoritmo de ajuste dos parâmetros que assegure esta convergência de uma maneira rápida e confiável.

### 2.3.3 Princípio de Aprendizado de Hebb

Em 1949, Hebb propôs um princípio pelo qual o aprendizado em sistemas nervosos complexos poderia ser reduzido a um processo puramente local, em que a intensidade das conexões sinápticas é alterada apenas em função dos erros detectáveis localmente (KOVÁCS, 1996). Este princípio local, com as devidas adaptações ao discriminador linear, pode ser traduzido no seguinte algoritmo: ao se apresentar o  $l$ -ésimo exemplo, os parâmetros  $w_i$  devem ser atualizados segundo a regra:

$$w_i^{novo} = w_i^{velho} + \Delta w_i, \text{ com } \Delta w_i = \eta (y_l^d - y_l) x_{i,l}^d \quad (9)$$

Por esta regra, a alteração do  $i$ -ésimo parâmetro depende unicamente do produto da  $i$ -ésima entrada pelo erro de saída  $e_l = (y_l^d - y)$ , sendo  $y$  a saída com os parâmetros “velhos” do discriminador, ou seja:

$$y_i = \text{sgn} \left( \sum_{j=1}^n w_j^{\text{velho}} x_{i,j}^d - \Theta^{\text{velho}} \right). \quad (10)$$

O parâmetro  $\eta$  em (9) é referido como taxa de aprendizado, na medida em que reflete a taxa com que os ganhos são alterados em consequência dos erros. Esta é uma regra local, pois não depende dos valores das demais variáveis espalhadas pelo sistema. Para aplicar esta regra de atualização ao parâmetro  $\Theta$ , note que este pode ser incorporado à soma ponderada como um ganho  $w_{n+1}$ , associado a uma entrada  $x_{n+1}$ , que está constantemente no valor  $x_{n+1} = -1$ , resultando na seguinte regra de atualização:

$$\Theta^{\text{novos}} = \Theta^{\text{velho}} - \eta(y_i^d - y_i). \quad (11)$$

Com a incorporação do limite na soma ponderada,  $w_{n+1} = \Theta$ , a notação do discriminador passa a ser:

$$y = \text{sgn} \left( \sum_{i=1}^{n+1} w_i x_i \right) = \text{sgn}(w'x). \quad (12)$$

### 2.3.3.1 Lei de Aprendizado do Perceptron

O princípio Hebbiano de treinamento expresso por (9) e (10) foi empregado por Rosenblatt para alterar os parâmetros  $w$  de um discriminador linear (HAYKIN, 2001). Uma vez que os  $\Delta w_i$  serão todos nulos, a expressão (10) deixa claro que, se a saída  $y_i$  for igual a saída  $y_i^d$ , os parâmetros não sofrem alteração alguma. Entretanto, se a saída desejada e a saída do discriminador forem diferentes, então:

$$\begin{aligned} y_i^d = 1, y_i = -1 &\rightarrow \Delta w_i = 2 \cdot \eta \cdot x_{i,j}^d \\ y_i^d = -1, y_i = 1 &\rightarrow \Delta w_i = 2 \cdot \eta \cdot x_{i,j}^d \Rightarrow \begin{cases} \Delta w_i = 0 \text{ para } y_i^d = y_i \\ \Delta w_i = 2 \cdot \eta \cdot y_i^d \cdot x_{i,j}^d \text{ para } y_i^d \neq y_i \end{cases}, \end{aligned} \quad (13)$$

o que pode ser resumido com a seguinte expressão:

$$\Delta w_i = \eta \cdot (1 - y_i^d \cdot y_i) \cdot y_i^d \cdot x_{i,l}^d. \quad (14)$$

Agora, deseja-se um discriminador linear cujo hiperplano, além de separar as duas coleções de pontos, também mantenha uma distância de pelo menos  $\tau$  de qualquer dos pontos  $x_i^d$ . Ou seja, deve-se impor uma zona de exclusão de largura  $\tau$  em torno do hiperplano. Entretanto, o parâmetro  $\tau$  deve ter um valor suficientemente pequeno para que isso seja possível. Como a distância  $d(w,x)$  no espaço euclidiano  $\mathfrak{R}^n$  é definida por:

$$d(w,x) = \frac{|w'x|}{|w|}, \text{ então:}$$

$$\frac{|w'x_i^d|}{|w|} \geq \tau, \text{ para todo } l = 1, 2, \dots, L. \quad (15)$$

Considerando que os pontos das duas categorias  $(\Phi_1, \Phi_2)$ , a serem separadas, devem estar nos respectivos lados do hiperplano, e, aplicando a relação (2), se tem que:

$$\begin{cases} w'x_i^d > 0 \text{ para } y_i^d = 1 \rightarrow x \in \Phi_1 \\ w'x_i^d < 0 \text{ para } y_i^d = -1 \rightarrow x \in \Phi_2 \end{cases}, \quad (16)$$

o que pode ser resumido na condição:

$$y_i^d \cdot w' \cdot x_i^d \geq |w| \cdot \tau = \delta, \quad (17)$$

Com este requisito adicional, o algoritmo da expressão (14) passa a ser:

$$\Delta w_i = \eta \cdot H(\delta - y_i^d \cdot w' \cdot x_i^d) \cdot y_i^d \cdot x_{i,l}^d, \quad (18)$$

onde  $H(u)$  é a função degrau unitário.

A expressão (18) é a Lei de Aprendizagem do Perceptron proposto por Rosenblatt. Se as duas categorias  $\Phi_1$  e  $\Phi_2$ , representadas no conjunto de treinamento forem linearmente separáveis então o algoritmo (18) sempre converge em um número finito de iterações (KOVÁCS, 1996).

O fato de ser linearmente separável garante a existência de um hiperplano  $w^*$  no sentido da expressão (8). Partindo do valor inicial  $w_0 = 0$ , todos os exemplos do conjunto de treinamento são utilizados de maneira cíclica, e os  $w_i$  são atualizados de acordo com (18) apenas quando:

$$\delta - y_i^d \cdot w^t \cdot x_i^d > 0 \rightarrow y_i^d \cdot w^t \cdot x_i^d < \delta. \quad (19)$$

Assim, se um determinado momento  $i = jL$  do processo iterativo, quando já utilizou-se  $j$  vezes todo o conjunto de treinamento de  $L$  exemplos, o  $l$ -ésimo exemplo atualiza o vetor  $w$   $m_l$ . Neste momento,  $w$  será igual a:

$$w = \eta \cdot \sum_{l=1}^L m_l y_l^d x_l^d. \quad (20)$$

A Desigualdade de Cauchy-Schwartz impõe que:

$$(w^{*t} w)^2 \leq |w^*|^2 \cdot |w|^2. \quad (21)$$

Sendo assim, a partir de (20), conclui-se que:

$$\begin{aligned} w^{*t} \cdot w &= \eta \cdot \sum_{l=1}^L m_l \cdot y_l^d \cdot w^{*t} \cdot x_l^d \geq \eta \cdot \left( \sum_{l=1}^L m_l \right) \min(y_l^d \cdot w^{*t} \cdot x_l^d) = \\ &= \eta \cdot M \cdot \tau \cdot |w^*|, \text{ onde } M = \sum_{l=1}^L m_l \end{aligned} \quad (22)$$

Por outro lado, o incremento no módulo de  $w$  cada vez que é feita uma atualização, devido a um particular exemplo  $x_q^d$  é tal que:

$$\begin{aligned} \Delta |w|^2 &= |w + \eta \cdot y_q^d \cdot x_q^d|^2 - |w|^2 = \eta^2 \cdot |x_q^d|^2 + 2 \cdot \eta \cdot y_q^d \cdot w^t \cdot x_q^d \\ &\leq \eta^2 \cdot |x_q^d|^2 + 2 \cdot \eta \cdot \tau \cdot |w^*|^2 = \eta \cdot \left( \eta \cdot |x_q^d|^2 + 2 \cdot \tau \cdot |w^*|^2 \right) \end{aligned} \quad (23)$$

De acordo com (18), uma variação em  $w$  só pode ocorrer quando:  $y_q^d w^t x_q^d \leq \delta = \tau |w^*|$ . Somando os incrementos no módulo de  $w$  sobre todas as atualizações, tem-se que:

$$|w|^2 \leq M \cdot \eta \cdot \left( \eta \cdot X + 2 \cdot \tau \cdot |w^*|^2 \right), \quad (24)$$

onde  $X = \max |x_i^d|^2$

Substituindo-se (22) e (24) na Desigualdade de Cauchy-Schwartz:

$$1 \geq \frac{(w^* w)^2}{|w^*|^2 |w|^2} \geq \frac{\eta \cdot M \cdot \tau^2}{\eta \cdot X + 2 \cdot \tau \cdot |w^*|^2} \Rightarrow M \leq \frac{\eta \cdot X + 2 \cdot \tau \cdot |w^*|^2}{\eta \cdot \tau^2}. \quad (25)$$

Como todos os parâmetros envolvidos do lado direito da desigualdade são finitos e não nulos, obtém-se que  $M$  é necessariamente limitado por um valor finito. Isto é, o número total de atualizações não poderá crescer indefinidamente.

### 2.3.4 ADALINE e MADALINE

O método ADALINE (do inglês, *adaptive linear element*), que posteriormente foi generalizado para o multidimensional MADALINE (do inglês, *multiple adaptive linear element*) foi desenvolvido por Widrow na década de 1960. O adaline é um modelo muito simples, o que acabou por restringir seu uso ao contexto acadêmico. Mas houve uma importante contribuição, que foi a invenção de um princípio de treinamento muito poderoso, conhecido como Regra Delta. Esta regra foi, mais tarde, generalizada para redes com modelos neurais mais elaborados (HAYKIN, 2001).

Como em toda aproximação, é necessário se obter um mínimo, ou máximo, global. Entretanto, em dimensões maiores isto se torna um problema. A proposta de Widrow foi de obter o ponto mínimo através de um processo de iteração local, utilizando um exemplo do conjunto de treinamento de cada vez. O método resultante é conhecido como método do gradiente, ou regra delta.

O erro quadrático para um discriminador aleatório de parâmetros  $w$  é dado por:

$$E(w) = \sum_{l=1}^L \left( \sum_{i=1}^n w_i \cdot x_{i,l}^d - y_l^d \right)^2. \quad (26)$$

Pelo método do gradiente, partindo-se de um ponto arbitrário  $w(0)$ , pode-se caminhar pela superfície  $E(w)$  em direção ao ponto de mínimo, bastando para isso, evoluir sempre no sentido oposto ao do gradiente naquele ponto. Ou seja, dado  $w(k)$ , obtém-se  $w(k+1)$  por:

$$\begin{aligned} w(k+1) &= w(k) - \eta \cdot \text{grad}(E(w(k))), \text{ ou} \\ w(k+1) &= w(k) - \eta \cdot \frac{d}{dw} E(w(k)) \end{aligned} \quad (27)$$

onde o parâmetro  $\eta$  determina o tamanho do passo que é dado naquela direção.

Este algoritmo é a base para um procedimento de treinamento ainda mais poderoso no uso de RNAs, que é o algoritmo da *retropropagação*.

### 2.3.5 Retropropagação

De acordo com Kovács (1996) este é o mais poderoso procedimento conhecido para treinamento de redes neurais. Ele parte de uma regra delta mais ampla, chamada de Regra Delta Generalizada de Rumelhart, Hinton e Williams. É basicamente um adaline de Widrow seguido de uma função de ativação  $g(v)$ . Ou seja:

$$y = g\left(\sum_{i=1}^n w_i \cdot x_i\right). \quad (28)$$

De forma geral, após algumas manipulações matemáticas, chega-se a formulação da regra delta generalizada, dada por:

$$w(k+1) = w(k) + 2 \cdot \eta \cdot \sum_{l=1}^L \delta_l^d \cdot x_l^d. \quad (29)$$

Regra esta, que tem a mesma forma da regra delta de Widrow, demonstrada pela expressão (30):

$$w(k+1) = w(k) + \eta \cdot \sum_{l=1}^L (x_l^d)(y_l^d - y_l) = w(k+1) = w(k) + \eta \cdot \sum_{l=1}^L \delta_l \cdot x_l^d. \quad (30)$$

De acordo com Kovács (1996) e Selau (2000), o algoritmo de retropropagação é definido a seguir.

### 2.3.5.1 Algoritmo de Retropropagação

Dado um conjunto de treinamento de  $L$  exemplos:

$$\Psi = \left\{ \left( x_l^d, y_l^d \right) \right\}_{l=1}^L, \quad (31)$$

uma rede de  $M$  camadas com  $\{J_0, J_1, \dots, J_M\}$  elementos em cada camada é treinada sobre  $\Psi$  pelo seguinte algoritmo:

- Escolhe-se um conjunto de parâmetros iniciais  $W(0)$ , e uma taxa de aprendizado  $\eta$ ;
- Estabelece-se algum critério de parada, seja este um número máximo de interações  $K_{\max}$ , ou um fator de correção  $\lambda$ , ou ainda uma combinação dos dois;
- Calculam-se os erros de saída, da camada de saída e das camadas subseqüentes com  $l = 1, 2, \dots, L$ ,  $\delta_{y(jM),l} = (y_{jM,l}^d - y_{jM,l})$  e:

$$\delta_{jM,l} = \delta_{y(jM),l} \cdot \frac{d}{dv_{jM}} g(v_{jM}), \quad (32)$$

variando  $m = (M-1), (M-2), \dots, 1, 0$ :

$$\delta_{jm,l} = \left( \sum_{j(m+1)=1}^{J(m+1)} \delta_{j(m+1),l} \cdot w_{m+1,j(m+1),jm} \right) \frac{d}{dh_{jm}} g(h_{jm}); \quad (33)$$

- Atualizam-se os parâmetros mediante:

$$\Delta w_{m,jm,j(m-1)} = 2 \cdot \eta \cdot \sum_{l=1}^L \delta_{jm,l} u_{j(m-1),l}^d, \quad (34)$$

notando que  $u_{j(0),l}^d = x_{j(0),l}^d$ ;



- e) Recalcula-se  $W(1) = W(0) + \Delta W$  e retoma-se o ciclo a partir do passo (c), até que o critério de parada seja satisfeito.

### 2.3.6 Redes Neurais Artificiais na análise de séries temporais

Segundo Haykin (2001),

“O tempo [grifo do autor] constitui um ingrediente essencial do processo de aprendizagem.(...) É através da incorporação do tempo na operação de uma rede neural que ela é capacitada a seguir as variações estatísticas em processos não-estacionários.”

Existem duas formas distintas de incorporar o tempo na operação de uma rede neural: representação implícita ou representação explícita. No caso da representação explícita, o tempo recebe sua particular representação. Este tipo de representação é muito utilizado em sistemas biológicos, como por exemplo, o sistema de ecolocalização de um morcego. Porém este caso não será abordado neste trabalho.

Entretanto, na representação implícita, o tempo é representado pelo efeito que tem sobre o processamento de sinais de uma maneira implícita, por isso o nome. Neste tipo de representação o tempo não é incluído diretamente no modelo, mas sim de uma forma indireta. Desta forma, a rede neural estática (como definida anteriormente) é suprida com propriedades dinâmicas, tornando-a sensível quanto a estrutura temporal dos sinais portadores de informação (HAYKIN, 2001).

Mas para que isso ocorra, é necessário que a rede neural possua memória. Esta memória pode ser de curto prazo ou de longo prazo. Segundo Haykin (2001), a memória de longo prazo é inserida numa rede neural através de algum método de aprendizagem supervisionada. Assim, o conteúdo de informação do conjunto de dados de treinamento é armazenado, total ou parcialmente, nos pesos sinápticos da rede. Entretanto, se a tarefa considerada possuir uma dimensão temporal, é necessário que alguma forma de memória de curto prazo seja inserida, tornando o modelo dinâmico. Uma forma simples de inserir memória de curto prazo na estrutura de uma rede neural é através de atrasos no tempo (*lags*), que podem ser

implementados a nível sináptico, ou seja, dentro de uma das camadas ocultas da rede, ou na camada de entrada da rede.

A utilização de uma rede neural estática (como as vistas anteriormente) é basicamente de reconhecimento estrutural de padrões ou de classificação. O reconhecimento temporal de padrões requer o processamento de padrões que evoluem no tempo, com a resposta em um dado instante de tempo depender não só do valor presente da entrada, mas também de seus valores passados. Se estes valores passados forem informados na camada de entrada então esta rede é conhecida como TLFN focada, (do original em inglês *focused time lagged feedforward network*). De acordo com Haykin (2001), a TLFN focada é um filtro não-linear mais poderoso, consistindo de uma memória de linha de atraso derivada de ordem  $p$  e um perceptron de múltiplas camadas, conforme a equação (35).

$$y(n) = \sum_{j=1}^{m_1} w_j y_j(n) = \sum_{j=1}^{m_1} w_j \varphi \left( \sum_{l=0}^p w_j(l) x(n-l) \right) \quad (35)$$

onde se assume que o neurônio de saída na TLFN focada é linear; os pesos sinápticos do neurônio de saída são representados pelo conjunto  $\{w_j\}_{j=1}^{m_1}$ , onde  $m_1$  é o tamanho da camada oculta e  $\varphi(\cdot)$  é a função de ativação do neurônio  $j$ .

Desta forma uma rede neural pode ser utilizada para captar as variações temporais existentes num sistema dinâmico, ou seja, num sistema que se altera ao longo do tempo. A função  $\varphi(\cdot)$  utilizada na modelagem de séries temporais por redes neurais costuma ser a função logística, conforme equação (36).

$$\varphi(x) = \frac{1}{1 + e^{-x}} \quad (36)$$

Na função *nnet* utilizada no pacote R não fica clara qual função é utilizada como função de ativação nos neurônios da camada oculta. Acredita-se que, pelo padrão apresentado pela literatura e artigos revisados, seja uma função do mesmo modo que a equação (36). Mas isso não pode ser afirmado com certeza.

### *2.3.7 Redes Neurais Artificiais e a Estatística*

As RNAs também são conhecidas como aproximadores de funções. Isto se deve ao fato de, mesmo em sistemas mais complexos (como sistemas não-lineares, por exemplo), seus resultados serem melhores, em termos de erros, do que outros métodos. Muitos artigos comparam as estimativas feitas por uma RNA contra outros métodos conhecidos, indo desde a modelagem por Box-Jenkins (PINTO; MEDEIROS, 2005) até comparações com modelos tipicamente não lineares, como os ARCH/GARCH (SHLENS, 1999).

Entretanto, apesar deste trabalho estar focado nas qualidades preditivas das RNAs, seu uso têm se estendido na Estatística, principalmente no campo da Análise Multivariada. Neste ambiente é utilizada como função classificatória, de forma a separar a população em grupos distintos (como visto no neurônio de McCulloch). Definições mais específicas sobre a modelagem utilizada, algoritmo, arquitetura, entre outras informações serão tratadas no próximo capítulo.

Para uma maior compreensão sobre os tipos de redes, arquiteturas, métodos e algoritmos observar Haykin (2001), Kovács (1996), Hair (2005), Fernandes (1995), Selau (2000), Zhang (2000) e Shlens (1999).

### 3 ESTUDO DE CASO

#### 3.1 Apresentação da empresa

A empresa deste estudo de caso surgiu em meados de 1847 com a abertura de uma oficina mecânica pelo Sr. Daniel Massey em Ontário, Canadá. Em 1874, se transformou na Massey Manufacturing Company quando os direitos de patente do Self-Dumping Wheel-Rake da Sharp foram adquiridos e de onde começou a fabricar e vender mais de 250.000 unidades de maquinário agrícola. Em 1891 acontece a fusão da Massey Company e A Harris Son & Company, as duas empresas de maior sucesso no setor agrícola do Canadá, para formarem a Massey-Harris. Em 1918 esta companhia introduziu o trator Massey-Harris Número 1. Já em 1920 a Harry Ferguson aperfeiçoa o sistema Ferguson de integrar o trator e implementos e por volta de 1930 acontece a introdução do primeiro trator com tração nas quatro rodas. No ano de 1947 após a produção de 300.000 tratores termina a parceria Ford-Ferguson e é formada a Massey-Harris Ferguson Company que será abreviado posteriormente para Massey-Ferguson com a fabricação contínua de tratores e implementos agrícolas. Em 1975 aconteceu o lançamento do MF 275, o trator mais vendido da história da mecanização brasileira. A empresa vem a ser adquirida, com todos os direitos de distribuição, em 1994, pela AGCO Company.

A AGCO, desde 1960, vem desenvolvendo sistemas voltados para a qualidade de seus processos de manufatura e de seus produtos. Em 1961 a empresa atinge a liderança de mercado no segmento produção de tratores, atingindo em 1962 um recorde de vendas no mundo ocidental. No ano de 1993 a AGCO adquire os direitos de distribuição dos produtos Massey Ferguson na América do Norte, e em 1994 adquire a Massey Ferguson Group Limited, aderindo no mesmo ano ao programa Gaúcho da Qualidade e Produtividade e sendo a primeira empresa de máquinas agrícolas brasileiras a obter a Certificação ISO 9001.

Um outro marco para a empresa foi o lançamento da série 5000 dos tratores Massey Ferguson em 1998, seguido em 1999 pelo lançamento das colheitadeiras MF 34 e MF 38.

Em 2001 a empresa comemorou seus 40 anos consecutivos de liderança no Brasil e se consolidou como líder brasileira em exportação de máquinas agrícolas. No ano de 2002, visando dar continuidade aos processos e a atender sempre melhor, a empresa lançou o programa de gestão de Pós-Venda.

A AGCO é uma companhia norte-americana que fabrica e distribui uma grande variedade de máquinas e implementos agrícolas e é responsável por 25% da produção mundial de máquinas agrícolas. Entre alguns de seus produtos estão tratores de 51 a 306 CV, colheitadeiras, retro-escavadeiras, pás carregadeiras, tratores de esteira e pulverizadores. As marcas que a AGCO representa atualmente estão na Figura 4.



Figura 4 - Marcas que a AGCO representa  
Fonte: AGCO do Brasil Indústria e Comércio Ltda.

A empresa está presente em mais de 140 países, sendo que, no Brasil, possui fábricas nos municípios de Canoas e Santa Rosa, no Rio Grande do Sul. A fábrica de Santa Rosa é exclusiva de colheitadeiras. A empresa também se orgulha de

possuir a maior rede de concessionárias do país no seu ramo, com 220 lojas espalhadas pelo país. Além disso, até hoje mantêm o primeiro programa de marketing de relacionamento do setor agrícola, que é usado para desenvolver soluções cada vez mais adequadas junto aos seus clientes.

A AGCO do Brasil também é responsável pelos negócios da Corporação em toda a América do Sul, América Central e o Caribe. Seus produtos são exportados, atualmente, para a América do Norte, Oriente Médio, Europa, Oceania e países do Pacífico. Para tanto a empresa conta com 3 centros de distribuição na América Latina: um em Santos-SP, outro em Canoas-RS e um em Haedo, em Buenos Aires na Argentina. Em termos quantitativos, a AGCO do Brasil conta com aproximadamente 2000 colaboradores, entre fornecedores e representantes no Brasil e cerca de 3000 funcionários nas fábricas de Canoas e Santa Rosa. A AGCO do Brasil também é a maior fabricante de tratores do país e a maior exportadora.

Recentemente a AGCO investiu 52 milhões de dólares nas suas fábricas brasileiras para novos equipamentos, melhoras em logística e reorganização de alguns de seus processos de produção. Com isso, a empresa hoje possui capacidade para produzir 140 tratores por dia e 1600 colheitadeiras por ano, com um parque industrial de 80 mil metros quadrados somando as duas unidades no Rio Grande do Sul.

### 3.2 Modelo da produção da empresa

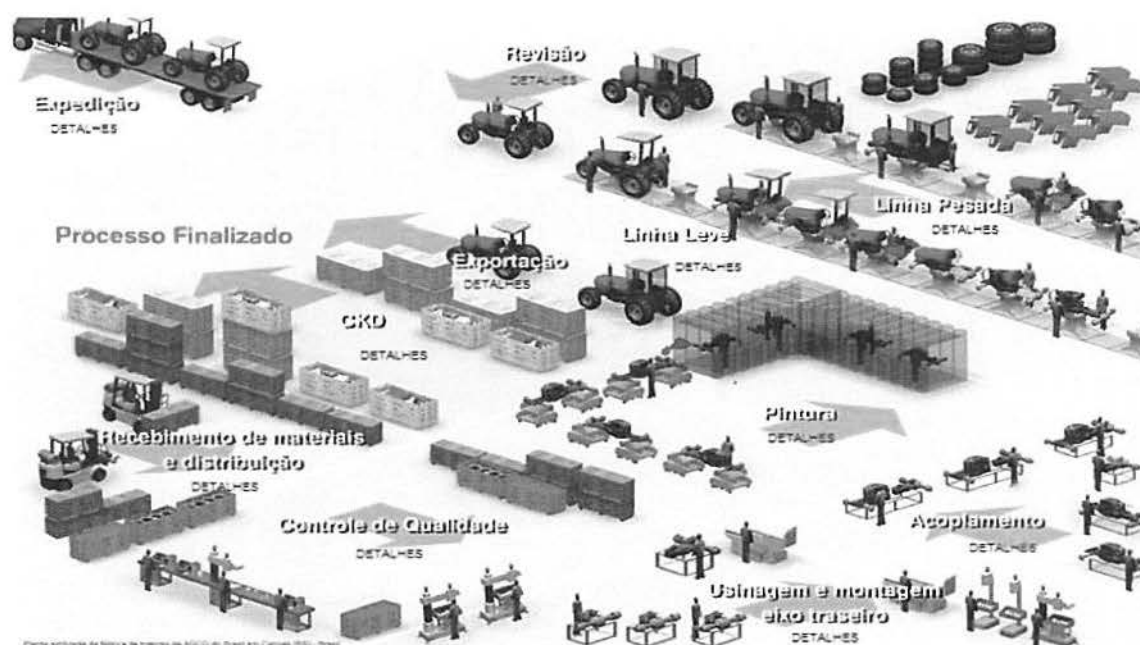
UFRGS  
SISTEMA DE BIBLIOTECAS  
BIBLIOTECA SETORIAL DE MATEMÁTICA

“Prover a capacitação de satisfazer a demanda atual e futura é uma responsabilidade fundamental da administração de produção. Um equilíbrio adequado entre capacidade e demanda pode gerar altos lucros e clientes satisfeitos, enquanto o equilíbrio errado pode ser potencialmente desastroso”. (SLACK, 2002)

A demanda de tratores, devido à abertura das exportações, aumentou. Por isto, a AGCO trabalha hoje com uma capacidade máxima de 2.500 tratores/mês, e aproximadamente, 20.000 tratores por ano, operando a fábrica com sua capacidade máxima desde 2004, após o aumento de sua planta de produção em Canoas-RS.

A fábrica trabalha com produção contínua e, em casos especiais, por projeto. Quando há a entrada de um pedido por projeto, existe uma análise com os engenheiros para a possibilidade de produção e apresentação de custos. Em caso de aprovação, o pedido é entregue em até 90 dias.

A empresa trabalha por células em cada um dos processos envolvidos, e a produção é puxada a partir do momento de montagem das peças. Cada célula atende por uma determinada função, ou parte do produto, permitindo uma maior variedade no produto sem que seja necessário removê-lo da linha de produção. A divisão por celular também permite um trabalho independente que não é possível numa linha de produção única. Em relação a gargalos de produção, um dos maiores é a falta de peças, que muitas vezes não são entregues no prazo ou apresentam falhas de qualidade. A Figura 5 ilustra todo o processo de produção



**Figura 5 - Processo de produção da AGCO de forma simplificada**  
 Fonte: AGCO do Brasil Indústria e Comércio Ltda.

### 3.3 Método de previsão de vendas atualmente utilizado

A previsão de demanda é o que a AGCO espera que seja vendido no mês seguinte e no próximo. A empresa, até este momento, se vale única e

exclusivamente da experiência de seus gerentes de produção, vendas e marketing. De acordo com o especialista entrevistado, o método para previsão consiste em observar e analisar as seguintes informações:

- a) O total de vendas nos últimos dois anos;
- b) As vendas para o mês de previsão nos dois últimos anos;
- c) Gastos com propagandas e possíveis feiras que venham a ocorrer;
- d) Pedidos de consórcios e exportações<sup>1</sup>;
- e) Pedidos de tratores especiais;
- f) Alguma promoção que venha a ser feita.

De posse destas informações, os departamentos de Marketing, Vendas e Produção se reúnem e decidem quantos e quais são os produtos e serem produzidos.

### 3.4 Apresentação do produto escolhido

O modelo de trator escolhido para as previsões de vendas é o MF 275. O MF do nome se deve ao fato de ser um trator da marca Massey Ferguson. Este é um trator com praticamente 100% de peças nacionais. Com exceção de algumas peças de rolamento e outras pequenas partes importadas.

A escolha deste modelo se deve tanto às suas vendas (um dos líderes de vendas) quanto a sua versatilidade. O MF 275 pode ser utilizado em praticamente qualquer tipo de lavoura, devido a sua potência de 75 cv (cavalos de força) e seu tamanho reduzido. A sua versatilidade também se deve a grande quantidade de opcionais. As especificações técnicas do modelo são apresentadas na Tabela 1.

---

<sup>1</sup> Não foram utilizadas estas informações nos modelos, pois o objetivo é modelar as vendas por atacado.



Tabela 1 - Especificações Técnicas do MF 275

PERFORMANCE		TRANSMISSÃO		CONTROLE REMOTO																																								
Potência do motor na rotação nominal - kw (cv)	55 (75)*	Tipo	Deslizante 8x2	Tipo	Independente																																							
Potência Máxima na TDP - kw (cv)	49 (64)*		Engrenamento constante 12x4	Número de válvulas																																								
Torque máximo do motor @ 1400 rpm - Nm (mkgf)	289 (29,5)*		Sincronizada 12x4	- Básico	1																																							
Rotação nominal do motor - rpm	2200	Redutor de velocidades (Creeper)	Opcional	- Opcional	2																																							
(*) ISO 14396			4x2 ou 4x4 lateral	Vazão máxima - L/min	42																																							
MOTOR		TOMADA DE POTÊNCIA		RODAGEM																																								
Marca	Perkins	Tipo	Dependente	<table border="1"> <thead> <tr> <th>TRAÇÃO</th> <th>TRASEIRO</th> <th>DIANTEIRO</th> </tr> </thead> <tbody> <tr> <td>4x2(Básico)</td> <td>18 4-30R1</td> <td>7 50-16F2</td> </tr> <tr> <td>4x2</td> <td>14 9-26R1</td> <td>7 50-16F2</td> </tr> <tr> <td>4x2</td> <td>14 9-26R2</td> <td>7 50-16F2</td> </tr> <tr> <td>4x2</td> <td>16 9-30R1</td> <td>7 50-16F2</td> </tr> <tr> <td>4x2</td> <td>16 4-30R2</td> <td>7 50-16F2</td> </tr> <tr> <td>4x2</td> <td>13 6-38R1</td> <td>7 50-16F2</td> </tr> <tr> <td>4x2</td> <td>23 1-26R2</td> <td>9 00-16F2</td> </tr> <tr> <td>4x4(Básico)</td> <td>18 4-30R1</td> <td>12 40-24R1</td> </tr> <tr> <td>4x4</td> <td>14 9-26R1</td> <td>6 30-24R1</td> </tr> <tr> <td>4x4</td> <td>23 1-26R2</td> <td>14 9-24R1</td> </tr> <tr> <td>4x4</td> <td>13 6-38R1</td> <td>12 4-24R1</td> </tr> <tr> <td>4x4</td> <td>18 4-30R2</td> <td>12 4-24R1</td> </tr> </tbody> </table>		TRAÇÃO	TRASEIRO	DIANTEIRO	4x2(Básico)	18 4-30R1	7 50-16F2	4x2	14 9-26R1	7 50-16F2	4x2	14 9-26R2	7 50-16F2	4x2	16 9-30R1	7 50-16F2	4x2	16 4-30R2	7 50-16F2	4x2	13 6-38R1	7 50-16F2	4x2	23 1-26R2	9 00-16F2	4x4(Básico)	18 4-30R1	12 40-24R1	4x4	14 9-26R1	6 30-24R1	4x4	23 1-26R2	14 9-24R1	4x4	13 6-38R1	12 4-24R1	4x4	18 4-30R2	12 4-24R1
TRAÇÃO	TRASEIRO	DIANTEIRO																																										
4x2(Básico)	18 4-30R1	7 50-16F2																																										
4x2	14 9-26R1	7 50-16F2																																										
4x2	14 9-26R2	7 50-16F2																																										
4x2	16 9-30R1	7 50-16F2																																										
4x2	16 4-30R2	7 50-16F2																																										
4x2	13 6-38R1	7 50-16F2																																										
4x2	23 1-26R2	9 00-16F2																																										
4x4(Básico)	18 4-30R1	12 40-24R1																																										
4x4	14 9-26R1	6 30-24R1																																										
4x4	23 1-26R2	14 9-24R1																																										
4x4	13 6-38R1	12 4-24R1																																										
4x4	18 4-30R2	12 4-24R1																																										
Modelo	P4000	- Básico	Independente	Tanque de combustível - L - Plataformado 75 - Cabinado 100 Carter do motor - L 6 Transmissão / hidráulico - L - 4x2 38 - 4x4 39																																								
Nº de cilindros	4	- Opcional	540 ou 540/1000	<b>EQUIPAMENTOS BÁSICOS</b> Barra de tração escalante HD, caixa de ferramentas com chave de rodas, suporte dos pesos dianteiros, assento regulável e suspensão, pré-filtro, controle remoto com retorno por mola																																								
Cilindrada - cm	4100	Rotação nominal da TDP - rpm	540 ou 540/1000	<b>EQUIPAMENTOS VARIÁVEIS</b> Aquecedor de partida, pesos dianteiros, pesos traseiros, bloco óptico, defletor de calor e pó, tampa de óleo para carreta, proteção da TDP, catálogo de peças, tomada de luz para carreta, kit de ferramentas, bateria de irrigar, buzina, EPCC, soldo, creeper, sistema PWT de rodagem traseira (18 4-30R1 e 18 4-30R2) cabine																																								
Aspiração	Natural	Rotação do motor, à rotação nominal da TDP - rpm	1900																																									
AMBIENTE DO OPERADOR		SISTEMA DE DIREÇÃO																																										
Semi-plataforma	Básico	Tipo	Hidroestática																																									
Cabina	Opcional	EIXO DIANTEIRO																																										
Obs: Não disponível para tratores estreitos.		Tipo	4x2																																									
SISTEMA ELÉTRICO		- Acionamento	Central ou Lateral																																									
Alternador - A		- Bloqueio do diferencial	Automático																																									
- Plataformado	55	EIXO TRASEIRO																																										
- Cabinado	75	Redução final	Epicíclica																																									
Tensão - V	12	Bloqueio do diferencial	Acionamento por pedal																																									
Bateria - A/h	95	FREIOS																																										
EMBREGEM		Serviço																																										
Tipo	Dupla ou Split torque	- Nº de discos	4 de cada lado																																									
Acionamento	Contato permanente	- Tipo	Discos em banho de óleo																																									
Material	Orgânico	- Acionamento	Mecânico																																									
Dímetro disco - mm		Estacionamento																																										
- Transmissão	305	- Tipo	Trava manual nos pedais																																									
- TDP	254	SISTEMAS HIDRÁULICOS																																										
TABELA DE VELOCIDADES		Controle	Mecânico																																									
Rotação do motor - rpm	2200	Capacidade de levante na rotula - kgf																																										
Rodado traseiro	18 4-30R1	- Básico	2100																																									
Frente	12x4	- Variável	2500																																									
1	2,54	Vazão máxima - L/min	17 ou 27																																									
2	3,26	Pressão máxima - bar (kgf/cm <sup>2</sup> )	210(214)																																									
3	3,95	Funções de controle	Posição, profundidade, reação, transporte e bombeamento constante																																									
4	4,88																																											
5	7,25																																											
6	8,95																																											
7	10,55																																											
8	13,02																																											
9	15,82																																											
10	19,53																																											
11	29,00																																											
12	35,81																																											
Ré	Ré																																											
1	3,67																																											
2	4,53																																											
3	14,69																																											
4	18,14																																											
PESOS E DIMENSÕES		4x2		4x4																																								
Peso de embarque - kg		2580		2990																																								
Peso máximo com lastro - kg		3564		4054																																								
Comprimento total - mm		3920		4170																																								
Altura máxima no toldo - mm		2600		2600																																								
Distância entre eixos - mm		2135		2370																																								
Vão livre médio - mm		380		290																																								
Bitolas (mínima/máxima) - mm																																												
- Dianteira		1390/1390		1645/2040																																								
- Traseira		1560/2165		1560/2185																																								
Raio de giro médio - mm																																												
- Com freio		3550		3750																																								
- Sem freio		4000		4220																																								

Fonte: AGCO do Brasil Indústria e Comércio Ltda.

### 3.5 Apresentação da série de vendas

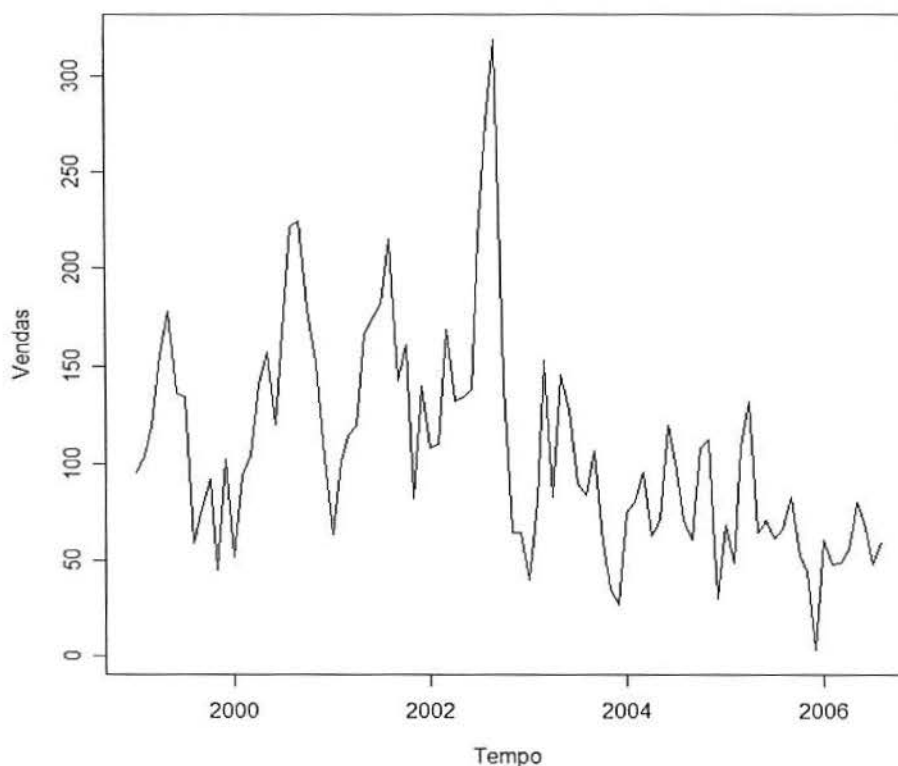
UFRGS  
SISTEMA DE BIBLIOTECAS  
BIBLIOTECA SETORIAL DE MATEMÁTICA

A Tabela 2 e a Figura 6, apresentam a série de vendas do trator MF 275, modelo 4x2, segundo dados da Associação Nacional de Fabricantes de Veículos Automotores (ANFAVEA). Estes dados foram obtidos no site da associação ([www.anfavea.com.br](http://www.anfavea.com.br)).

**Tabela 2 – Unidades vendidas de tratores MF 275 por atacado**

Ano	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
1999	95	104	119	156	178	136	134	59	77	92	45	102
2000	52	94	104	141	157	120	168	221	224	179	152	103
2001	64	101	115	119	167	174	182	215	143	161	82	140
2002	108	110	169	132	134	138	215	276	319	140	65	65
2003	40	81	153	83	146	128	90	84	107	60	35	27
2004	75	80	96	63	71	120	101	71	61	108	113	30
2005	69	49	109	132	65	71	62	67	83	53	44	3
2006	61	48	49	56	80	68	48	60	-	-	-	-

Fonte: ANFAVEA (2006)



**Figura 6 - Gráfico da série de vendas dos tratores MF 275 por atacado**  
**Fonte: elaboração própria com dados obtidos junto a ANFAVEA.**

Como se percebe, a série apresenta um comportamento bem irregular. As vendas nos meses de setembro de 2002 e dezembro de 2005 são bem atípicas. A análise é feita no tópico a seguir.

### 3.6 Análise prévia da série

Como se trata de uma série de vendas, diversas variáveis poderiam ser usadas para explicar o comportamento. Dentre algumas delas se pode citar: preço de venda,

demanda de mercado, custeio de produção, clima, entre outras. Entretanto algumas variáveis são demasiadamente complicadas para se obter. Variáveis como o clima, por exemplo. Utiliza-se a média de chuvas do mês, ou a temperatura média do mês, ou uma variável *dummy* indicando se houve ou não tempestades, entre outras que podem ser citadas. Desta forma, estas variáveis não teriam como ser utilizadas na modelagem. Como nem todas as variáveis poderiam ser testadas quanto a sua contribuição para o modelo, então nenhuma delas é utilizada

Entretanto, observando-se a Figura 6, a série aparenta possuir dois níveis distintos: um deles até o mês de setembro de 2002 e o outro a partir de outubro do mesmo ano. Sendo assim, deseja-se verificar a linearidade da série. Para isso foi utilizado o Teste de Rede Neural para Não Linearidade de Teraesvirta (TERAESVIRTA; LIN; GRANGER, 1993). Para este teste foi utilizado a configuração padrão, com lag igual a 1 e a estatística de teste Qui-quadrado. Segundo o teste, a série deve ser não linear em relação a média (p-valor = 0,02473), conforme Figura 7.

```

Teraesvirta Neural Network Test

data: ts.dados
X-squared = 7.3997, df = 2, p-value = 0.02473

```

**Figura 7 - Saída do teste de Teraesvirta no R**  
**Fonte: elaboração própria com dados da ANFAVEA.**

O programa utilizado para as análises e para a elaboração dos modelos é o R, um programa estatístico distribuído gratuitamente, e os seguintes pacotes do mesmo programa: *nnet* e *tseries*.

### 3.7 Aplicação da técnica de Redes Neurais Artificiais

O próximo passo é preparar a série para a análise. A metodologia utilizada é a mesma de utilizada por vários autores, como, por exemplo, Fernandes (1995) e Mantovani (2004) entre outros. Para a utilização de redes neurais é necessário que seja realizado um *treinamento* de cada modelo proposto de RNA.

Alguns dos últimos valores da série foram separados para o teste de predição do melhor modelo (de acordo com os critérios utilizados). A série completa apresenta 92 valores. Os últimos 6 foram retirados para averiguação das predições, restando então 86 valores para treinamento. A série, com estes 86 valores, teve sorteadas 30 amostras de 60 valores para o treinamento de cada modelo. Como os procedimentos de amostragem e de atribuição de pesos iniciais para a rede neural são procedimentos aleatórios, foi definida uma semente inicial para o programa, para posterior reprodução do experimento.

Os modelos utilizados possuem três características básicas para diferenciá-los: o número de variáveis usadas na camada de entrada, o tipo de variáveis usadas e o número de neurônios na camada oculta. Para as variáveis de entrada foram realizadas seis diferentes configurações:

- a) Apenas o mês imediatamente anterior ( $X_{t-1}$ );
- b) O mês imediatamente anterior e variáveis indicadoras do mês em questão ( $X_{t-1}$  e mais ind.);
- c) O mesmo mês, só que do ano anterior ( $X_{t-12}$ );
- d) O mês imediatamente anterior, o mesmo mês no ano anterior e ainda o mês imediatamente anterior no ano anterior ( $X_{t-1}$ ,  $X_{t-12}$  e  $X_{t-13}$ );
- e) O mês imediatamente anterior e o mesmo mês no ano anterior ( $X_{t-1}$  e  $X_{t-12}$ );
- f) E finalmente um modelo com o mês imediatamente anterior, o mesmo mês no ano anterior e mais as variáveis indicadoras do mês anterior ( $X_{t-1}$ ,  $X_{t-12}$  e ind.).

Estes modelos foram obtidos junto a análises empíricas realizadas na série. Entretanto, como a análise realizada com redes neurais não possui nenhum algoritmo ou metodologia definida, muitos outros modelos são possíveis. As análises empíricas foram baseadas em ajustes por modelos ARIMA e SARIMA. Mesmo a série sendo não linear estes modelos mostraram possíveis caminhos a seguir quanto as variáveis da camada de entrada.

Para definir o número ideal de neurônios na camada oculta também se enfrenta o mesmo problema de falta de métodos específicos ou claros para sua seleção. De acordo com Min Qi e Zhang (2000), costuma-se utilizar critérios como AIC (*Akaike information criteria*, do original em inglês) e BIC (*Bayesian information criteria*, do

original em inglês) para este fim. Segundo estes autores, estes critérios (assim como outros) podem ser usados para selecionar o melhor modelo. Entretanto, os autores sugerem mais de uma forma para calcular o AIC e o BIC. Para a seleção de modelos neste trabalho, foram utilizados os seguintes critérios:

$$SSE = \sum_{i=1}^T (y_i - \hat{y}_i)^2, \quad (37)$$

$$AIC1 = \log\left(\frac{SSE}{T}\right) + \frac{2\log(m)}{T}, \quad (38)$$

$$BIC3 = \log\left(\frac{SSE}{T}\right) + \frac{m \log(T)}{T}, \quad (39)$$

onde  $SSE$  é a soma quadrática do erro,  $T$  é o tamanho da série utilizada,  $m$  é o número de parâmetros do modelo,  $y_i$  é o valor original da série e  $\hat{y}_i$  o valor estimado pelo modelo.

O número de parâmetros é calculado da seguinte forma:

$$m = k(p + 2) + 1, \quad (40)$$

onde  $k$  é o número de neurônios na camada oculta e  $p$  o número de neurônios na camada de entrada, ou seja, o número de variáveis usadas no modelo.

Os modelos citados acima foram feitos com quatro diferentes configurações em relação a quantidade de neurônios na camada oculta, totalizando 24 modelos feitos. A Tabela 3 e a Tabela 4 apresentam os valores obtidos pelos critérios AIC1 e BIC3, respectivamente, para seleção de modelos.

**Tabela 3 - Critérios AIC dos modelos**

Neurônios na camada oculta	$X(t-1)$	$X(t-1)$ e $Ind$	$X(t-12)$	$X(t-1)$ , $X(t-12)$ e $X(t-13)$	$X(t-1)$ e $X(t-12)$	$X(t-1)$ , $X(t-12)$ e $ind$ .
3	7,259	7,188	7,834	7,203	7,423	7,048
4	7,358	7,181	7,849	6,947	7,275	5,911
5	7,223	6,792	7,867	6,884	7,167	5,614
6	7,178	5,549	7,785	7,182	6,837	3,504

Fonte: elaboração própria

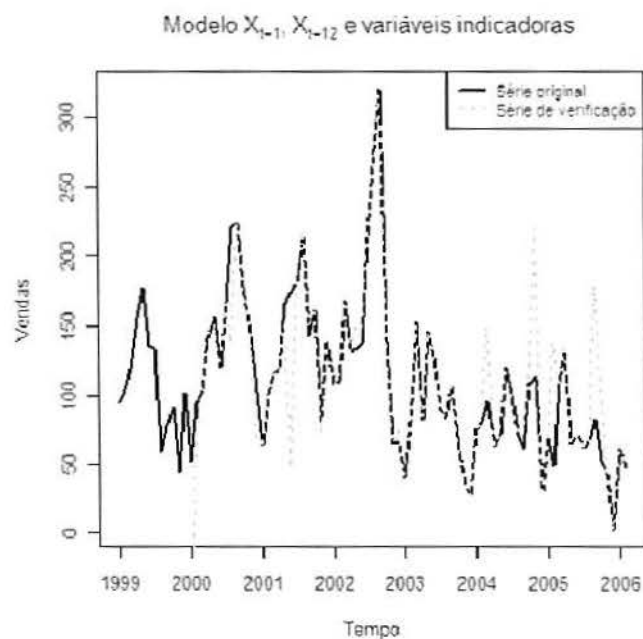
Tabela 4 - Critérios BIC dos modelos

Neurônios na camada oculta	$X(t-1)$	$X(t-1)$ e $Ind$	$X(t-12)$	$X(t-1), X(t-12)$ e $X(t-13)$	$X(t-1)$ e $X(t-12)$	$X(t-1), X(t-12)$ e $ind.$
3	7,830	10,024	8,405	8,146	8,179	10,075
4	8,114	10,972	8,605	8,202	8,280	9,957
5	8,165	11,540	8,809	8,454	8,422	10,680
6	8,308	11,254	8,915	9,068	8,343	9,593

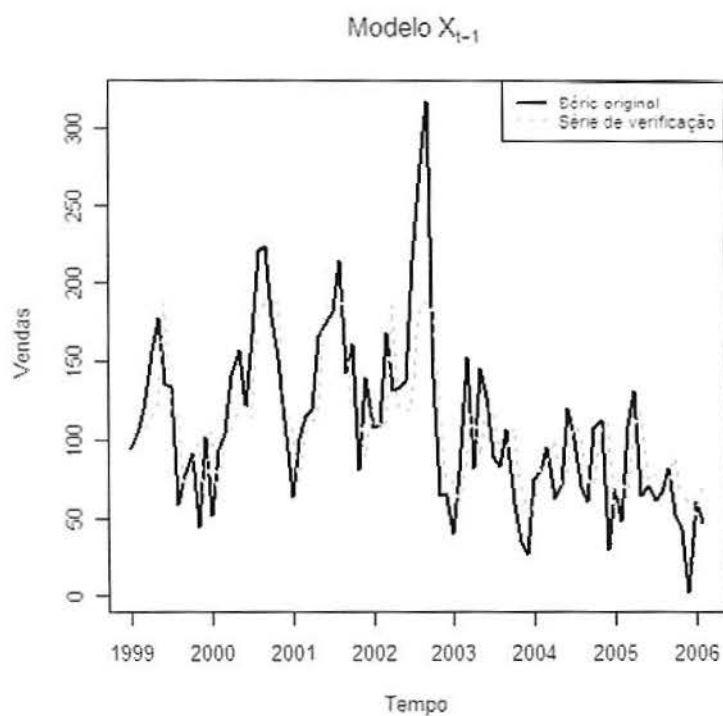
Fonte: elaboração própria

O modelo, ou modelos, a serem escolhidos nesta primeira análise devem ser os que possuem os menores valores de AIC e BIC. Entretanto, a escolha de um ou mais modelos dentre os 24 feitos é subjetiva devido ao fato de que não basta obter uma média entre os critérios para se decidir por um ou outro modelo. Um bom exemplo disso é que o modelo com menor AIC ( $X(t-1)$ ,  $X(t-12)$  e *variáveis indicadoras*) não é o mesmo que o modelo com menor BIC ( $X(t-1)$ ). Entretanto, assim como outras técnicas não paramétricas, baixos AIC e BIC não necessariamente apresentam o melhor modelo. Modelos com AIC (ou BIC) muito baixos podem sofrer de *super treinamento*, ou seja, a série é muito bem modelada, mas perde poder de previsão. Isso costuma ocorrer quando uma parte demasiadamente grande da série é usada no treino. Como exemplo extremo, se a série toda fosse usada no treinamento, provavelmente o modelo se adequaria com perfeição, mas seu poder de previsão seria reduzido. Há uma regra geral (*rule of thumb*) que se deve usar entre 60% e 80% da amostra para treino. A amostra de treino foi de 65 valores dentre os 86 disponíveis, algo em torno de 75,5% dos dados disponíveis.

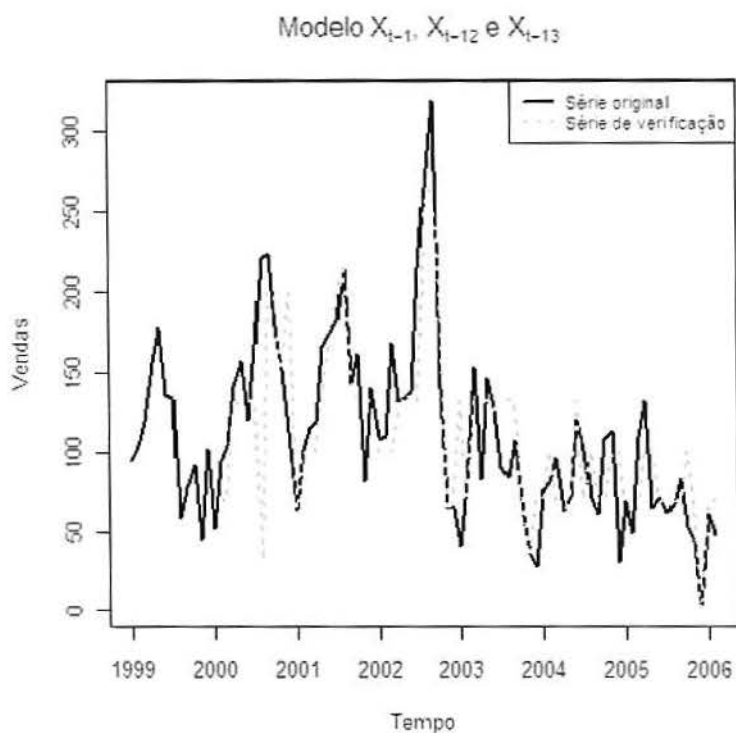
Sendo assim, é necessário escolher algum modelo que contemple tanto o critério de seleção baixo (AIC e BIC) e que possua boas estimativas. Explicar por que foram escolhidos os 4 modelos (referenciar os valores de AIC e BIC nas tabelas) Da Figura 8 até a Figura 11 são apresentados os gráficos dos 4 modelos selecionados.



**Figura 8 - Modelo com 6 neurônios na camada oculta**  
**Fonte: elaboração própria baseada nos dados obtidos junto a ANFAVEA.**

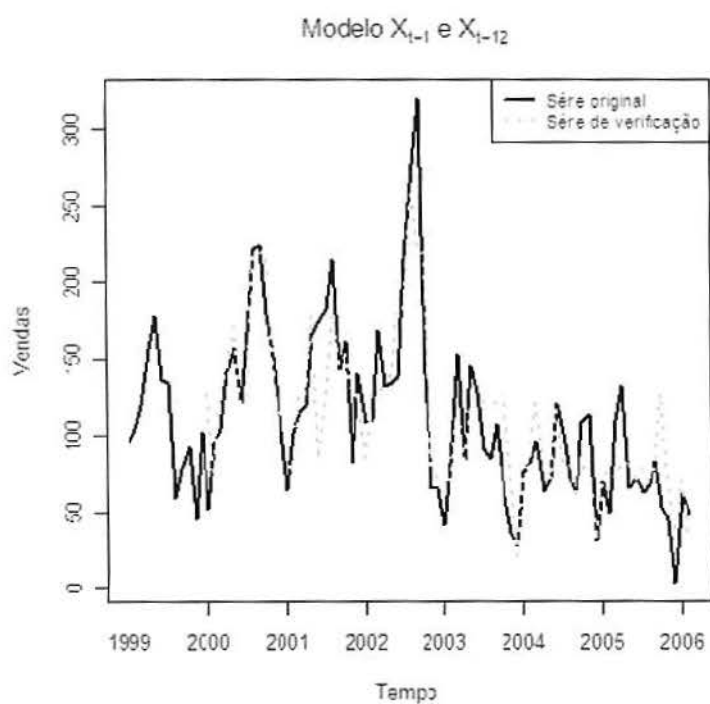


**Figura 9 - Modelo com 3 neurônios na camada oculta**  
**Fonte: elaboração própria baseada nos dados obtidos junto a ANFAVEA.**



**Figura 10 - Modelo com 4 neurônios na camada oculta**

Fonte: elaboração própria baseada nos dados obtidos junto a ANFAVEA.



**Figura 11 - Modelo com 6 neurônios na camada oculta**

Fonte: elaboração própria baseada nos dados obtidos junto a ANFAVEA.



### 3.8 Escolha do melhor modelo e as previsões

Pelos gráficos se nota que estes três modelos se ajustaram bem a série de validação. O próximo passo agora é verificar as previsões.

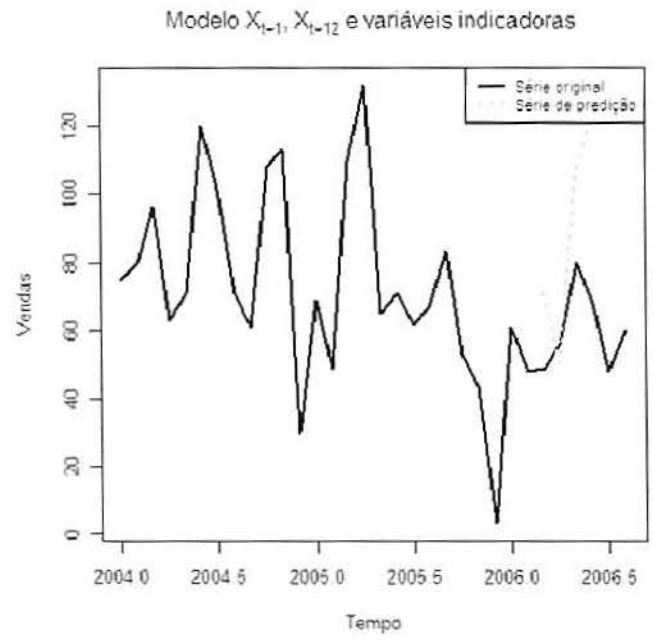
Quanto as previsões, os 4 modelos selecionados apresentaram boa acurácia. Segundo o especialista da própria empresa, erros percentuais de previsão inferiores a 50% já seriam adequados. A Tabela 5 mostra os verdadeiros valores da série seguidos das previsões realizadas pelos modelos, lembrando que, para isso, foi suposto que série de vendas não existia após fevereiro de 2006.

**Tabela 5 - Previsões e erros dos modelos selecionados**

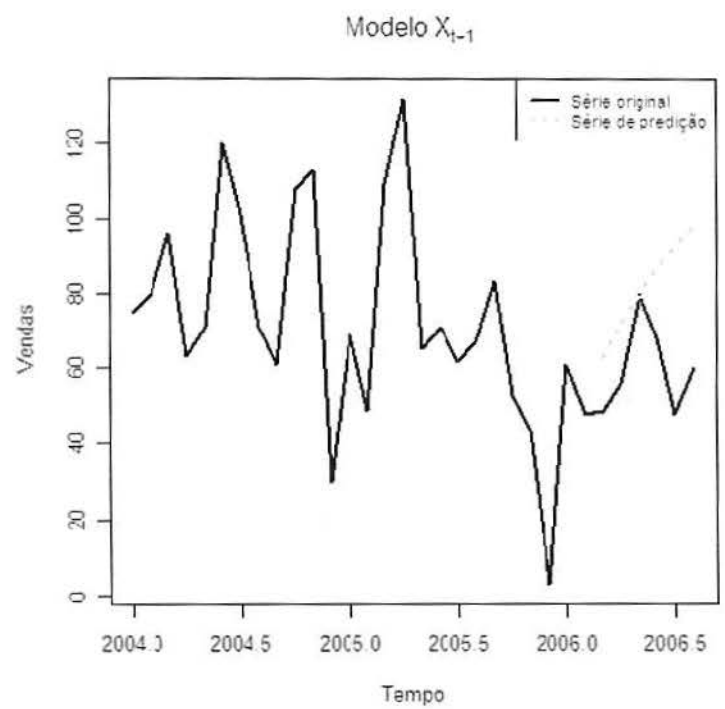
Data	$X(t-1)$ c/ 3 neurônios		$X(t-1), X(t-12)$ e $X(t-13)$ c/ 4 neurônios		$X(t-1)$ e $X(t-12)$ c/ 6 neurônios		$X(t-1), X(t-12)$ mais indicadoras c/ 6 neurônios		Valores verdadeiros
	Previsão	Erro absoluto %	Previsão	Erro absoluto %	Previsão	Erro absoluto %	Previsão	Erro absoluto %	
março-06	63	28,57%	70	42,86%	70	42,86%	72	46,94%	49
abril-06	72	28,57%	70	25,00%	113	101,79%	51	8,93%	56
maio-06	80	0,00%	1	98,75%	81	1,25%	106	32,50%	80
junho-06	87	27,94%	70	2,94%	103	51,47%	123	80,88%	68
julho-06	93	93,75%	70	45,83%	81	68,75%	181	277,08%	48
agosto-06	98	63,33%	101	68,33%	77	28,33%	137	128,33%	60

Fonte: elaboração própria e ANFAVEA (2006)

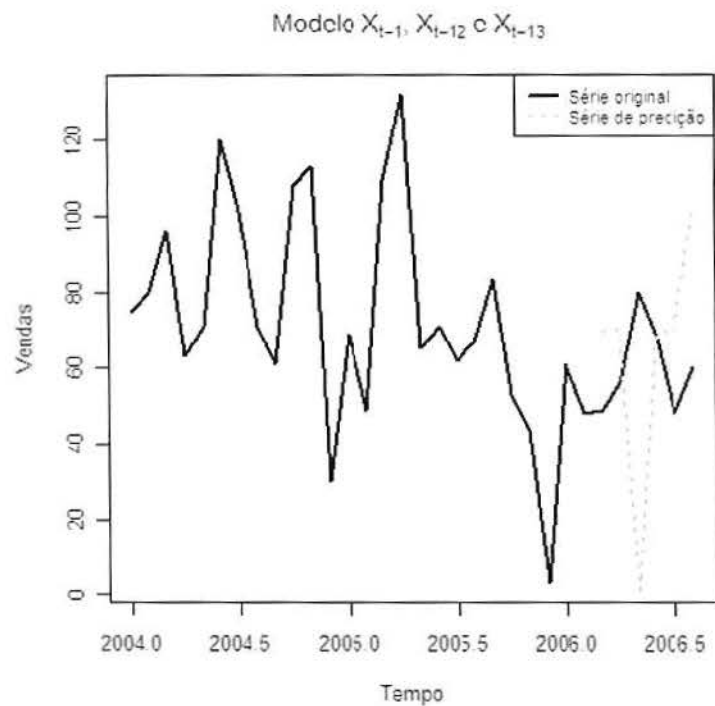
Observando-se os erros absolutos percentuais e os valores preditos, os modelos estão bem se comparados ao que a empresa obtém com as suas estimativas. Com exceção do modelo  $X(t-1), X(t-12)$  mais indicadoras que obteve três estimativas com erros superiores ao que a empresa obtém e também pelo modelo com  $X(t-1), X(t-12)$  e  $X(t-13)$ , que obteve, basicamente, apenas 3 valores de estimativas de vendas. Quanto aos outros dois modelos, os mesmos obtiveram pelo menos uma boa e uma péssima estimativa das vendas. Mas também foi realizada uma análise visual destas predições. Da Figura 12 a Figura 15 são apresentados os gráficos destes modelos com relação as previsões.



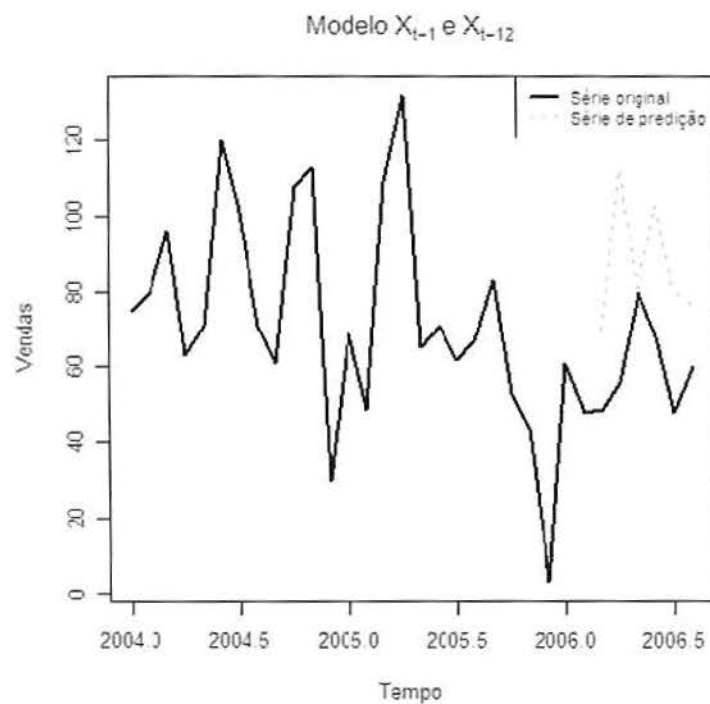
**Figura 12 - Gráfico com as previsões do modelo  $X(t-1)$ ,  $X(t-12)$  e indicadores**  
Fonte: elaboração própria baseada nos dados obtidos junto a ANFAVEA.



**Figura 13 - Gráfico com as previsões do modelo  $X(t-1)$**   
Fonte: elaboração própria baseada nos dados obtidos junto a ANFAVEA.



**Figura 14 - Gráfico com as previsões do modelo  $X(t-1)$ ,  $X(t-12)$  e  $X(t-13)$**   
**Fonte: elaboração própria baseada nos dados obtidos junto a ANFAVEA.**



**Figura 15 - Gráfico com as previsões do modelo  $X(t-1)$  e  $X(t-12)$**   
**Fonte: elaboração própria baseada nos dados obtidos junto a ANFAVEA.**

Mesmo com essa análise visual, ainda pode permanecer a dúvida entre qual o melhor modelo, principalmente com relação ao modelo  $X(t-1)$  e o modelo  $X(t-1)$  e  $X(t-12)$ . Sendo assim, mais uma análise é feita, comparando estes dois modelos. Como as previsões da empresa são feitas para, no máximo, dois meses a frente, então o mesmo deve ser feito com os modelos. A Tabela 6 apresenta as previsões com dois passos a frente de cada um dos dois modelos restantes e, logo a seguir, a Tabela 7 apresenta os critérios AIC, BIC e SSE de cada um dos modelos.

**Tabela 6 - Erros e previsões dos dois modelos escolhidos**

Data	$X(t-1)$ c/ 3 neurônios		$X(t-1)$ e $X(t-12)$ c/ 6 neurônios		Valores verdadeiros
	Previsão	Erro absoluto %	Previsão	Erro absoluto %	
março-06	63	28,57%	70	42,86%	49
abril-06	72	28,57%	113	101,79%	56
maio-06	73	8,75%	77	3,75%	80
junho-06	76	11,76%	79	16,18%	68
julho-06	93	93,75%	81	68,75%	48
agosto-06	84	40,00%	45	25,00%	60

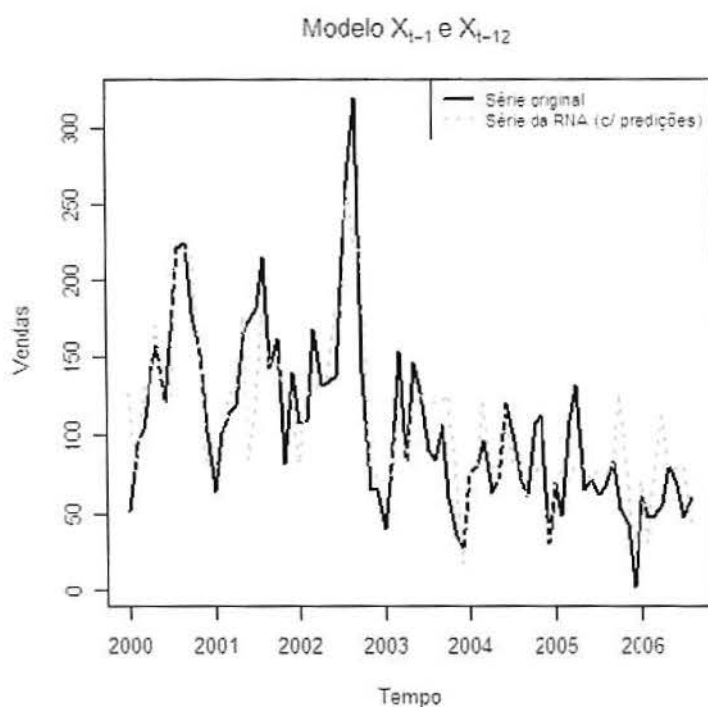
Fonte: elaboração própria e ANFAVEA (2006)

**Tabela 7 - Critérios de seleção dos modelos**

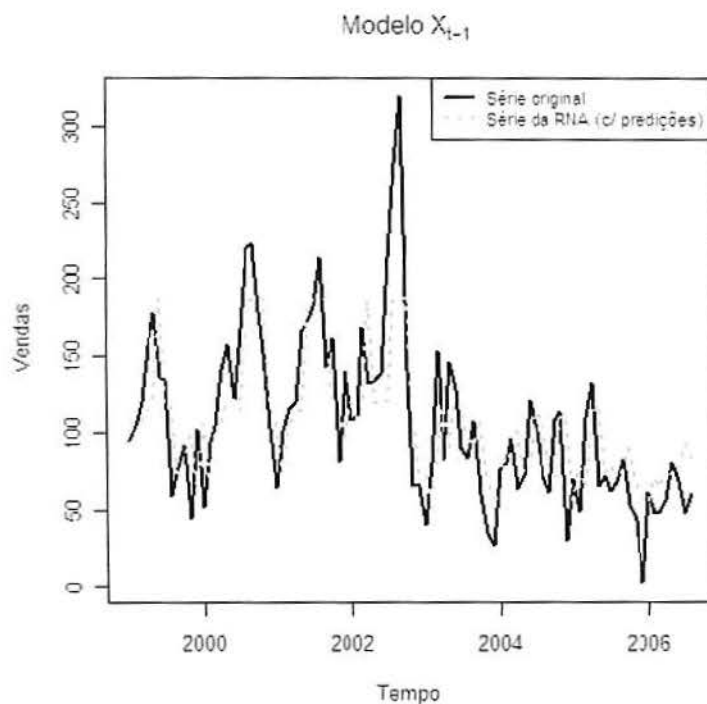
Modelos	Critérios		
	AIC	BIC	SSE das previsões
$X(t-1)$ c/ 3 neurônios	7,259	7,830	3.111,827
$X(t-1)$ e $X(t-12)$ c/ 6 neurônios	6,837	8,343	5.037,572

Fonte: elaboração própria

De posse destas destas informações, o modelo escolhido, entre os dois restantes, fica sendo o modelo  $X(t-1)$  com 3 neurônios na camada oculta. Este também é o modelo com menor número de parâmetros (1 neurônio na camada de entrada e 3 na camada oculta), totalizando 10 parâmetros. Além disso, a soma do erro quadrático das previsões do modelo  $X(t-1)$  é menor do que a do outro modelo. O único critério que dá alguma vantagem ao modelo  $X(t-1)$  e  $X(t-12)$  é o critério AIC. Porém apesar deste critério penalizar o modelo com relação ao número de parâmetros, o critério BIC é mais punitivo quanto a quantidade de parâmetros. Este modelo escolhido é muito semelhante, devido a variável usada no neurônio da camada de entrada, a um AR(1), um modelo da família ARIMA de modelos. A seguir estão os gráficos dos dois modelos, já com as previsões de 2 passos adiante.



**Figura 16 - Gráfico do modelo  $X(t-1)$  e  $X(t-12)$  já com as previsões, com 2 passos a frente**  
**Fonte: elaboração própria baseada nos dados obtidos junto a ANFAVEA.**



**Figura 17 - Gráfico do modelo  $X(t-1)$  já com as previsões, com 2 passos a frente**  
**Fonte: elaboração própria baseada nos dados obtidos junto a ANFAVEA.**

## CONCLUSÃO

Este trabalho teve como objetivo principal obter um modelo de redes neurais capaz de realizar previsões de vendas para um produto específico da AGCO do Brasil Comércio e Indústria Ltda. O produto escolhido foi um trator da marca Massey Ferguson, modelo 275, que tem como nome comercial MF 275. Para prever as vendas deste produto foram feitos 24 modelos diferentes. Ainda fazia parte deste objetivo que o modelo fosse simples, permitindo seu uso junto à empresa. Trabalhos anteriores, como Shlens (1999) e Fernandes et al. (1995), já indicavam que este tipo de modelagem costuma obter boas previsões.

Entretanto, ao contrário do que mostra o estudo de Fernandes et al. (1995), os modelos com variáveis indicadoras acabaram sendo relegados por um modelo mais simples, apesar de ter obtido um AIC bem inferior aos demais modelos. Suas previsões estavam inadequadas quanto ao objetivo de se obter um modelo mais preciso que o que a empresa utiliza (erros percentuais absolutos inferiores a 50%). Tais variáveis indicadoras foram usadas por Fernandes et al. (1995) para captar a sazonalidade da série.

O modelo apresentado pela empresa se mostrou inferior, em termos da precisão da previsão, em relação ao modelo final obtido. Em resumo, o modelo final escolhido teve boas previsões (com baixo erro absoluto percentual) se comparadas ao que a empresa costuma obter (em torno de 50% de erro absoluto percentual), mesmo a série sendo de um tamanho relativamente pequeno para o uso da modelagem por RNA.

O modelo final escolhido não só é o mais simples dentre os apresentados, como não se utiliza de variáveis externas à série (como custeio de produção, preço, demanda, entre outras), que poderiam tornar demasiado pesado e complexo para seu uso junto à empresa. A modelagem por redes neurais artificiais é adequada, neste caso pelo menos, exatamente por não exigir variáveis externas a própria série.

Em relação a pesquisas futuras, como sugestão, a idéia de se abordar outros modelos, com mais ou menos neurônios na camada oculta e com outras variáveis

para os neurônios da camada de entrada. Modelos com mais de uma camada oculta também devem ser analisados.

Apesar do modelo escolhido possuir erros absolutos percentuais menores que o modelo utilizado pela empresa, os mesmos poderiam melhorar significativamente se outras variáveis fossem utilizadas, como o clima, custeio de produção e outras que foram citadas durante o trabalho, mas que não foram verificadas. Tais contribuições podem ser investigadas futuramente.

## REFERÊNCIAS

- ABRAHAM, B; LEDOLTER, J. **Statistical methods for forecasting** (Wiley series in probability and statistics). New York: John Wiley & Sons, 2005. 445 p.
- ANDERSON, T. W. **The statistical analysis of time series** (Wiley series in probability and statistics). New York: John Wiley & Sons, 1971. 704 p.
- ARMSTRONG, J. S., COLLOPY, F.; YOKUM, J. T. Decomposition by casual forces: a procedure for forecasting complex time series. **International Journal of Forecasting**. Volume 21, 2005, p.25-36.
- ARMSTRONG, J. S., MORWITZ, V. G.; KUMAR, V. Sales forecasts for existing consumer products and services: Do purchases intentions contribute to accuracy? **International Journal of Forecasting**. Volume 16, 2000, p.383-397.
- BOYD, M.; KAASTRA, I. Designing a neural network for forecasting financial and economic time series. **Neurocomputing**, Volume 10, 1996. p. 215-236.
- BOX, G.; JENKINS, G. **Time series analysis: forecasting and control**. San Francisco: Holden-Day, 1976.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis: forecasting and control**. 3rd. ed. Englewood Cliffs: Prentice Hall, 1994. 592 p.
- BRAUERS, J.; WEBER, M. A New Method of Scenario Analysis for Strategic Planning. **Journal of Forecasting**, Volume 7, Jan-Mar 1988, p. 31-47.
- CONGDON, P. **Applied Bayesian modelling** (Wiley series in probability and statistics). New York: John Wiley & Sons, 2003. 465 p.
- CONOVER, W. J. **Practical Nonparametric Statistics**. New York: John Wiley & Sons, 1971. 491 p.
- DINARDO, J.; JOHNSTON, J. **Métodos econométricos**. Tradução e revisão: Manuela Magalhães Hill, Fátima Ferrão, Rui Menezes. 4. ed. Lisboa: McGraw-Hill, 2001. Título original: Econometric methods
- EDMUNDSON, B., LAWRENCE, M.; O'CONNOR, M. The Use of Non-time Series Information in Sales Forecasting: a Case Study. **Journal of Forecasting**, Volume 7, Jul-Sep 1988, p. 201-211.
- ENDERS, W. **Applied econometric time series** (Wiley series in probability and statistics). New York: John Wiley & Sons, 1995. 433 p.



FERNANDES, L. G. L.; PORTUGAL, M. S.; NAVAUX, P. O. A. Previsão de series de tempo: redes neurais artificiais e modelos estruturais. **Pesquisa e Planejamento Econômico**, Volume 26, n. 2, 1996, p. 253-276.

FERNANDES, F. C. F.; SANTORO, M. C. Avaliação do grau de prioridade e do foco do planejamento e controle da produção (PCP): modelos e estudos de casos. **Gestão & Produção**, Volume 12, n. 1, p. 25-38.

FRANSES, P. H. **Time series models for business and economic forecasting**. Cambridge: Cambridge University Press, 1998. 288 p.

FREES, E. W.; MILLER, T. W. Sales forecasting using longitudinal data models. **International Journal of Forecasting**, Volume 20, 2004, p.99-114.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002. 175 p.

GONÇALVES FILHO, E. V.; MARÇOLA, J. A. Annualized hours as a capacity planning tool in make-to-order or assemble-to-order environment: an agricultural implements company case. **Production Planning & Control**, Volume 12, n. 4, p. 388-398.

HAMILTON, J. D. **Time series analysis**. Princeton: Princeton University Press, 1994. 799 p.

HAMILTON, J. D.; RAJ, B. **Advances in Markov-switching models**: applications in business cycle research and finance (Studies in empirical economics). 1st. ed. Heidelberg: Physica-Verlag, 2004.

HAYKIN, S. **Redes neurais**: princípios e práticas. Tradução de Paulo Martins Engle. 2. ed. Porto Alegre: Bookman, 2001. 900 p. Título original: Neural networks: a comprehensive foundation, 2/E.

JAIN, C. L. How to determine the approach to forecasting. **The Journal of Business Forecasting Methods & Systems**, Volume 14, 2005, p. 28-29.

KIM, C. J.; NELSON, C. R. **State-space models with regime switching**: classical and Gibbs-sampling approaches with applications. Cambridge: MIT Press, 1999.

KOTLER, P. **Administração de marketing**: análise, planejamento, implementação e controle. Tradução de Ailton Bomfim Brandão. 5. ed. São Paulo: Atlas, 1998. 725 p. Título original: Marketing management: analysis, planning, implementation, and control.

KOVÁCS, Z. L. **Redes neurais artificiais**: fundamentos e aplicações. São Paulo: Edição Acadêmica, 1996. 164 p.

KROLZIG, H. M. **Markov-switching vector autoregressions**: modelling, statistical inference, and application to business cycle analysis (Lecture notes in economics and mathematical systems). Telos: Springer-Verlag, 1997.

KUYVEN, P. S. **Proposta de um método para a análise de demanda**: aplicação numa indústria de brinquedos. 2004. 172 p. Dissertação (Mestrado em Engenharia da Produção). Escola de Engenharia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2004.

LAWRENCE, M.; O'CONNOR, M. Sales forecasting updates: how good are they in practice? **International Journal of Forecasting**, Volume 16, 2000, p. 369-382.

LEMOS, F. O. **Metodologia para seleção de métodos de previsão de demanda**. 2006. 183 p. Dissertação (Mestrado em Engenharia de Produção). Escola de Engenharia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006.

LUCKO, G. **A statistical analysis and model of the residual value of different types of heavy construction equipment**. 2003. 292 p. Dissertação (Doctor of Philosophy in Civil Engineering). Faculty of Virginia Polytechnic Institute and State University, Blacksburg, 2003.

MANCUZO, F. **Análise e previsão de demanda**: estudo de caso em uma empresa distribuidora de rolamentos. 2003. 142 p. Trabalho de conclusão de curso (Mestrado Profissionalizante com Ênfase em Logística). Escola de Engenharia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2003.

MANTOVANI, G. F. **Previsão de séries temporais: redes neurais artificiais vs. modelos arima**. 2004. 62 p. Monografia (Graduação em Estatística). Instituto de Matemática, Departamento de Estatística, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006.

MATHER, H. **Competitive manufacturing**. Englewood Cliffs: Prentice Hall, 1988, 320 p.

MIN QI, ZHANG, G. P. An investigation of model selection criteria for neural network time series forecasting. **European Journal of Operational Research**, n. 132, 2001, p. 666-680.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. 2. ed. São Paulo: Edgard Blücher, 2006. 538 p.

NETER, J. et al. **Applied linear statistical models**. 4th. ed. New York: McGraw-Hill/Irwin, c1996. 1396 p.

PASQUAL, C. A.; PEDROZO, E. A. **O posicionamento logístico de máquinas e implementos agrícolas na mesorregião noroeste do estado do Rio Grande do Sul**. 2005. 20 p. Texto para discussão n. 8. Universidade de Passo Fundo, Faculdade de Ciências Econômicas, Administrativas e Contábeis, Centro de Pesquisa e Extensão da FEAC, Passo Fundo, 2005. Disponível em: <<http://www.upf.br/cepeac>>. Acesso em: outubro, 2006.

PINTO, F. C. A., MEDEIROS, M. C. Misspecified neural network models and linear time series forecasting. **Learning and Nonlinear Models – Revista Brasileira de Redes Neurais**, Volume 1, n. 5, 2005, p. 317-326.

POLLOCK, D. S. G. **A handbook of time-series analysis, signal processing and dynamics**. San Diego: Academic Press, 1999. 808 p.

PRICE, D. H. R.; SHARP, J. A. The Impact on the Performance of Growth Curve Models of Changes in Parameter Re-estimation Period and Other Factors. **Journal of Forecasting**, Volume 7, Oct-Dec 1988, p. 245-258.

R Development Core Team. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2006. URL <http://www.R-project.org>.

SAS/ETS User's Guide. **The AUTOREG procedure**. Disponível em: <<http://www.sussex.ac.uk/its/help/guides/sas/ets/index.htm>>. Acesso em: 8 de outubro de 2006.

SELAU, L. P. R. **Redes neurais artificiais no contexto estatístico**. 2000. 108 p. Monografia (Graduação em Estatística). Instituto de Matemática, Departamento de Estatística, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2000.

SHINGO, S. **Study of Toyota production systems**. Tokyo: Japan Management Association, 1981.

SHLENS, J. Time series prediction with artificial neural networks, 1999. Arquivo no formato pdf obtido na URL <http://citeseer.ist.psu.edu/shlens99time.html>. Acessado em 25 de outubro de 2006.

SHUMWAY, R. H. **Applied statistical time series analysis**. Englewood Cliffs: Prentice Hall, 1988. 379 p.

SLACK, N.; CHAMBERS, S.; JOHNSTON, R. **Administração da produção**. Tradução de Maria Teresa Corrêa de Oliveira, Fábio Alher; revisão técnica Henrique Luiz Corrêa. 2. ed. São Paulo: Atlas, 2002. 747 p. Título original: Operations management.

SNYDER, R. D., KOEHLER, A. B.; ORD, J. K. Forecasting for inventory control with exponential smoothing. **International Journal of Forecasting**, Volume 18, 2002, p. 5-18.

TERAESVIRTA, T. LIN, C.F., GRANGER, C. W. J. Power of the neural network linearity test. **Journal of Time Series Analysis**, Volume 14, 1993, p. 209-220.

TRAPLETTI, A., HORNIK, K. **tseries**: time series analysis and computacional finance. R package version 0.10-4, 2006.

VENABLES, W. N., RIPLEY, B. D. **Modern applied statistics with S**. 4th ed. New York: Springer, 2002. 495 p.

WANKE, P. Metodologia para a gestão de estoques de peças de reposição: um estudo de caso em empresa brasileira. **Revista Tecnológica**, dezembro 2005, p. 60-65.

YELLAND, P. M. Stable seasonal pattern models for forecast revision: A comparative study. **International Journal of Forecasting**, (ainda não publicado). Disponível em: <[www.sciencedirect.com](http://www.sciencedirect.com)> desde 28 de agosto de 2006. Acesso em: 8 de setembro de 2006.