

---

**COOPERAÇÃO ENTRE REDES NEURAIS ARTIFICIAIS E  
TÉCNICAS ‘CLÁSSICAS’ PARA PREVISÃO DE DEMANDA DE  
UMA SÉRIE DE VENDAS DE CERVEJA NA AUSTRÁLIA**

**Guilherme Marques Calôba \***

EP/COPPE  
Universidade Federal do Rio de Janeiro  
Rio de Janeiro – RJ  
E-mail: gcaloba@mail.com

**Luiz Pereira Calôba**

EP/COPPE  
Universidade Federal do Rio de Janeiro  
Rio de Janeiro – RJ

**Eduardo Saliby**

COPPEAD  
Universidade Federal do Rio de Janeiro  
Rio de Janeiro – RJ

*\* Corresponding author/autor para quem as correspondências devem ser encaminhadas*

*Recebido em 07/2002, aceito em 12/2002 após 1 revisão*

**Resumo**

O principal objetivo deste trabalho é avaliar a complementação do uso de técnicas de previsão de vendas com Redes Neurais Artificiais. Como aplicação, foi analisada a demanda industrial no setor de cerveja da Austrália. A maioria dos modelos de previsão atuam de forma isolada, ou seja, tratando problemas por enfoques que se excluem. A sugestão deste trabalho é utilizar estes métodos de forma cooperativa, buscando alcançar melhores resultados que aqueles obtidos individualmente. O trabalho inicia-se com considerações sobre previsões de séries temporais, caracterização breve sobre redes neurais e outros métodos de previsão. A seguir, o estudo de caso: uma série histórica de vendas de cerveja na Austrália. Encerra-se com uma conclusão e recomendações, e lista de referências bibliográficas.

**Palavras-chave:** previsão de demanda, redes neurais artificiais, cooperação.

**Abstract**

The main goal of this article is to evaluate the complementation between forecasting techniques, in particular Neural Networks. The environment of the application is a foreign industrial sector, sales of beer in Australia. Most of the forecasting models act in a separate way, that is, treating problems through different views that exclude one another. The suggestion of this work is to use the methods in a cooperative way, looking forward to achieve better results. The article begins with some considerations on Time Series Studies, followed by a brief characterization of Neural Networks and the other methods used on this work. Continuing, the case is presented and discussed and finishing, conclusions and a list of references.

**Keywords:** demand forecasting, artificial neural networks, cooperation.

## 1. Introdução

O principal objetivo deste artigo é avaliar a complementação do uso de técnicas de previsão de vendas mais conhecidas com outras menos difundidas, em particular Redes Neurais Artificiais.

Dado que existe um planejamento em uma organização, quer empresa ou país, existe espaço para uma previsão de demanda. A maioria das organizações existe para servir ao consumidor, em suas mais diversas formas. Prever a demanda significa estimar a resposta do consumidor e planejar medidas para atender estas necessidades.

Este trabalho buscou combinar técnicas lineares e não-lineares para a melhor previsão de demanda, obtendo resultados positivos e interessantes na cooperação destas técnicas.

## 2. Considerações sobre Previsão de Séries Temporais

Todo problema de previsão possui uma metodologia própria de resolução. Neste trabalho, serão utilizadas algumas características e definições particulares. Para se ter uma idéia do quanto uma previsão é “precisa”, é necessário haver uma forma de estimar o quanto se está errando, em média. Estes valores de imprecisão geralmente se originam da Estatística, como é ilustrado, por exemplo, em Wonnacott & Wonnacott (1990). Este erro é calculado de duas formas distintas neste presente trabalho:

a) Erro Médio Percentual (MAPE): Consiste na seguinte expressão:

$$MAPE = \left( \frac{1}{N} \right) \sum_{i=1}^N |(y_i - y_{pi}) / y_i| \quad (1)$$

onde

- $y_i$  é o valor real do período  $i$ ;
- $y_{pi}$  é a previsão para o período  $i$  e
- $N$  é o número de períodos no conjunto de validação.

b) Raiz do Erro Médio Quadrático Percentual (REMQP): Consiste no cálculo do desvio padrão percentual observado, segundo a fórmula:

$$REMQP = \left\{ \left( \frac{1}{N} \right) \sum_{i=1}^N [(y_i - y_{pi}) / y_i]^2 \right\}^{1/2} \quad (2)$$

onde

- $y_i$  é o valor real do período  $i$ ;
- $y_{pi}$  é a previsão para o período  $i$  e
- $N$  é o número de períodos no conjunto de validação.

No restante deste trabalho, a raiz do erro médio quadrático percentual será denotada como EMQP. Foram utilizados neste trabalho dois métodos de previsão clássica, a saber: a regressão múltipla e a decomposição clássica, incluindo cálculo de tendência, sazonalidade aditiva e remoção de ciclos por análise de frequência (Transformada de Fourier).

### 3. Caracterização das Redes Neurais Utilizadas no Trabalho

A rede utilizada neste trabalho é do tipo *feedforward*. Esse é um tipo de rede já bastante conhecido e experimentado, e o mais largamente utilizado atualmente. Em redes do tipo *feedforward* os neurônios são dispostos em camadas, usualmente duas camadas ativas. Os neurônios de uma camada só se conectam com os da camada subsequente. Pode ser destacada na estrutura de uma rede desse gênero 3 tipos de “camadas”.

A primeira camada, ou “camada de entrada”, não é composta por neurônios, mas simplesmente pelas conexões dos sinais de entrada. A camada seguinte, conhecida como “camada intermediária” ou “camada escondida”, é composta por  $N$  neurônios, e realiza algumas transformações. Ao fim existe uma camada conhecida como “camada de saída”. Existem tantos neurônios nessa camada quanto o número de saídas desejadas, cada saída representando uma variável que se deseja prever.

A seguir tem-se a figura 1 representa uma rede do tipo *feedforward* hipotética, apresentando  $n$  entradas, uma única camada intermediária, com  $Q$  neurônios e  $m$  neurônios na camada de saída.

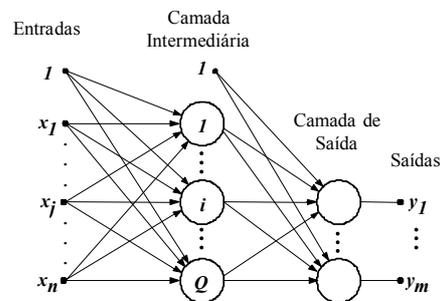


Figura 1 – Rede neural *feedforward* de duas camadas

Prova-se que a rede neural com duas camadas de neurônios possui a capacidade de realizar qualquer mapeamento entrada-saída desejado, desde que unívoco, e desde que o número de neurônios na camada intermediária seja suficiente, como cita, por exemplo, Hecht-Nielsen (1990).

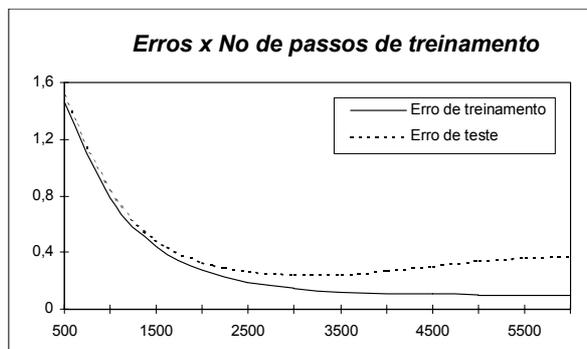
#### 3.1 Conjuntos de Teste, Treinamento e Validação

Em qualquer problema a ser resolvido através do uso de redes neurais, é necessário o uso de pares de entrada-saída já conhecidos. Muitas vezes a quantidade de pares disponíveis não é muito grande, mesmo assim deve-se separar esses dados em dois conjuntos bem definidos, o conjunto de treinamento e o conjunto de teste.

O treinamento da rede é feito utilizando o conjunto de treinamento. É, entretanto, necessário medir a performance da rede, considerando como ela responde a pares entrada-saída não apresentados durante o treinamento (o conjunto de teste).

Isso é importante pois um treinamento prolongado demais gera um problema conhecido como *overtraining*, que pode levar a uma super especialização da rede (principalmente quando se dispõe de muito poucos dados) e a uma perda da capacidade da rede de responder bem a dados nunca apresentados (perda da capacidade de generalização). A figura a seguir mostra claramente a evolução típica do erro de uma rede ao longo do processo de treinamento, tanto do erro do conjunto de treinamento como do conjunto de teste. A partir de um certo número

de passos, a performance do conjunto de teste pára de decair e começa a piorar, mesmo que a do conjunto de treinamento continue decaindo. Inicia-se a super especialização. É virtualmente aceito que a melhor rede é aquela que fornece o menor erro no conjunto de validação (definido a seguir) é aquela que fornece o mínimo erro para o conjunto de teste.



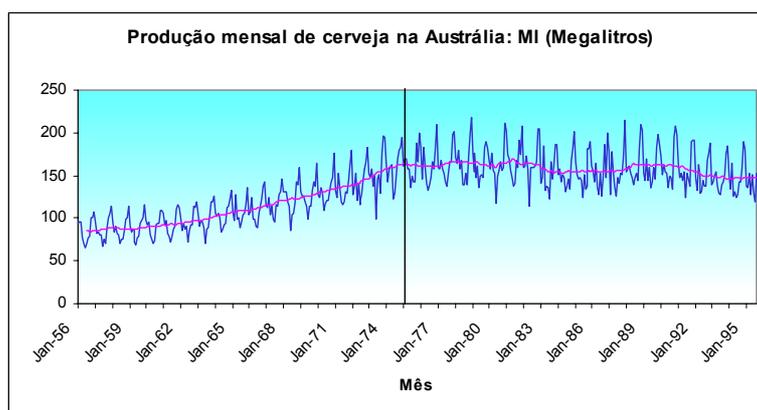
**Figura 2** – Evolução do erro durante o treinamento

O conjunto de validação corresponde a pares entrada-saída que não são apresentados à rede durante o treinamento e não são parâmetro para encerrar o treinamento, ou seja, representam um conjunto totalmente novo a ser apresentado à rede para avaliação de desempenho.

#### 4. Estudo de Previsão de Vendas: Vendas de Cerveja na Austrália

Este trabalho contempla a previsão comparativa em produção de cerveja na Austrália. Desta forma, a exemplificação dos métodos e sua comparação nos casos específicos mostra-se de forma mais clara. A proposta do trabalho é utilizar primeiramente os métodos clássicos de previsão e, em seguida, a Rede Neural.

A série utilizada para este exemplo consiste em observações da produção de cerveja, em megalitros, em todo o território da Austrália. Baseado em uma série de mais de quarenta anos, se partiu para a previsão dos 20 últimos meses. A seguir, a série de vendas.



**Figura 3** – Série de Vendas de Cerveja na Austrália

Esta série possui sazonalidade muito marcante, o que leva a crer que um método como a decomposição consiga alcançar um resultado satisfatório no tratamento. O estudo de componentes de ciclo e de tendência deve conduzir a um resultado ainda melhor.

Dado o fato que a série é muito longa e possui uma tendência que se modifica claramente aproximadamente em sua metade, foi utilizada para previsão apenas informações de 1976 em diante, o que ainda fornece uma gigantesca base de dados, em comparação ao que geralmente é disponível para previsões. Esta separação está denotada pela linha vertical no meio do gráfico (ver figura 3).

#### 4.1 Tratamento Estatístico Convencional

Inicialmente, utilizando o método dos mínimos quadrados, foi estimada a reta de tendência para os pontos da série. Esta reta é a melhor estimativa linear do conjunto de pontos. Estes pontos foram numerados de 1 a 216, variando desde o mês de janeiro de 1976 até agosto de 1995. A retirada de tendência gerou a seguinte distribuição gráfica para os pontos:

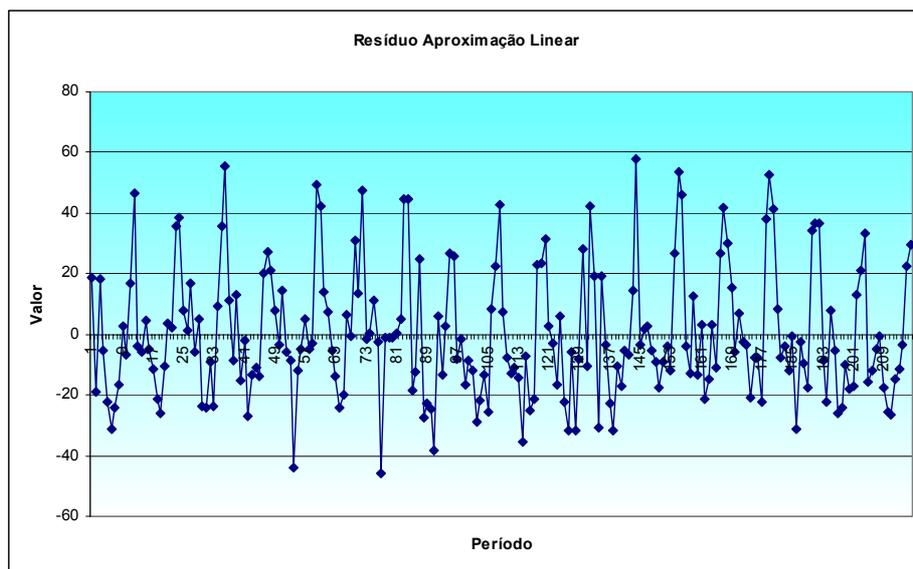


Figura 4 – Resíduo após retirada de nível e tendência

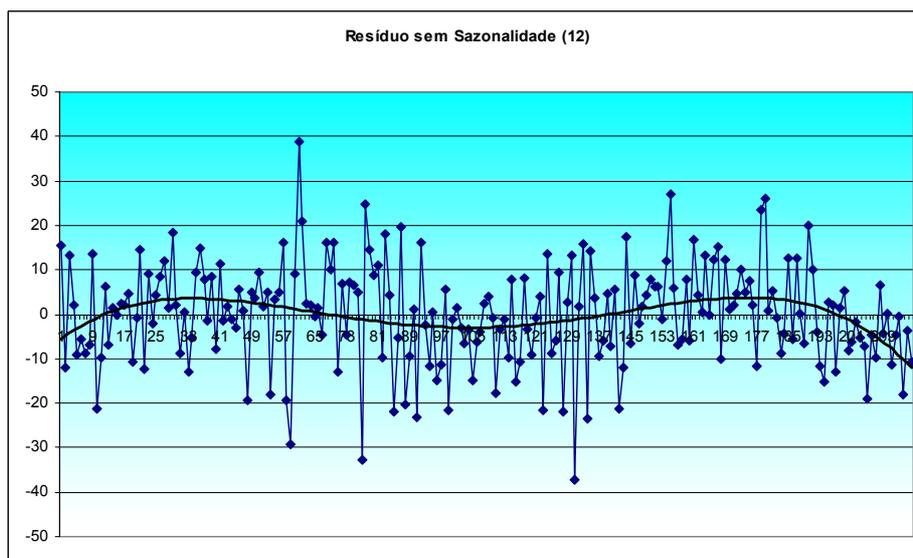
A eliminação da tendência conseguiu reduzir bem a dispersão dos pontos, levando a um somatório quadrado do “resíduo” bem menor. A raiz quadrada da média dos quadrados dos resíduos, que era de 159, passou a ser 21, mostrando que a reta em questão “explica” estatisticamente parte do modelo. Esta estatística (raiz quadrada da média dos quadrados dos resíduos) será denotada deste ponto em diante apenas como “resíduo”.

Verifica-se, no estudo da sazonalidade aditiva, um padrão com uma queda no meio do ano, com uma tendência de maior consumo no verão (final do ano), o que parece uma tendência bastante razoável. O padrão sazonal, representando o quanto acima ou abaixo da média anual cada mês está, em média, pode ser visto abaixo, discriminado para cada mês do ano.



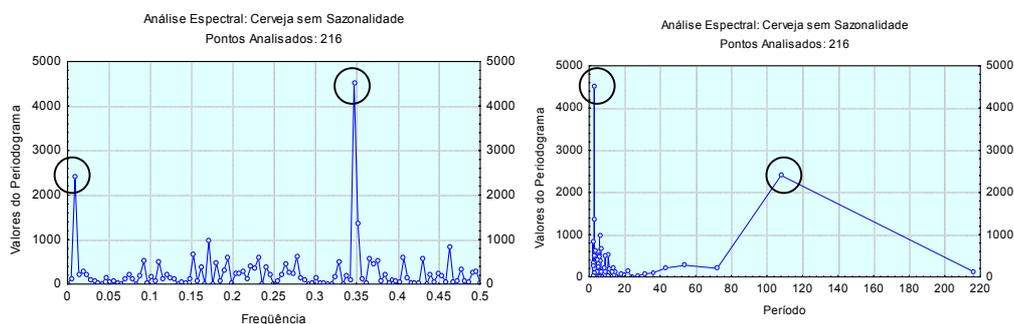
**Figura 5 – Padrão Sazonal**

Uma vez subtraídos os padrões sazonais à série, esta tomou a seguinte distribuição:



**Figura 6 – Resíduo após retirada de sazonalidade anual**

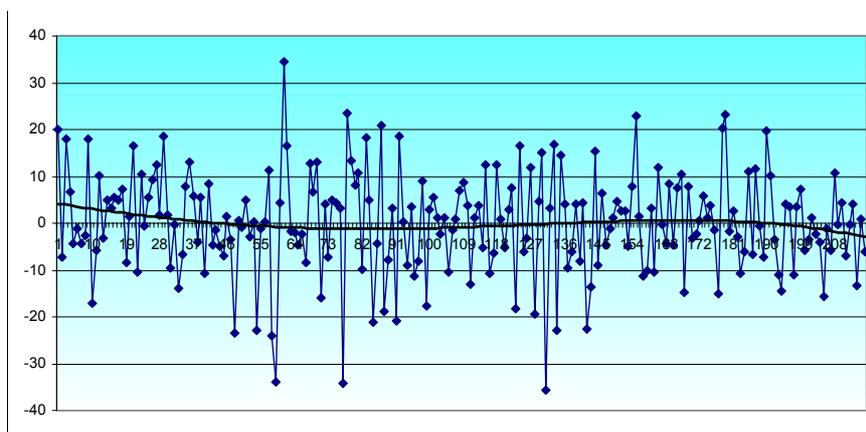
Houve, novamente, queda do resíduo, para 11,44. Em muitos estudos, a decomposição da série se encerraria neste ponto. Neste caso, no entanto, buscou-se refinar a análise através do estudo de ciclos, viabilizando a possível eliminação de sazonalidades não inteiras dentro do espectro e também de padrões cíclicos muito longos, não tratáveis pelo estudo clássico de sazonalidade. A análise de Fourier da série indicou o seguinte periodograma, no qual aparecem em destaque os valores com maiores picos de sazonalidade.



**Figura 7** – Periodograma para a série desazonalizada da cerveja

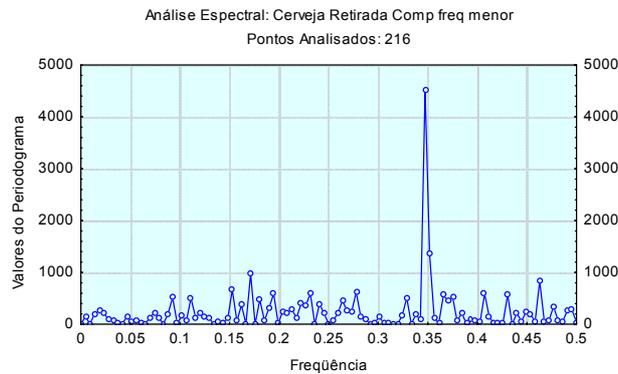
Dois picos de sazonalidade foram estudados: o primeiro, com uma frequência bem alta, de período 2,88, e outro sendo um ciclo de metade da série, período 108 (pontos com destaque na figura 7). É necessário ter muito cuidado no tratamento de ciclo dentro da série, pois alterando períodos muito curtos, pode se introduzir componentes instáveis, levando a problemas na análise. Foi o que ocorreu no tratamento da componente de período 2,88. Neste caso, o desvio padrão dos resíduos, que valia 11,44, passou a ser de 13,26, devido à introdução de componentes não desejadas no espectro.

A componente removida a princípio foi a de período maior, cuja frequência é igual a 0,0093. Seu coeficiente de seno é igual a  $-0,4605$  e o do cosseno é  $-4,6958$ . Esta componente foi removida da forma descrita anteriormente, através da função dada pelos coeficientes e pela frequência. Após este tratamento desta componente, encontrou-se a seguinte distribuição para os pontos da série:



**Figura 8** – Série após retirada da componente de frequência baixa

Após este tratamento, o resíduo caiu novamente para 10,95. É importante ressaltar o fato de que apesar desta queda de resíduo não ser tão significativa, no tratamento posterior do problema através das redes neurais, é fundamental. O periodograma da série está exposto na próxima figura:

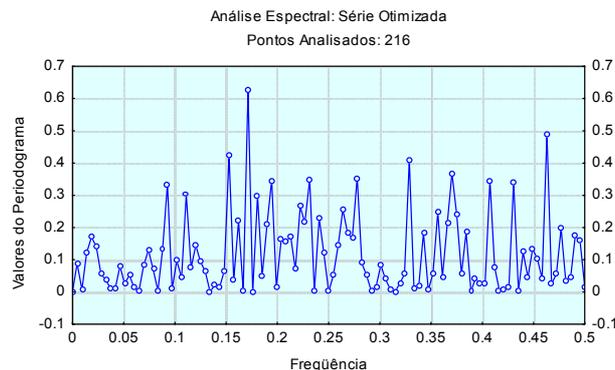


**Figura 9** – Periodograma da Série após retirada da componente de frequência menor

Verifica-se que a componente de menor frequência teve seu sinal totalmente eliminado. A componente de frequência maior, 0,34722, no entanto, não foi removida neste passo. Ao ser removida, a componente aparecia novamente no novo espectro, ainda maior. Observando mais cuidadosamente a figura 9, observa-se a presença de uma outra componente com valor espectral elevado, de frequência pouco acima da componente de maior pico. A hipótese levantada foi de que a componente com valor espectral alto era, na verdade, uma componente intermediária. O embasamento teórico para isso é que, como observado em Chatfield (1989), a análise FFT possui limitações em precisão. Só podem ser representadas por esta análise componentes cujo período pertença ao conjunto  $n/i$ , onde  $n$  é o número de valores observados na série e  $i$  é um número inteiro, variando de 1 a  $n/2$ .

Encontrou-se, no entanto, uma forma de contornar o problema: através de uma otimização realizada com o objetivo de minimizar a soma do resíduo. No caso, a previsão se daria pela soma de tendência, sazonalidade e da componente de frequência menor, já retirada anteriormente.

O otimizador variou coeficiente de senos, cossenos e frequência, procurando minimizar o resíduo. A única condição de contorno inserida foi para a frequência, que deveria se localizar entre 0,34722 e 0,351851, os dois picos observados no periodograma. Após algumas iterações, encontrou-se  $a = 0,1678$ ,  $b = 0,0705$  e  $\omega = 0,34858$ . O valor observado para a função objetivo (resíduo) foi reduzido para  $\varepsilon = 9,6659$ , já multiplicado por 40 (escala original). A prova final do sucesso desta otimização foi dado pela análise espectral está na figura a seguir:



**Figura 10** – Periodograma da Série após retirada da componente otimizada

Todos os valores de densidade espectral estão em uma mesma ordem de grandeza. Neste instante, pode-se dizer que o tratamento de ciclo se encerrou.

Uma vez encerrado o tratamento da série pela decomposição, é possível executar a previsão dos 20 pontos reservados à validação para este exemplo. Esta previsão é feita através da extensão da reta de tendência, somada ao índice sazonal para cada mês e às componentes de ciclo, também extrapoladas para o mês corrente.

Os resultados a cada passo de previsão foram os seguintes:

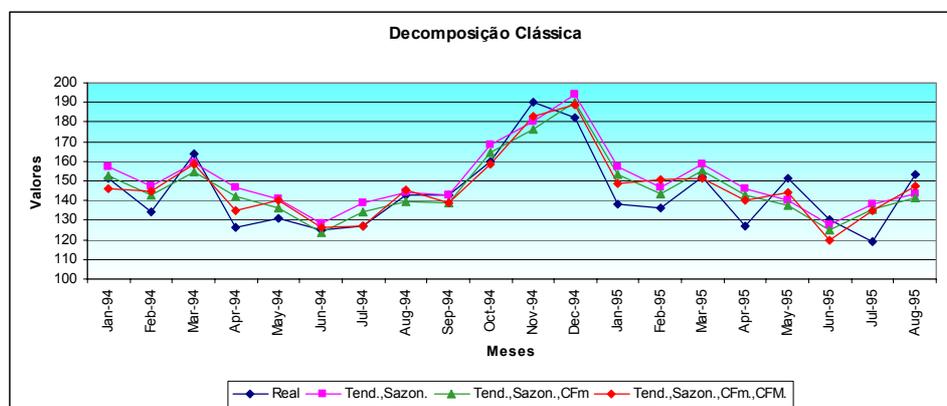
**Tabela 1** – Tabela comparativa de erros

Erros na Previsão		
	EMQP	MAPE
<i>Decomposição Clássica</i>	8,065%	7,159%
<i>Dec. Clás + CFm*</i>	6,930%	6,107%
<i>Dec. Clás + CFm + CFM**</i>	5,758%	5,049%

\* Componente de Frequência Menor

\*\* Componente de Frequência Maior

Nesta tabela, a primeira linha exhibe os erros pela previsão utilizando apenas sazonalidade e tendência; a segunda linha representa o uso adicional da série de frequência mais baixa; e a terceira, o acréscimo da componente interpolada de frequência maior. O gráfico abaixo mostra a comparação entre o que foi previsto pela decomposição e os dados reais.



**Figura 11** – Comparação entre previsto e realizado – Decomposição

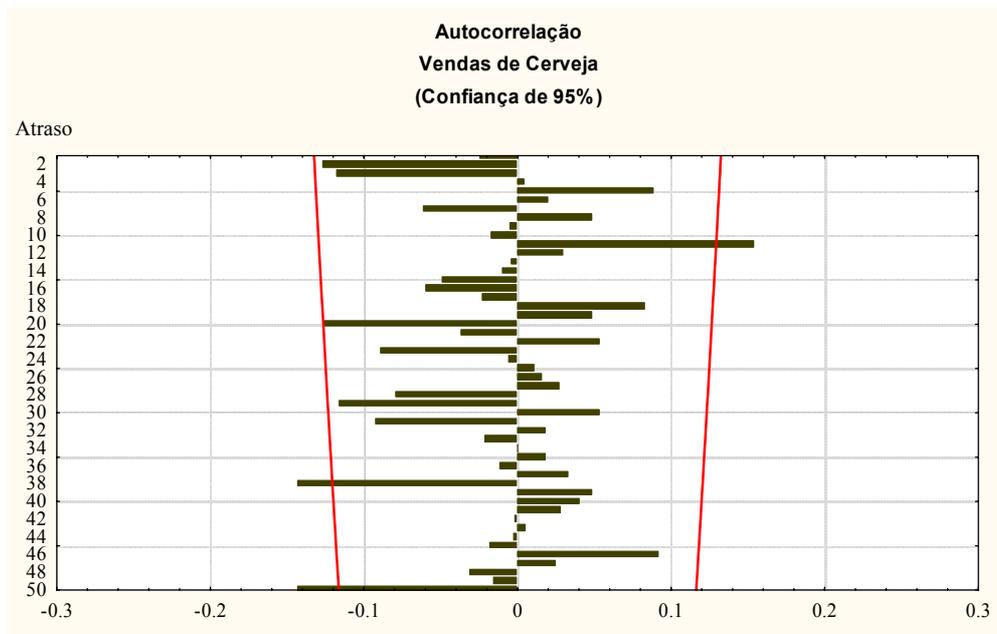
## 4.2 Redes Neurais

Como já foi exposto anteriormente, foram utilizados na previsão através de Redes Neurais os valores da série após a retirada de tendência, sazonalidade e ciclo. Esta decisão foi tomada por deixar para a rede apenas o problema de identificar padrões não-lineares, o que esta vai

fazer melhor através de dados cujos padrões lineares já tenham sido excluídos. Desta forma, a rede vai tratar a não-linearidade, deixando para os métodos clássicos as componentes lineares. Esta propriedade das redes neurais artificiais é salientada por Calôba (1988), Wasserman (1989) e Haykin (1999).

O ponto de partida para utilização da rede foi a preparação dos pontos para entrada na rede. Inicialmente, todos os valores devem estar numa escala entre  $-1$  e  $1$ . Isto se dá porque a função de ativação dos neurônios utilizada foi a tangente hiperbólica, que possui seu intervalo mais significativo de variação entre  $-1$  e  $1$ . Valores neste intervalo contribuem mais para o treinamento.

A seguir, foi feito estudo de autocorrelação da série, utilizando pacote de software estatístico. Através deste gráfico, podem ser identificados os meses que mais estão correlacionados com o valor atual da série. As linhas verticais correspondem a dois desvios padrões da correlação de uma série randômica, ou seja, pontos com correlação superior tem nível de significância menor que 5% para a hipótese de ausência de correlação. Como resultado, apresentou-se o gráfico abaixo:



**Figura 12** – Autocorrelação da Série Decomposta

No gráfico, a abscissa representa a escala das autocorrelações, enquanto a escala vertical identifica, para as barras apresentadas, os meses de atraso com relação à série principal, ou seja, a defasagem da autocorrelação calculada. Três valores mostram-se particularmente interessantes: as vendas defasadas de 11, 38 e 50 meses possuem uma correlação acima do limite de confiança da série randômica. Além disso, existe um efeito combinado possível entre os meses 2 e 3, com correlações quase no limite de significância, e no mesmo sentido. Este efeito combinado poderá trazer alguma contribuição significativa.

Assim, para cada ponto de treinamento, teste e validação, haverá três pontos de entrada, que serão estes valores atrasados da série. É através destes valores que deve ser reconhecido o padrão da série. Logo que forem separados os pares de treinamento e teste, cuja abscissa são os três valores atrasados de vendas e a ordenada é o valor a ser previsto, a rede já está praticamente pronta para ser utilizada.

#### 4.2.1 Seleção de Arquitetura e Treinamento da Rede

Para este caso, em cada ano do período de treinamento e teste, foram selecionados aleatoriamente 4 pontos de teste e oito de treinamento, sem meses pré-determinados para teste ou treinamento. A tentativa é que os conjuntos de teste e treinamento sejam ambos semelhantes ao padrão da série, homogêneos tanto com relação aos meses do ano quanto com os anos utilizados. Isto gerou um conjunto de 110 pontos de treinamento, 56 de teste e 20 de validação.

Uma vez selecionados os conjuntos de treinamento e teste, falta apenas definir a arquitetura da rede a ser utilizada. Isto consiste em determinar parâmetros como a taxa de treinamento e a quantidade de neurônios na camada intermediária. A taxa utilizada foi de 0,001, baixo o suficiente para evitar divergência. Como o método de treinamento é por lote, não deve se ter receio de convergências muito lentas.

A determinação do número de neurônios na camada intermediária não pode ser feita sem alguma experimentação. O uso de poucos neurônios na camada intermediária leva a uma capacidade de generalização maior, mas pode estar simplificando demais o modelo. Já o uso de muitos neurônios pode introduzir uma complexidade que o problema não possui, e fazer com que a rede fique muito restrita, e propensa a *overtraining*.

#### 4.2.2 Treinando e Testando a Rede Neural

Seguindo ao pré-processamento, começaram as experiências com a rede propriamente dita. Foram utilizados as séries atrasadas 2, 3, 11, 38 e 50 meses, seguindo uma ordem lógica de descarte de piores resultados, como segue abaixo. Para cada experiência, é necessário procurar a melhor configuração, ou seja, no caso descrito, testar diversos números de neurônios na camada intermediária.

A primeira aplicação realizada utilizou apenas os dados de vendas 2 e 3 meses antes do valor a ser previsto. Iniciou-se assim pois estas são as entradas que possuem menor correlação. O melhor resultado obtido foi com 2 neurônios na camada intermediária, alcançando MAPE de 5,01% e EMQP de 5,71%.

A seguir, agregou-se a essa configuração a série atrasada de 11 meses. Para esta experiência, com 2 neurônios, o MAPE desceu para 4,92% e o EMQP subiu para 5,73%. Há indícios, então, que a inserção da série com defasagem de 11 meses pode trazer benefícios à previsão.

De modo a esclarecer as dúvidas, a rede foi treinada apenas com a entrada de 11 meses atrasados. O melhor resultado obtido foi com 3 neurônios: MAPE de 4,86% e EMQP de 5,54%.

Assim, descartaram-se as entradas de 2 e 3 meses. Anexaram-se as correlações mais fortes, 38 e 50 meses defasadas, e novamente foram testadas diversas redes, com números distintos

de neurônios. O melhor resultado se deu para 6 neurônios na camada intermediária, alcançando MAPE de 4,50% e EMQP de 5,23%. Este foi o melhor resultado encontrado empregando-se a rede neural. O gráfico a seguir mostra o comportamento dos erros durante o treinamento no caso ótimo, no qual não ocorreu validação cruzada:

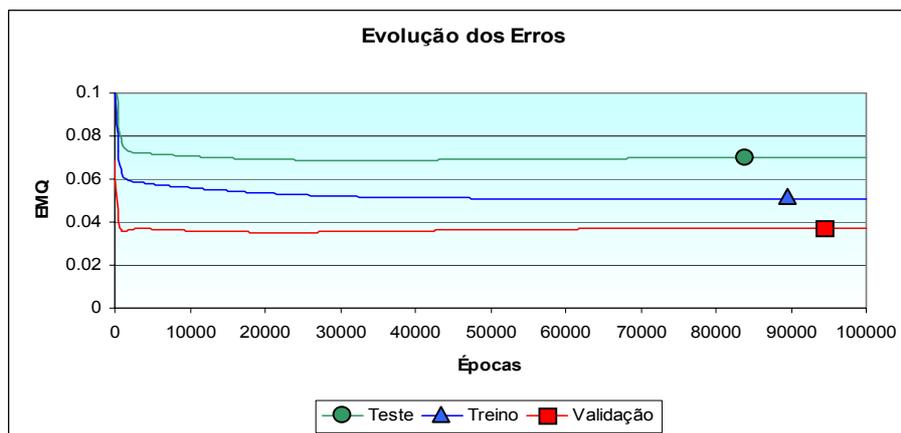


Figura 13 – Evolução dos Erros no Treinamento com 6 Neurônios na camada intermediária

### 4.3 Análise de Resultados

Adicionalmente à decomposição clássica e as Redes Neurais, foi utilizado o método do amortecimento exponencial triplo, também conhecido como amortecimento de Holt-Winters, detalhado, por exemplo, em Morettin & Toloi (1987), que é considerado um bom método de previsão de séries e foi aplicado de forma isolada. A seguir, uma tabela comparativa com os resultados dos métodos:

Tabela 2 – Comparação dos Métodos de Previsão

Erros na Previsão		
	EMQP	MAPE
<i>Decomposição Clássica</i>	8,06%	7,16%
<i>Dec. Clás + CFm</i>	6,93%	6,11%
<i>Dec. Clás + CFm + CFM</i>	5,76%	5,05%
<i>Holt-Winters</i>	6,94%	5,25%
<i>Redes Neurais (11,38,50) n=6, α=0.001</i>	5,23%	4,50%

Como é observável, a Rede Neural complementou de forma positiva a decomposição, chegando a resultados mais precisos que seu uso isolado. Encontrou-se não só menor média de erro, mas também desvio-padrão mais reduzido. Abaixo é exposto um gráfico com a comparação dos pontos reais e os pontos previstos pela rede neural, bem como pelos outros métodos aplicados.

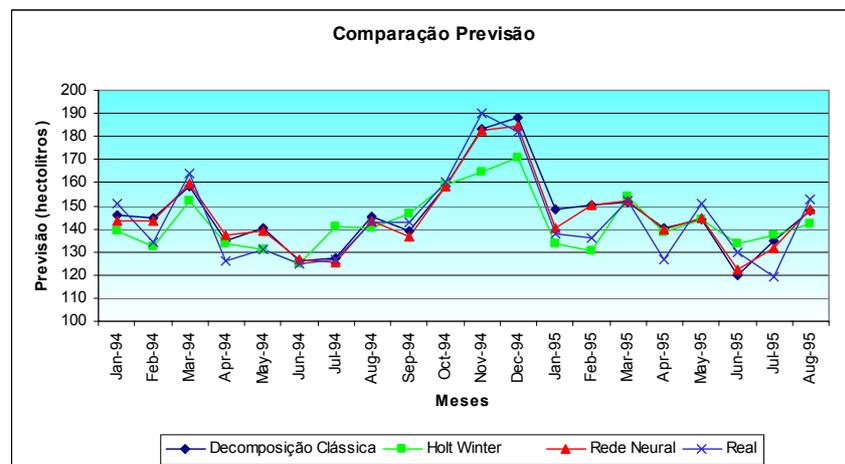


Figura 14 – Comparação de resultados obtidos pelos diversos métodos na previsão

## 5. Conclusões

A série de vendas de cerveja foi tratada por duas diferentes técnicas de previsão. A adição de Redes Neurais ao método de decomposição clássica alcançou resultados superiores. O método aqui empregado (RN e decomposição) alcançou não só a menor média entre os três, como também o menor desvio-padrão da amostra de erros mensais de previsão.

Conclui-se que a Rede Neural conseguiu contribuir de forma significativa para este problema de previsão, reduzindo o erro de uma série já tratada nos domínios do tempo e da frequência em 10%, o que se torna um mérito importante para o método.

As Redes Neurais conseguem eficiência destacada quando tratam de não-linearidades. A postura durante todo este artigo foi utilizar este método aliado a outras técnicas de previsão, no caso, a decomposição clássica. Através desta combinação, é possível executar melhores previsões, pois a rede trata a parte não-linear do problema, enquanto os outros métodos se encarregam da previsão de padrões lineares. Assim, a Rede Neural fica colocada como um instrumento que possibilita uma margem de melhoria na previsão dos métodos clássicos.

Os resultados encontrados foram importantes para a colocação de três proposições importantes:

### A) *Redes Neurais não são a melhor solução sempre*

Muitas vezes, quando uma nova técnica de previsão ou otimização aparece com resultados interessantes, esta se torna rapidamente a “melhor” técnica existente. Isto não é verdade quando se fala de Redes Neurais. Assim como todas as técnicas, existem casos em que seu uso pode ser perfeitamente dispensável.

Os resultados encontrados evidenciam que as séries trabalhadas já possuíam uma ótima aproximação pelos métodos lineares. Assim sendo, o acréscimo da rede não proporciona um ganho tão grande na redução de resíduos, já que o padrão não-linear restante é menor que com a perturbação aleatória da série.

B) *A não-linearidade é um problema que pode ser tratado convenientemente por Redes Neurais*

Se por um lado a rede neural não pode resolver tudo, nos casos em que métodos clássicos de previsão não conseguem resultados bons devido a não linearidades, a rede pode conseguir performances significativas, não só no campo da previsão, como na classificação e otimização.

C) *Associar métodos clássicos lineares e redes neurais parece levar a bons resultados no tratamento de problemas (fracamente) não lineares*

Como se pôde ver durante este trabalho, a combinação de redes neurais e outras técnicas de previsão compôs um método de previsão melhor que o uso separada das técnicas. Este é o ponto chave deste trabalho: a cooperação entre redes neurais e outros métodos para uma previsão mais acurada.

### **Referências Bibliográficas**

- (1) Calôba, L.P. (1988). Introdução à Computação Neuronal. *Anais 9º CBA, UFES – ES*.
- (2) Chatfield, C. (1989). *The Analysis of Time Series – An Introduction*. Chapman and Hall, New York.
- (3) Haykin, S. (1999). *Neural networks, a Comprehensive Foundation*. 2<sup>nd</sup> ed. Prentice Hall, Englewood Cliffs, New Jersey.
- (4) Hecht-Nielsen, R. (1990). *Neurocomputing*. Addison-Wesley Publishing Company, New York.
- (5) Morettin, P.A. & Toloi, C.M. de C. (1987). *Previsão de Séries Temporais*. Atual Editora, São Paulo.
- (6) Wasserman, P.D. (1989). *Neural Computing, Theory and Practice*. Van Nostrand Reinhold, New York.
- (7) Wonnacott, T. & Wonnacott, R. (1990). *Introductory Statistics for Business and Economics*. John Wiley & Sons, New York.