



## 16º Congresso da APDR

Universidade da Madeira, Funchal

Colégio dos Jesuítas, 8 a 10 Julho 2010

### NOVA ABORDAGEM DA METODOLOGIA DE REDES NEURONAIS ARTIFICIAIS PARA A PREVISÃO DE SÉRIES TEMPORAIS DE TURISMO: A DATA COMO ÍNDICE. APLICAÇÃO À REGIÃO DA MADEIRA

João Paulo Teixeira ([joaopt@ipb.pt](mailto:joaopt@ipb.pt)); Paula Odete Fernandes ([pof@ipb.pt](mailto:pof@ipb.pt)).

Instituto Politécnico de Bragança

Campus de Santa Apolónia, Apartado 1038

5301-854 Bragança - Portugal.

Telefone: +351 273 303103

Fax: +351 273 325405

#### RESUMO

Em trabalhos anteriores os autores relataram os seus trabalhos com Redes Neuronais Artificiais (RNA) para realizarem a previsão da série temporal ‘Dormidas Mensais em Hotéis’ das regiões Norte, Centro e Portugal Continental. A metodologia de RNA tem provado fazer previsões com melhor precisão que outras metodologias. Como consequência do aumento da procura turística nos últimos anos, estas séries registaram uma tendência significativamente crescente. Como esta metodologia usa o passado no seu treino tem-se tornado cada vez mais difícil para este modelo prever valores futuros com uma dimensão nunca vista no passado. Os autores propõem neste trabalho uma nova abordagem usando o tempo como parâmetro de entrada em vez de usarem apenas os últimos 12 valores registados no ano anterior. Com este novo parâmetro na entrada pretendem capturar a variação temporal destas séries ao longo dos anos.

Neste trabalho foi usada a série temporal da Região Autónoma da Madeira usando o mês e o ano como índices na entrada da RNA em diferentes combinações de acordo com modelos já experimentados com a série da região Norte. Os modelos confirmaram o benefício da utilização dos índices temporais reduzindo o valor do erro relativo médio e também do coeficiente de correlação.

**Palavras chave:** Previsão, Séries Temporais, Redes Neuronais Artificiais.

#### 1. INTRODUÇÃO

A evolução do turismo na Madeira tem seguido muito de perto a tendência nacional e mesmo mundial, pois tem sido um sector que tem vindo a assumir uma importância significativa em termos de atracção de divisas, registando assim um desempenho



globalmente positivo, mesmo sendo um destino turístico maduro e reconhecido internacionalmente.

A Região Autónoma da Madeira de acordo com dados disponibilizados pelo Instituto Nacional de Estatística Português (INE, 2009), consolidou, nos últimos anos, o seu estatuto de terceiro destino turístico português mais procurado, com a particularidade de ter mantido o seu peso no Produto Interno Bruto (PIB), pois cerca de um quarto do PIB da Madeira resulta do turismo.

De acordo com o Plano Estratégico Nacional do Turismo (2006-2015), para o período que o mesmo contempla, prevê, como um dos principais objectivos para a Madeira, que o número de dormidas de estrangeiros se situe entre os 6,4 e 6,5 milhões de dormidas, ou seja, prevê-se um aumento de 2,7% ao ano do número hóspedes estrangeiros e, as dormidas de turistas nacionais deverão crescer, no mesmo período, a uma taxa anual de 2,1% (PENT, 2007).

Distintos estudos empíricos têm vindo a ser desenvolvidos e publicados na área do turismo onde reflectem que a previsão da procura turística assume um papel relevante no processo de planeamento e tomada de decisões (Witt & Witt, 1995; Wong, 2002; Fernandes, 2005). Estes estudos têm apresentado diferentes modelos de previsão, lineares e não lineares, com diversas características e níveis de complexidade (Yu & Schwartz, 2006; Fernandes *et al.*, 2008; Machado *et al.*, 2010).

De referir que os autores, do presente trabalho, têm vindo a desenvolver estudos sobre a modelação da procura turística, utilizando séries temporais das regiões do Norte, Centro e de Portugal Continental, utilizando diversos modelos, desde a aplicação de modelos de regressão linear simples, modelos ARIMA e modelos utilizando a metodologia das Redes Neurais Artificiais (RNA) (Fernandes 2005; Fernandes *et al.*, 2008; Machado *et al.* 2010). Estes últimos modelos apresentaram qualidades estatísticas e de ajustamento satisfatórias evidenciando serem adequados para a modelação e previsão da procura turística, para as séries das regiões referidas anteriormente. Contudo e porque nos últimos anos, antes de 2009, as séries apresentaram um aumento significativo da procura turística e uma vez que a metodologia das RNA utiliza os valores históricos para prever o futuro, não contemplando no seu treino os valores máximos que se registaram nos últimos anos, tem-se tornado cada vez mais difícil, para este modelo,



produzir valores futuros com uma dimensão nunca vista no passado. Assim e para resolver este problema, os autores propõem uma nova abordagem utilizando a dimensão tempo como parâmetro de entrada, no sentido de captar a variação temporal destas séries ao longo dos anos.

Seguindo esta linha de raciocínio e porque estudos sobre a modelação da procura turística para a Região Autónoma da Madeira ainda é escassa, surgiu o desenvolvimento deste trabalho de investigação. Pois, será de todo vantajoso contribuir para o enriquecimento de estudos nesta área, desenvolvendo mecanismos que permitam antever a evolução da procura turística, no sentido de criar condições favoráveis aos turistas que se deslocam à Região em análise. Acresce o facto de poder proporcionar aos principais agentes da oferta hoteleira, para que possam efectuar uma gestão por objectivos, um instrumento que lhes permita auxiliar nas previsões do número de dormidas, com vista a obter erros mínimos entre as dormidas previstas e efectivas.

Para tal, o principal objectivo do presente estudo centrou-se numa análise ao comportamento da série temporal de turismo: “Dormidas Mensais nos Estabelecimentos Hoteleiros”, registadas no período de Janeiro de 1987 a Dezembro de 2009, na Região Autónoma da Madeira. Para a modelação e previsão dos valores desta série utilizou-se a metodologia das Redes Neuronais Artificiais. Tal como já foi referido os autores propõem-se trabalhar a série sob uma nova abordagem que assenta na utilização da variável tempo, mês e ano, como índices na entrada da RNA e como saída da rede o valor previsto para esse mesmo mês. Em trabalhos anteriores, os autores utilizavam como entrada da RNA as 12 observações do ano anterior e a saída consistia no valor para o mês seguinte.

O artigo está estruturado da seguinte forma: a secção dois apresenta e descreve o comportamento da série em estudo, na secção três é abordada a metodologia utilizada, na secção seguinte são detalhados os modelos experimentados e as variáveis usadas. Os resultados dos modelos são apresentados e discutidos na secção cinco. As conclusões são elaboradas numa secção final.

## **2. APRESENTAÇÃO E ANÁLISE DO COMPORTAMENTO DA SÉRIE**

A série “Dormidas Mensais de Turistas nos Estabelecimentos Hoteleiros”, é considerada como significativa da actividade turística devido às suas especificidades



uma vez que contempla visitantes que usufruíram das capacidades turísticas. Neste trabalho foi considerada a série da região Autónoma da Madeira.

Os dados observados tem uma periodicidade mensal e reportam-se ao horizonte temporal compreendido entre Janeiro de 1987 e Dezembro de 2009, constituindo assim um total de 276 observações mensais ao longo de 23 anos (figura 1). Os valores da série foram facultados pelo Instituto Nacional de Estatísticas de Portugal (INE, 1987-2009). De referir que os valores para o ano de 2009, são dados preliminares apresentados pelo Instituto Nacional de Estatística.

Pela análise à figura 1, a série temporal regista uma sazonalidade anual, sendo mais evidente a partir de Janeiro de 1999. Atinge o seu pico máximo, tipicamente em Agosto, seguindo-se de um outro pico em Abril. O seu pico mínimo é atingido em Dezembro ou Janeiro e ainda um mínimo local em Junho. A origem do pico máximo em Abril pode estar relacionada com a tradicional Festa da Flor, que se prende com a celebração da Primavera e o exuberante florescer típico da época, em que a cidade do Funchal é o palco deste espectáculo que tradicionalmente atrai muitos turistas à região. Já em Agosto, além dos turistas procurarem a ilha da Madeira como produto turístico de sol e mar ou turismo da natureza, o evento ‘Rali Vinho Madeira’ vem contribuir para o aumento das dormidas uma vez que o mesmo constitui uma das mais importantes provas do Campeonato Europeu de automobilismo e a maior manifestação automobilística da Região. Embora em Setembro os valores comecem a diminuir, apresentam ainda valores elevados quando comparados com os meses das estações do Outono e Inverno. Mais uma vez a Ilha da Madeira promove dois eventos sendo eles a ‘Festa do Vinho da Madeira’ e o ‘Festival Colombo’, sendo que ambos os acontecimentos captam turistas, essencialmente de mercados emissores externos. Também o mês de Junho começou a registar valores mais elevados ao longo dos últimos anos o que poderá ser reflexo do evento ‘Festival do Atlântico’, convidando turistas para a região, uma vez que o evento integra o Festival de Música da Madeira, o Concurso Internacional de Fogo-de-artifício (este concurso apresenta, em todos os sábados do mês de Junho, espectáculos de fogo-de-artifício conjugados harmoniosamente com a música) e a actuação de Bandas Filarmónicas.



## 16º Congresso da APDR

Universidade da Madeira, Funchal

Colégio dos Jesuítas, 8 a 10 Julho 2010

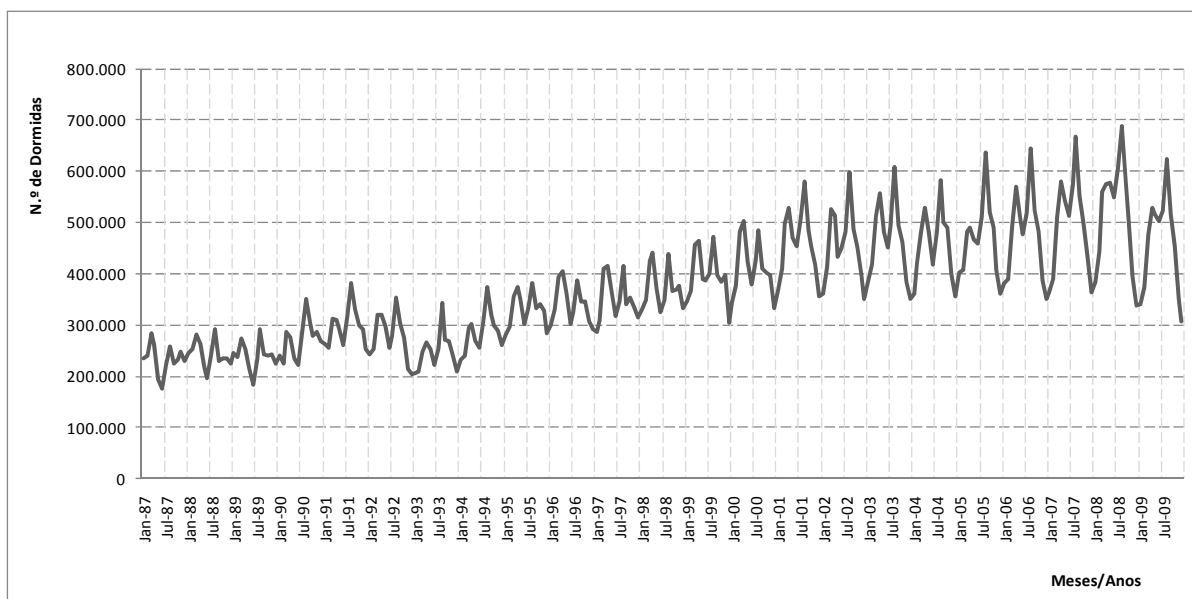


Figura 1: Dormidas dos turistas nas Unidades de Alojamento da Região Madeira, no intervalo [Jan-87:Dez-09] (INE, 1987-2009).

Relativamente ao comportamento da série, tendo por base a figura anterior, a mesma apresenta um crescimento acentuado a partir de 1994 até 2003, com um decréscimo não muito significativo em 2004, mas com valores ainda superiores aos registados nos anos antecedentes a 2000. Apresenta, ainda, um crescimento assinalável nos anos de 2005 a 2008. Este crescimento pode ser resultado de investimentos realizados na promoção e imagem da região a uma escala nacional e internacional, associada aos eventos que a Região Autónoma da Madeira tem vindo a realizar, alguns deles já foram referidos no parágrafo anterior. Os decréscimos registados no ano de 2004 podem ser reflexo das atenções estarem voltadas para o Continente, devido à realização do evento desportivo EURO2004, e no ano de 2009, poderá ser resultado da desfavorável conjuntura económica bem como do efeito negativo na propensão para viajar, associado ao risco de propagação do vírus H1N1.

### 3. METODOLOGIA

A previsão de séries temporais de turismo com RNA tem sido realizada habitualmente com um modelo que aqui denominamos de modelo Clássico (Fernandes 2005; Fernandes e Teixeira 2007) que consiste em usar na entrada um conjunto de valores da série de meses anteriores como informação para prever o mês seguinte. Em estudos



## 16º Congresso da APDR

Universidade da Madeira, Funchal

Colégio dos Jesuítas, 8 a 10 Julho 2010

efectuados por Fernandes e Teixeira, verificou-se experimentalmente que melhores previsões são realizadas quando na entrada estão os valores dos 12 meses anteriores. Estes consistem nos valores correspondentes ao ano anterior e completam um ciclo da sazonalidade destas séries, pelo que são um bom conjunto de factores para prever o valor da série do mês seguinte.

Contudo as séries de dormidas nos Hotéis de Portugal (Machado *et al.* 2010), região Norte e Região Centro de Portugal (Fernandes, *et al.*, 2008) bem como a série da região da Madeira, aqui trabalhada, denotam um crescimento acentuado nos últimos anos como se pode confirmar pelas Figuras 2, 3 e 4, para as regiões Norte, Centro e Portugal Continental, respectivamente.

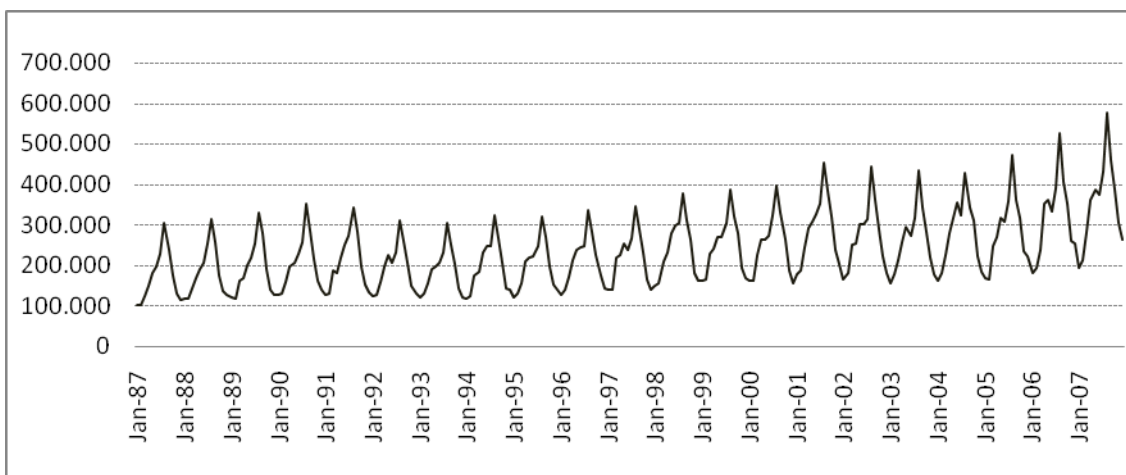


Figura 2: Dormidas dos turistas nas Unidades de Alojamento da Região Norte, no intervalo [Jan-87:Dez-07] (INE, 1987-2007).

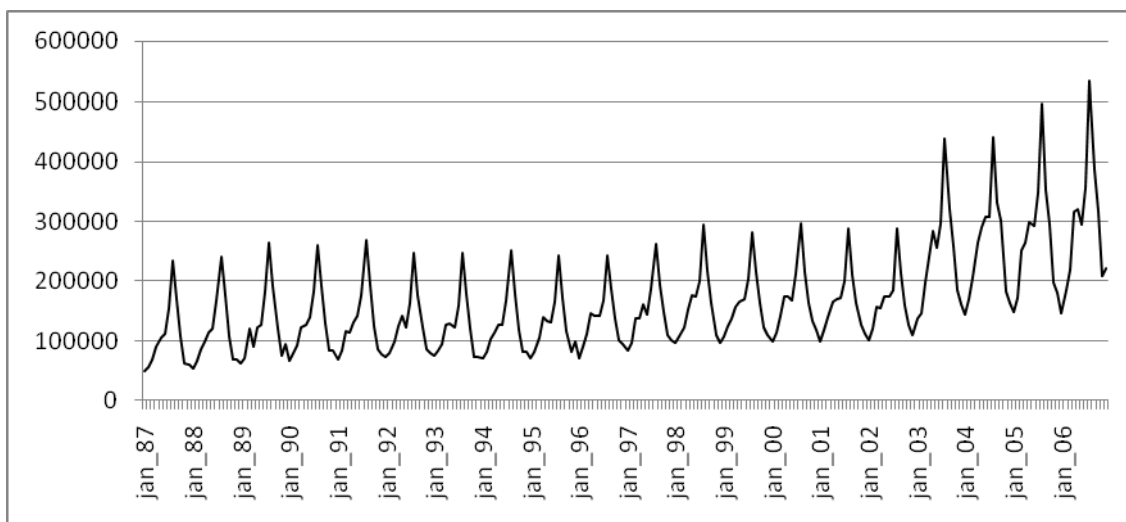


Figura 3: Dormidas dos turistas nas Unidades da Região Centro, no intervalo [Jan-87:Dez-06] (INE, 1987-2006).

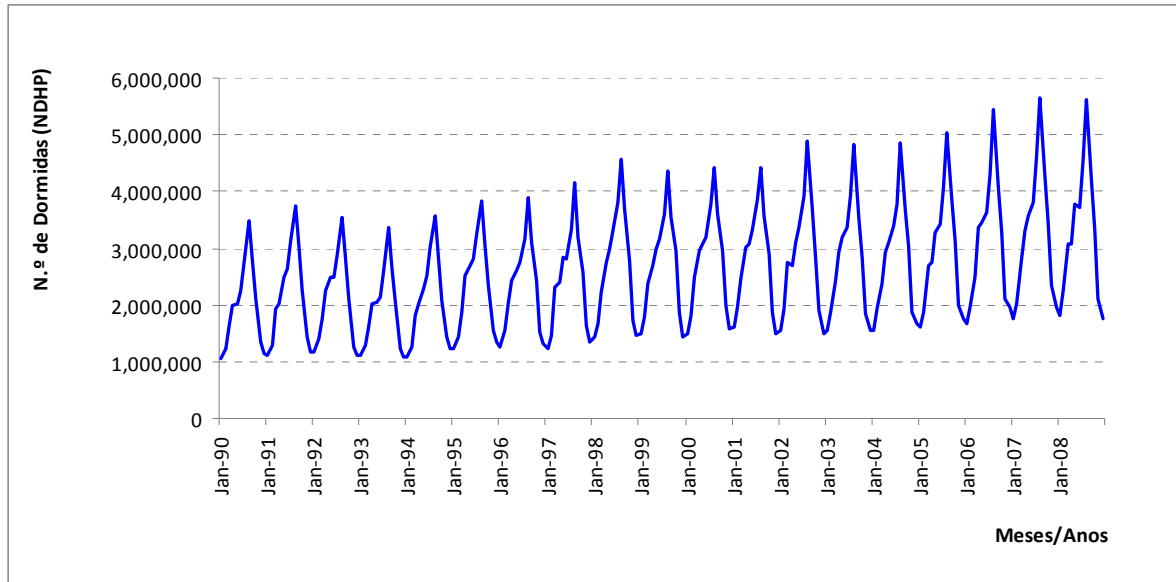


Figura 4: Dormidas dos turistas nas Unidades de Alojamento em Portugal Continental, no intervalo [Jan-87:Dez-08] (INE, 1987-2008).

Este crescimento para valores nunca ocorridos no passado acaba por ser uma dificuldade para o modelo de RNA, tornando-se difícil para um modelo que é estatístico, e que é treinado com valores do passado, supostamente semelhantes às novas situações a prever, conseguir fazer previsões com magnitudes que nunca foram atingidas nas situações de treino. Contudo tem-se verificado que esse crescimento tem uma tendência de crescimento aproximadamente linear. Então procura-se neste modelo aqui proposto usar uma nova variável na entrada de rede que faculte uma dimensão temporal ao modelo.

Assim, foram já testadas com sucesso em (Fernandes e Teixeira 2009) algumas combinações de entradas contendo esta dimensão usando a série da região Norte de Portugal. Nomeadamente, foram experimentadas as seguintes situações:

- Modelo A – usando na entrada apenas uma referência ao índice do mês ordenado desde o início da série. A rede contém apenas um nó na entrada.
- Modelo B – usando na entrada apenas uma referência ao ano e outra ao mês. A entrada da RNA consiste em apenas dois nós.



- Modelo C – consiste numa combinação do modelo A como modelo clássico. Assim a entrada da rede contém uma referência ao mês e os valores dos 12 meses anteriores, num total de 13 entradas.
- Modelo D – consiste também numa combinação do modelo clássico mas agora com o modelo B. Assim a entrada da rede contém os valores dos 12 meses anteriores, uma referência ao ano e outra ao mês do ano, num total de 14 entradas.

A tabela 1 apresenta um resumo dos resultados obtidos para os diferentes modelos experimentados. Pode verificar-se que os resultados com os modelos B, C e D foram melhorados em relação ao modelo clássico. Adicionalmente verificou-se que o modelo A não consegue fazer previsões a um nível de qualidade aceitável.

Tabela 1 – Erro Relativo Médio (ERM) para os modelos Clássico, A, B, C e D aplicado à série da Região Norte (Fernandes e Teixeira, 2009).

	<i>Modelo Clássico</i>	<i>Modelo A</i>	<i>Modelo B</i>	<i>Modelo C</i>	<i>Modelo D</i>
<i>ERM_Validação (%)</i>	5.08	15.17	4.62	5.97	5.56
<i>ERM_Teste (%)</i>	6.36	35.38	5.98	6.14	5.79

Neste trabalho procurou-se confirmar a utilidade dos modelos denominados de B, C e D com a série da Região Autónoma da Madeira. Este série, impõe um nível de dificuldade adicional, à partida, devido à alteração da sua forma ao longo dos anos, e com uma acentuada tendência de crescimento mas com poucas semelhanças com um crescimento linear.

## 4. MODELOS DE RNA

Nesta secção apresentam-se com detalhe os modelos comparados.

Todos os modelo são baseados em RNA, com pequenas alterações na sua arquitectura devido ao diferente número de nós na camada de entrada e também da informação fornecida à rede. Então começa-se por fazer uma breve descrição genérica do modelo de RNA, seguindo-se a descrição dos modelos usados iniciando-se pelo modelo denominado clássico usado como referência seguindo-se os modelos B, C e D. O modelo A usado em (Fernandes e Teixeira 2009) foi abandonado por se ter verificado a sua incapacidade para modelar adequadamente estas séries.

### 4.1 REDES NEURONAIAS ARTIFICIAIS





As RNA são usadas nas mais variadas áreas do conhecimento para fazer previsões com base em conhecimento do passado transmitido às suas sinapses durante um processo denominado de treino. Estas adquirem a capacidade de generalização para novas situações desde que tenham sido usadas situações nos dados de treino que lhe permitam caracterizar adequadamente o problema em causa (Haykin, 1999).

Neste trabalhos são usadas RNA com uma arquitectura *feed-forward* em três camadas e treinadas num processo supervisionado com algoritmos de treinos *back-propagation* (Rumelhard e McClelland 1986). A arquitectura é denominada *feed-forward* porque a saída é calculada a partir dos nós da camada de trás para os nós das camadas da frente, como se exemplifica na figura 5. Genericamente pode-se representar essa rede numa estrutura mais compacta como na figura 6, em que se representa uma rede com 2 nós na camada de entrada, 4 nós na camada escondida e 3 nós na camada de saída. Todos os nós de uma camada são aqui representados de forma compacta nessa camada.

Neste exemplo são usadas as funções de activação tangente hiperbólica na camada escondida e função linear na camada de saída.

A saída de cada nó de cada camada é apresentado na figura 6 como  $a_1$  e  $a_2$ . Em que  $a_1$  é dado pela tangente hiperbólica do somatório do produto da matriz de pesos  $IW$  da camada escondida pelo vector de entrada  $p$  mais os desvios  $b_1$  de cada nó dessa camada. A saída  $a_2$  é determinada pelo somatório do produto da matriz de pesos  $LW$  da camada de saída pelo vector  $a_1$  resultante da camada anterior, mais o vector de desvios  $b_2$  dos nós da camada de saída.

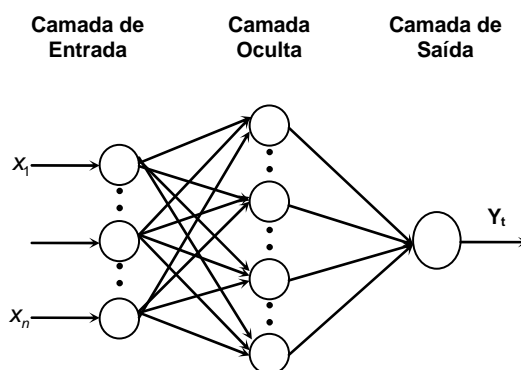


Figura 5: Exemplo de rede neuronal *feed-forward* multicamada.

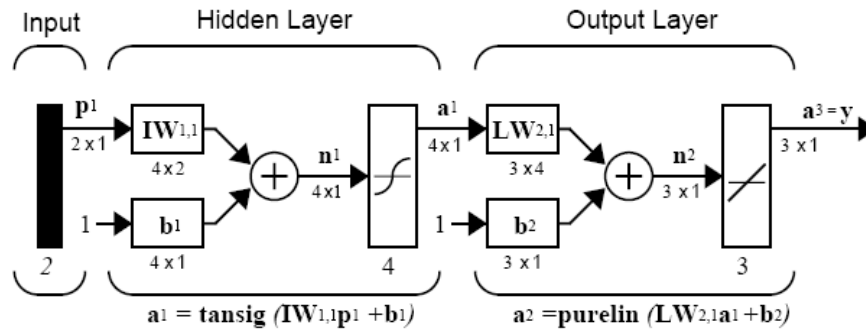


Figura 6: Representação compacta de uma rede *feed-forward* multicamada. Fonte: Demuth, Howard *et al.* (2008:5-11).

As matrizes de pesos  $IW$  e  $LW$  bem como os vectores de desvios  $b_1$  e  $b_2$ , são ajustados durante o processo de treino, com um algoritmo de back-propagation ou retro-propagação. Assim denominado porque o erro entre a saída pretendida e a saída da rede é propagado das camadas mais à frente para as camadas de trás. Este algoritmo, apresentado em 1986 por Rumelhard e McClelland, vieram permitir o treino de redes multi-camada para resolver problemas não lineares. Trata-se de um treino supervisionado porque durante o treino é fornecido um conjunto de dados de entrada e os respectivos valores pretendidos para a saída. Desta forma o treino consiste em ajustar os pesos e os desvios que permitem minimizar o erro entre os vectores de saída da rede e os vectores de valores alvo pretendidos para todas as situações usadas no treino. Este algoritmo foi objecto de alterações para melhorar o seu desempenho em termos de rapidez para chegar à solução e em termos de capacidade de obtenção da melhor solução. Os diversos algoritmos desenvolvidos são adequados a situações e problemas diferentes no que concerne à complexidade do problema e à dimensão da rede usada. Para o problema em estudo o algoritmo Levenberg Maquardt (Hagan 1994) demonstrou ser o mais adequado e por isso foi usado em todos os modelos descritos nas secções seguintes.

O processo de treino é interrompido usando um processo de validação cruzada. Este processo consiste na utilização de um conjunto denominado de validação e que é usado durante as diversas iterações para avaliar o erro neste conjunto. O treino é interrompido quando durante um número  $N$  sucessivo de iterações o erro neste conjunto não diminui. Este processo impede que a rede neuronal se ajuste em demasia ao conjunto de



situações dado no conjunto de treino perdendo a capacidade de generalização. Neste trabalho foi usado um valor de  $N=6$  para interromper o treino.

Foi ainda usado um conjunto denominado de conjunto de teste que não é usado no processo de treino e que serve para no final avaliar a performance do modelo com dados novos, não ‘vistos’ no processo de construção do modelo.

Os conjuntos de treino validação e teste para cada modelo são apresentados na tabela II.

Tabela 2 – Divisão da série nos conjuntos de treino, validação e teste para os modelos.

	Conjunto de Treino		Conjunto de Validação		Conjunto de Teste	
	Início	Fim	Início	Fim	Início	Fim
Modelos Clássico, C e D	Jan. 88	Dez. 06	Jan. 07	Dez. 07	Jan. 08	Dez. 08
Modelo B	Jan. 87	Dez. 06	Jan. 07	Dez. 07	Jan. 08	Dez. 08

Em todos os modelos o ano de 2007 foi usado para validação e o ano de 2008 para teste. O ano de 2009, não foi usado em nenhum dos conjunto de treino, validação ou teste, pela razão de os dados fornecidos pelo INE para este ano serem valores previstos pelo próprio INE. Assim foi considerado inadequado usar valores previstos por um modelo desconhecido deste trabalho, para comparar com os valores previstos pelos modelos aqui desenvolvidos. Contudo são apresentados os valores da série do INE e os valores previstos para cada modelo. O conjunto de treino é o mesmo para os modelos clássico, C e D uma vez que todos eles usam na entrada os valores dos 12 meses anteriores. Também por esta razão este conjunto só pode começar em Janeiro de 1988, perfazendo um total 228 meses. Já para o modelo B, o conjunto de treino começa em Janeiro de 1987, perfazendo um total de 240 meses.

A arquitectura da rede neuronal para todos os modelos aqui desenvolvidos é semelhante, variando apenas no número de nós da camada de entrada e na camada escondida. A figura 7, apresenta uma arquitectura genérica das redes dos modelos usados.

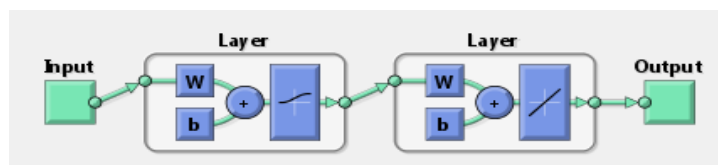


Figura 7: Representação genérica das redes neuronais usadas nos modelos.

Os nós de entrada e da camada escondida são diferentes em cada modelo e referidos nas secções seguintes. A camada de saída contém sempre um nó apenas. Esse nó



corresponde ao valor da série para o mês em causa. A função de activação na camada escondida foi sempre a função logística ou logarítmica hiperbólica (função *logsig* em Matlab), representada pela figura 8. Na camada de saída foi sempre usada a função de activação linear (função *purelin* em Matlab). Todas as redes foram treinadas com o algoritmo de treino Levenberg Marquardt (Hagan, 1994).

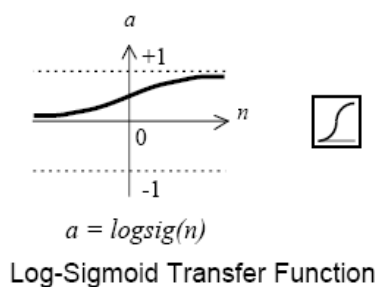


Figura 8: Função de activação *logsig*.

Os modelos foram desenvolvidos em ambiente Matlab recorrendo à toolbox de redes neuronais (Demuth, *et al.* 2008). De referir que os valores das séries nas entradas e nas saídas são normalizados entre -1 e 1, pelas funções das redes neuronais do Matlab, num processo transparente para o utilizador a partir da versão *release2008*.

Cada modelo foi treinado várias vezes e seleccionado o caso com menor erro no conjunto de validação. O erro medido foi o erro relativo médio apresentado na equação da secção seguinte.

#### 4.2 MODELO CLÁSSICO

Este modelo consiste numa RNA com 12 nós na camada de entrada e 6 nós na camada de saída. Cada nó da camada de entrada recebe directamente o valor da série de um dos 12 meses anteriores. Esta mesma arquitectura foi também usada no anterior trabalho (Fernandes e Teixeira, 2009) para a região Norte. Neste modelo a sazonalidade é supostamente dada pelos 12 meses anteriores presentes na entrada da rede e a tendência deverá ser inferida da magnitude dos seus valores.

#### 4.3 MODELO B

A entrada contém apenas dois nós para o ano e para o mês do ano. O ano é numerado por ordem cronológica de 1 para 1987 até 23 para o ano de 2009. O mês é numerado de 1 a 12 para os meses de Janeiro a Dezembro. A saída será o valor da série para o



ano/mês correspondente. Espera-se que a sazonalidade seja dada pelo mês e a tendência pelo ano. A estrutura deve capturar os valores da saída, que nunca são usados na entrada e são apenas vistos na saída no processo de treino. Isto requer uma estrutura mais complexa da rede, pelo que se chegou experimentalmente ao número de 20 nós na camada escondida.

#### 4.4 *MODELO C*

Este modelo consiste na utilização conjunta do índice do mês com as doze entradas do modelo clássico. O índice do mês é contado a partir de 1 para o primeiro mês (Janeiro de 1987) e vai até  $23 \times 12 = 276$  (Dezembro de 2009). É esperado que os valores dos 12 meses anteriores dêem a sazonalidade e o índice do mês a tendência. Neste caso o número de nós óptimo obtido experimentalmente foi de 7, enquanto para a série da região Norte havia sido de 16.

#### 4.5 *MODELO D*

Este modelo consiste numa combinação do modelo B com o modelo clássico. A entrada tem agora 14 nós. Dois para o ano e mês determinados da mesma forma que no modelo B e 12 para os valores dos 12 meses anteriores como no caso do modelo clássico. Para esta série foram usados 8 nós na camada escondida enquanto para a série da região Norte tinham sido usados 16 nós.

### 5. ANÁLISE E DISCUSSÃO DOS RESULTADOS

Apresentam-se nesta secção os resultados obtidos com cada modelo e a sua análise

A medida do erro usada foi o erro relativo médio (ERM) medido pela equação seguinte:

$$ERM = \frac{1}{N} \sum_{i=1}^N \left| \frac{T_i - P_i}{T_i} \right|$$

Em que N é o número de meses previstos, i o mês,  $T_i$  o valor alvo da série original para o mês i e  $P_i$  o valor previsto para o mês i.

Uma vez treinado cada um dos modelos foi realizada uma simulação para o conjunto de meses da série até Dezembro de 2009.

Para o modelo clássico apresenta-se na figura 9 os valores previstos pelo modelo bem como os da série original. Os três últimos períodos correspondem ao conjunto de validação (ano de 2007), ao conjunto de teste (ano de 2008) e ao ano de 2009. Refira-se



que em relação a este último ano não foram efectuadas medidas comparativas com os valores da série original, já que estes são também valores previstos pelo INE. Ainda assim, registam-se no gráfico da figura e pode verificar-se uma previsão de decréscimo nesse ano quer pelo INE quer com este modelo. Os valores representados para o ano de 1987 são os originais já que para este ano não foi realizada previsão por falta de dados de entrada anteriores. O mesmo acontece para as figuras relativas aos modelos C e D.

Os valores de ERM no conjunto de validação (usado no treino de forma indirecta) e no conjunto de teste (não usado no treino) situam-se num valor significativamente baixo de 2.92% e de 4,38%, respectivamente, denotando uma qualidade elevada de previsão deste modelo para esta série quer em valor absoluto quer quando comparado com o erro de previsão deste mesmo modelo para a série da região Norte em que se situou nos 6,4% (ver tabela 1).

Na figura 10, registam-se as rectas de regressão linear e os respectivos coeficientes de correlação,  $R$ , para cada um dos conjuntos de treino, validação e teste, bem como para o total da série. A figura, mostra que o modelo fez uma previsão com um  $R=0.965$  para o conjunto de teste e um  $R=0.987$  para o total dos meses da série.

A qualidade da previsão do modelo B é representada nas figuras 11 e 12, em que se apresentam as curvas com os valores da previsão e originais, e as rectas de regressão com os respectivos coeficientes de correlação.

Refira-se uma melhoria na qualidade da previsão realizada por este modelo com um ERM de 2.84% e 4.25% para os conjuntos de validação e teste, respectivamente. Também o coeficiente de correlação foi melhor com este modelo do que com o modelo clássico, com  $R$  de 0.976 e 0.989 para os conjunto de teste e para o total dos meses, respectivamente.

As figuras 13 e 14 mostram as curvas com os valores previstos e as rectas de regressão com os coeficientes de correlação para o modelo C.

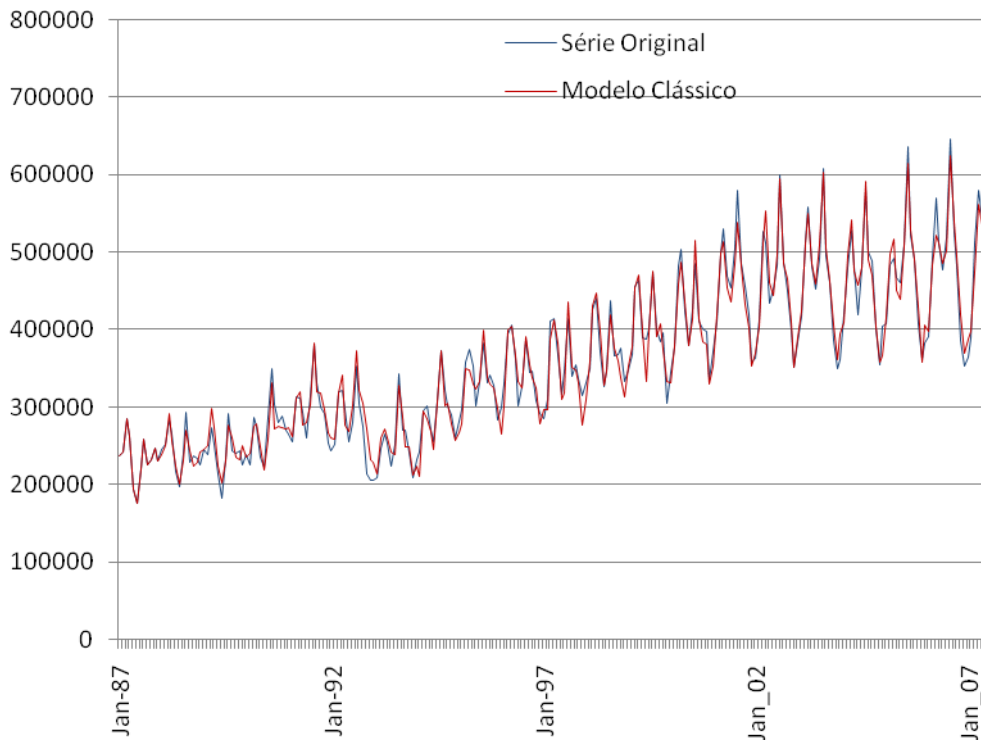


Figura 9: Valores previstos pelo modelo clássico.

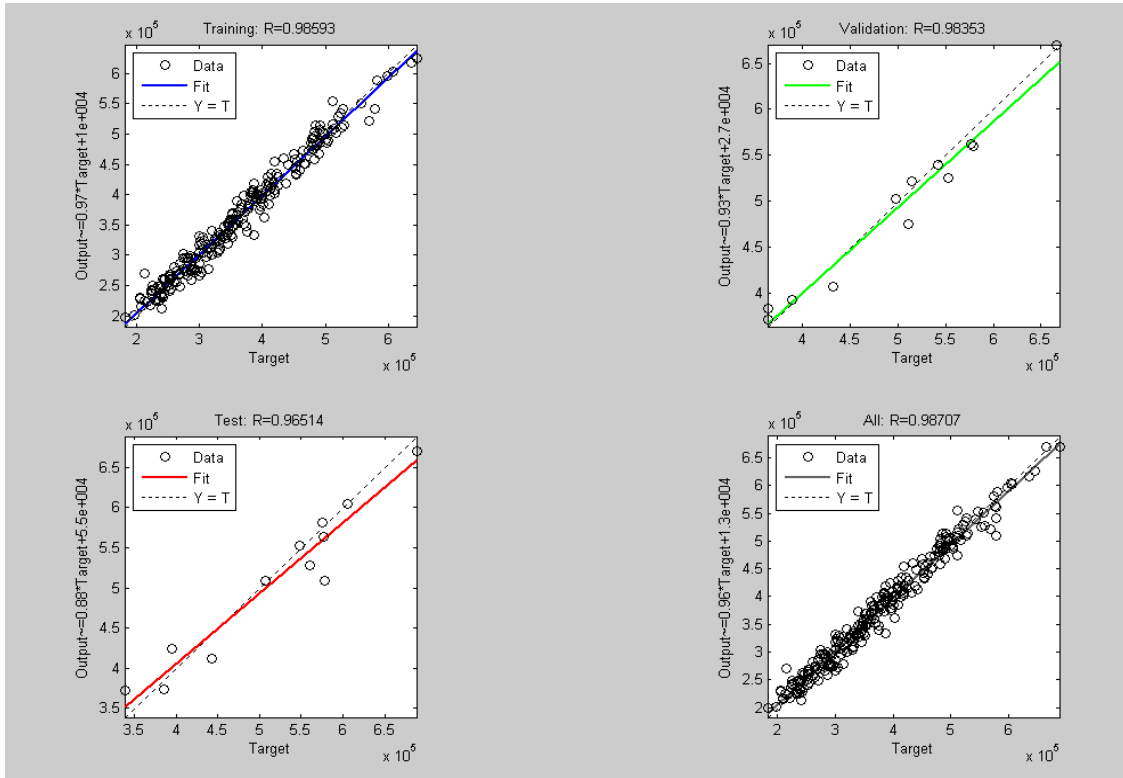


Figura 10: Rectas de regressão e coeficiente de correlação, R, para cada conjunto com o modelo clássico.



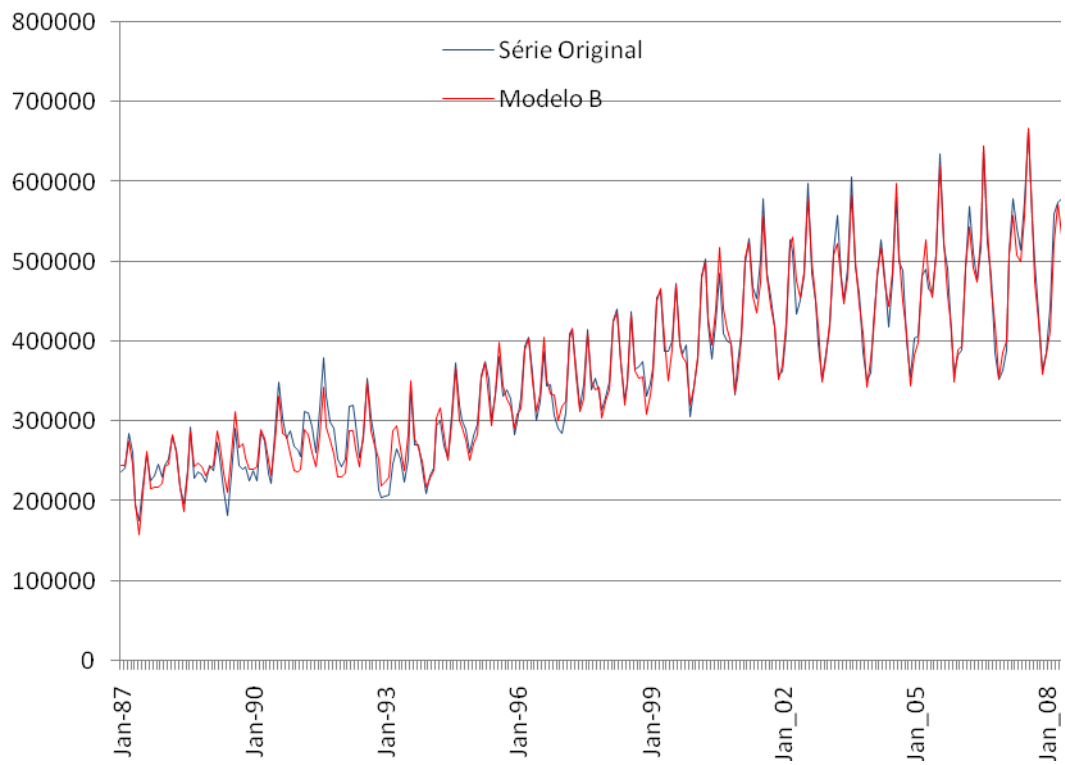


Figura 11: Valores previstos pelo modelo B.



# 16º Congresso da APDR

Universidade da Madeira, Funchal

Colégio dos Jesuítas, 8 a 10 Julho 2010

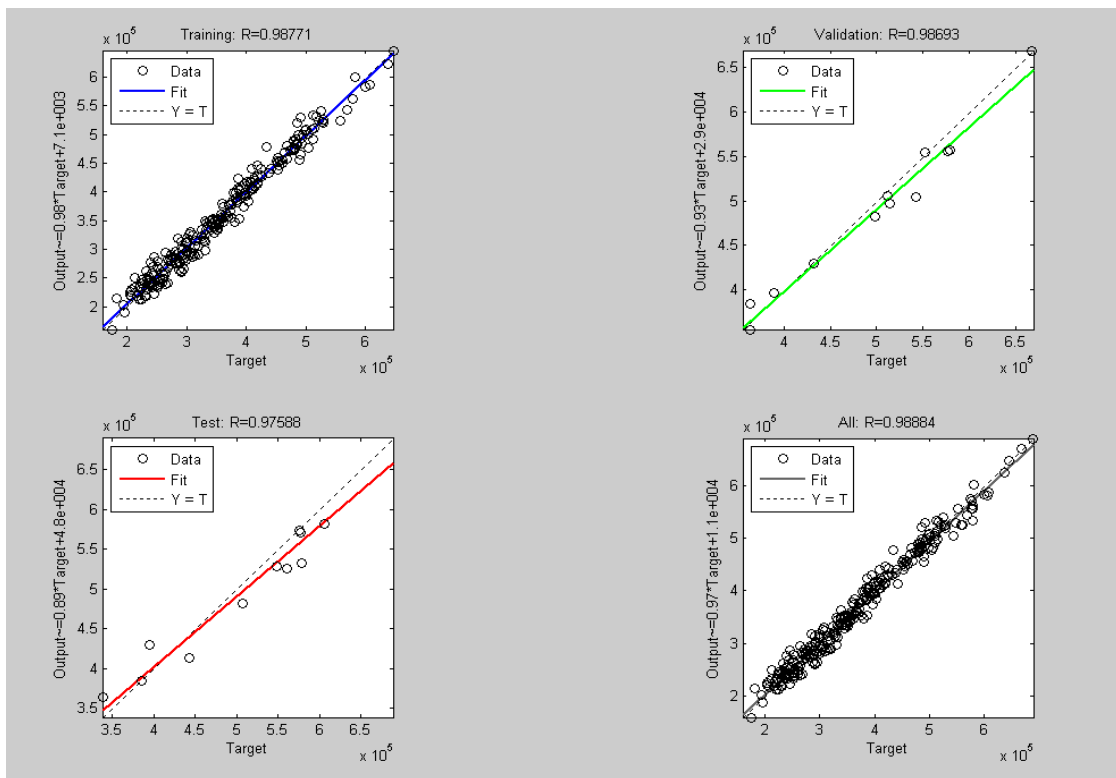


Figura 12: Rectas de regressão e coeficiente de correlação, R, para cada conjunto com o modelo B.



# 16º Congresso da APDR

Universidade da Madeira, Funchal

Colégio dos Jesuítas, 8 a 10 Julho 2010

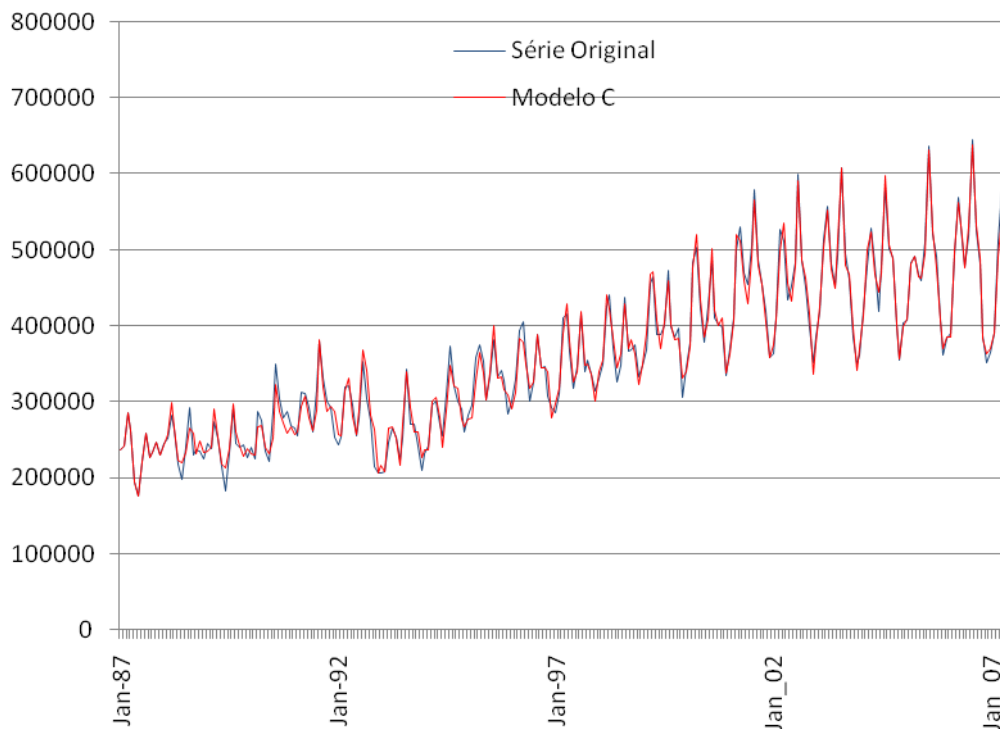


Figura 13: Valores previstos pelo modelo C.

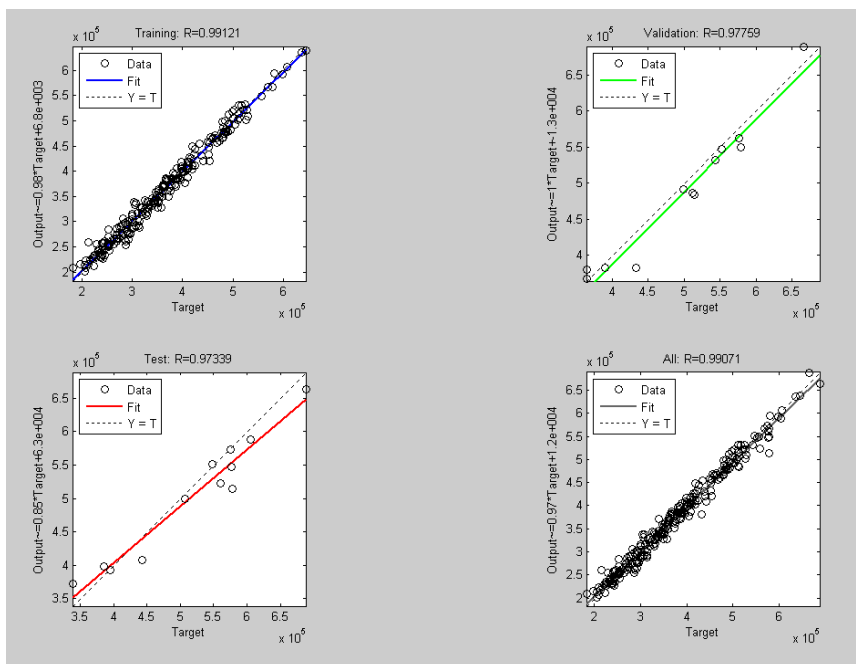


Figura 14: Rectas de regressão e coeficiente de correlação, R, para cada conjunto com o modelo C.



# 16º Congresso da APDR

Universidade da Madeira, Funchal

Colégio dos Jesuítas, 8 a 10 Julho 2010

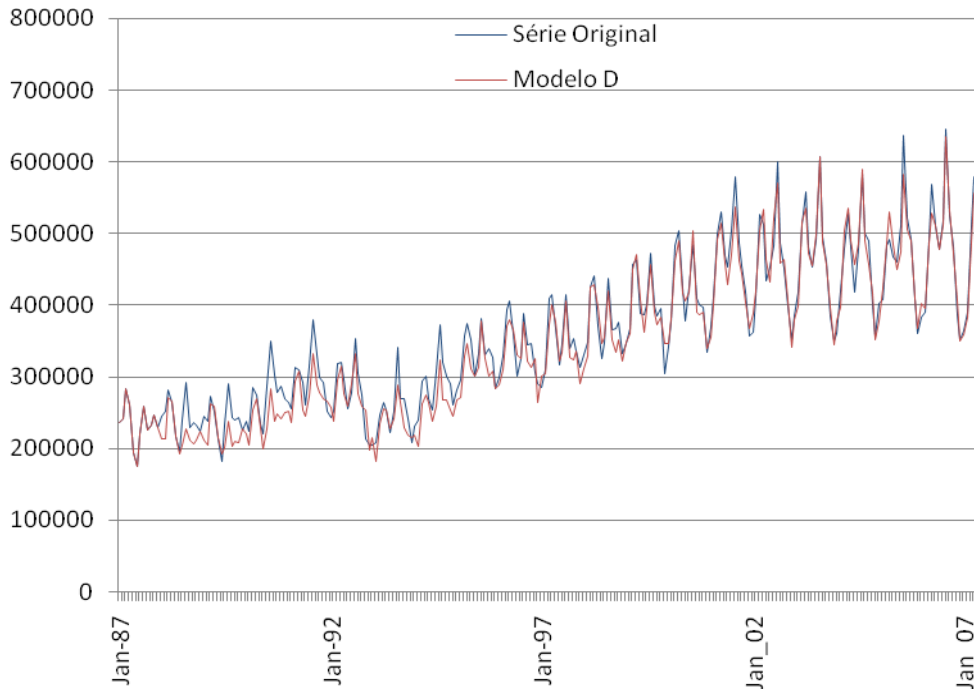


Figura 15: Valores previstos pelo modelo D.

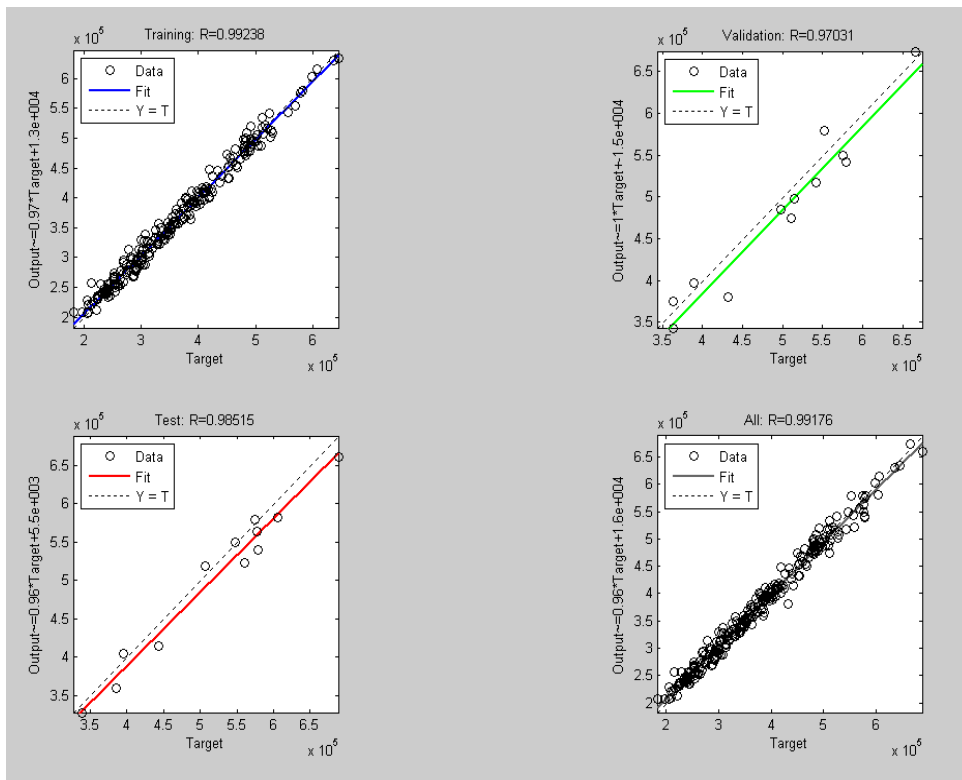


Figura 16: Rectas de regressão e coeficiente de correlação, R, para cada conjunto com o modelo D.



O ERM de 3.52% e 4.46% para os conjuntos de validação e teste, respectivamente, não mostra uma melhoria em relação ao modelo clássico, mas também não se afasta muito. No entanto o coeficiente de correlação apresenta melhores valores,  $R=0.973$  e  $R=0.991$ , para os conjuntos de teste e total dos meses, respectivamente.

As figuras 15 e 16 apresentam os valores previstos e as rectas de regressão com os coeficientes de correlação para o modelo D.

O modelo D apresenta um ERM de 3.07% e 3.83% para os conjuntos de validação e teste, respectivamente. Isto denota que foi o modelo que mais baixo erro apresentou no conjunto de teste. Também o coeficiente de correlação 0.985 e 0.992 para os conjuntos de teste e total dos meses representam uma melhoria em relação a qualquer dos restantes modelos.

A tabela 3 apresenta o resumo dos ERM nos conjuntos de validação e teste para os modelos experimentados.

Tabela 3 – ERM para os conjuntos de Validação e Teste com os modelos Clássico, B, C e D.

	<i>Modelo Clássico</i>	<i>Modelo B</i>	<i>Modelo C</i>	<i>Modelo D</i>
<i>ERM_Validação(%)</i>	2.92	2.84	3.52	3.07
<i>ERM_Teste (%)</i>	4.38	<b>4.25</b>	4.46	<b>3.83</b>

A tabela 4 resume os valores do coeficiente de correlação, R, para os conjuntos de teste e total dos meses. Esta medida de correlação é mais indicadora da qualidade de seguimento da forma da curva do que da qualidade da aproximação à magnitude dos valores. Um bom indicador desta aproximação da magnitude será o ERM.

Tabela 4 – Coeficiente de correlação R para o conjunto de Teste e para o total dos dados da série com os modelos Clássico, B, C e D.

	<i>Modelo Clássico</i>	<i>Modelo B</i>	<i>Modelo C</i>	<i>Modelo D</i>
<i>Total dos meses</i>	0.987	0.989	0.991	0.992
<i>Conjunto de Teste</i>	0.965	<b>0.976</b>	<b>0.973</b>	<b>0.985</b>

## 6. CONCLUSÕES

Em trabalhos anteriores dos autores foram experimentados modelos baseados em redes neuronais artificiais usando o índice relativo ao tempo, para tentar captar a tendência crescente das séries de turismo. Os modelos foram experimentados com a série da



região Norte com sucesso para o caso de 3 modelos usando índices para o ano e para o mês combinados de formas diferentes.

Neste trabalho os autores usam a série da Madeira para confirmar a melhoria da capacidade de previsão dos modelos referidos. A série da Madeira não tem uma sazonalidade tão bem definida no início, o que seria, à partida, um problema adicional para os modelos experimentados, contudo os modelos resolveram muito bem o problema e obtiveram inclusive uma melhoria na qualidade de previsão em relação à série da região Norte.

Foram usados: um modelo denominado de clássico usando nas entradas os valores dos últimos 12 meses; um modelo denominado modelo B tendo apenas duas entradas com os índices do ano e do mês do ano; um modelo denominado C tendo na entrada um índice para o mês combinado com as doze entradas do modelo clássico; e um modelo D combinando o modelo clássico e o modelo B.

A análise dos resultados realizada na secção anterior mostra que a utilização de índices temporais nos modelos B, C e D é adequada para que o modelo capture a evolução da tendência de crescimento destas séries de turismo.

O modelo C com resultados ligeiramente piores que os demais modelos, não é de descartar até porque apesar de ter um ERM pior que o modelo clássico, mas da mesma ordem, tem, ainda assim uma melhor correlação com os valores originais.

O modelo B mostra que apesar de ter apenas 2 nós na entrada, ao contrário dos 12 nós do modelo clássico, consegue capturar a sazonalidade e a tendência com melhores resultados que este, quer ao nível do ERM quer ao nível do coeficiente de correlação.

Finalmente, o modelo D, que combina os dois melhores modelos num só, consegue ser melhor que o modelo clássico ao nível do ERM e do coeficiente de correlação. Inclusivamente é o melhor dos modelos B, C e D. Contudo tem 14 nós na camada de entrada contra os 2 nós apenas do modelo B.

Assim, fica confirmado o interesse da utilização do índice temporal nos modelos baseados em RNA para previsão deste tipo de séries com uma tendência de evolução crescente.

## **BIBLIOGRAFIA**



Demuth, Howard; Beale, Mark and Hagan, Martin, 2008, *Neural Network, Toolbox™ 6: User's Guide*, United States.

Fernandes, Paula Odete; (2005); “*Modelling, Prediction and Behaviour Analysis of Tourism Demand in the North of Portugal*”; Ph.D. Thesis in Applied Economy and Regional Analysis; Valladolid University - Spain.

Fernandes, Paula Odete and Teixeira, João Paulo; (2007); “A new approach to modelling and forecasting monthly overnights in the Northern Region of Portugal”; *Proceedings of the 15th International Finance Conference (CD-ROM)*; Université de Cergy; Hammamet, Medina, Tunísia.

Fernandes, P. Odete e Teixeira, J. Paulo; (2009); “New Approach of the ANN Methodology for Forecasting Time Series: Use of Time Index”, *Proceeding of ICTDM, Kos-Grecia*.

Fernandes, P. Odete; Teixeira, J. Paulo; Ferreira, João M. and Azevedo, Susana G.; (2008); “Modelling Tourism Demand: A Comparative Study between Artificial Neural Networks and the Box-Jenkins Methodology”; *Romanian Journal of Economic Forecasting*, No 5(3), pp. 30/50. ISBN:978-84-92453-69-6.

Hagan, M. T., Menhaj, M., “Training feedforward networks with the Marquardt algorithm”, *IEEE Transactions on Neural Networks*, vol. 5, n 6, 1994.

Haykin, Simon; 1999; “*Neural Networks. A comprehensive foundation*”; New Jersey, Prentice Hall.

INE; (1987-2009); “*Anuários Estatísticos do Turismo de 1987 a 2009*”. Lisboa.

Machado, Telmo; Teixeira, João e Fernandes, Paula; (2010); “Modelação da procura turística em Portugal: regressão linear versus redes neuronais artificiais”; *Revista Turismo & Desenvolvimento*, Vol., pp.435/445. ISSN:1645-9261.

PENT; (2007); “*Plano Estratégico Nacional do Turismo. Para o Desenvolvimento do Turismo em Portugal*”; Ministério da Economia e Inovação; Lisboa.

Rumelhard, D. E. and McClelland, J. L.. (1986). *Parallel Distributed Processing – Explorations in the Microstructure of Cognition*. Volume 1 – Foundations, The Massachusetts Institute of Technology Press.

Witt, Stephen F. & Witt, Christine A.; (1995); “Forecasting tourism demand: a review of empirical research”; *International Journal of Forecasting*; No 11, pp.447/475.

Wong, K. F.; (2002); “Introduction: Tourism Forecasting State of the Art”; *Journal of Travel and Tourism Marketing*; No 13 (1/2), pp.1/3.

Yu, Gongmei & Schwartz, Zvi; (2006); “Forecasting Short Time-Series Tourism Demand with Artificial Intelligence Models”; *Journal of Travel Research*; No 45, pp. 194/203.