

# PREVISÃO DA SÉRIE TEMPORAL TURISMO COM REDES NEURONAIAS ARTIFICIAIS

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

Escola Superior de Tecnologia e Gestão (ESTiG)  
Instituto Politécnico de Bragança (IPB)

## RESUMO

O objectivo da investigação em curso consiste no desenvolvimento de modelos baseados em redes neuronais artificiais (RNA) e a sua aplicação de forma a prever a procura turística caracterizada na série temporal em estudo. A presente publicação contribui para os estudos económicos através da utilização destas redes. As RNA podem ser definidas como estruturas que compreendem a inter-conectividade adaptativa de elementos simples de processamento, com capacidade de computação paralela massiva para o processamento de dados de entrada que representam o conhecimento. As RNA têm a capacidade de aprender a partir de um conjunto de dados e experiências, de identificar padrões ou sequências e fazer generalizações para o futuro. A sua popularidade pode ser atribuída à sua capacidade única de simular uma grande variedade de dispositivos não lineares.

Este trabalho foca-se no tratamento, análise e modelação da série temporal que representa o número mensal de dormidas em hotéis na região norte de Portugal, entre Janeiro de 1987 e Dezembro de 2006. Foram testados vários processamentos prévios desta série, diferentes arquitecturas das RNA e diferentes algoritmos de treino destas redes. Foi ainda testada alguma variação no que concerne ao conjunto de dados de entrada da rede.

Na generalidade os resultados apresentados no processo de previsão pelos diversos modelos podem considerar-se a um nível razoavelmente próximo dos dados reais (objectivo). Segundo Lewis (1982) a generalidade dos modelos estudados consegue uma previsão com uma precisão elevada. Desta forma podem ser considerados adequados ao problema de previsão desta serei temporal como referência e de outras da mesma natureza.

## 1. INTRODUÇÃO

Diversos estudos que têm vindo a ser desenvolvidos e publicados na área do turismo são unânimes em considerar fundamental a previsão da procura turística para o processo de planeamento e tomada de decisão (Witt & Witt, 1995; Wong, 2002; Fernandes, 2005; Yu & Schwartz, 2006).

Actualmente estão disponíveis diversas metodologias para a tarefa de previsão em múltiplas aplicações apresentando diferentes características e níveis de complexidade. Os modelos de previsão de Box-Jenkins pertencem a uma família de modelos algébricos designados de ARIMA. Estes modelos permitem fazer a previsão de séries temporais estacionárias. Considera-se que a série temporal é uma provável realização de um determinado processo estocástico. A construção do modelo consiste em identificar o modelo matemático que melhor represente este processo estocástico desconhecido. Em alternativa, tem vindo a ser crescente a aplicação de modelos estatísticos baseados em redes neuronais artificiais (RNA) para modelar estes processos estocásticos. As RNA são modelos não lineares, respondendo bem a

problemas não lineares. O seu desenvolvimento surgiu como uma tentativa de simulação do funcionamento do cérebro humano, sendo o complexo sistema de neurónios biológicos a sua principal fonte de inspiração, ao ponto de se encontrarem diversas semelhanças na terminologia usada.

Neste trabalho procura-se explorar a modelização com redes neuronais na análise da procura turística. Esta metodologia tem vindo a suscitar o interesse na área das ciências económicas e empresariais tendo-se verificado ser uma alternativa valiosa aos métodos clássicos de previsão, e por vezes conseguindo dar resposta a situações que pelos métodos clássicos seriam de difícil tratamento (Thawornwong & Enke, 2004).

A metodologia utilizada centra-se na análise e modelação da série temporal de turismo: Dormidas Mensais nos Estabelecimentos Hoteleiros. Esta série é considerada como demonstradora da actividade turística, devido às suas características que contempla o número de visitantes que usufruíram das capacidades turísticas. O estudo foi realizado para a região Norte de Portugal (Fig. 1). Assim a análise desenvolvida neste trabalho recai sobre a modelação da série temporal Dormidas Mensais Registadas na Região Norte de Portugal. Os dados observados consideram o período compreendido entre Janeiro de 1987 e Dezembro de 2006, correspondendo a 240 observações mensais ao longo de 20 anos. Os valores da série foram facultados pelo Instituto Nacional de Estatísticas de Portugal (INE).

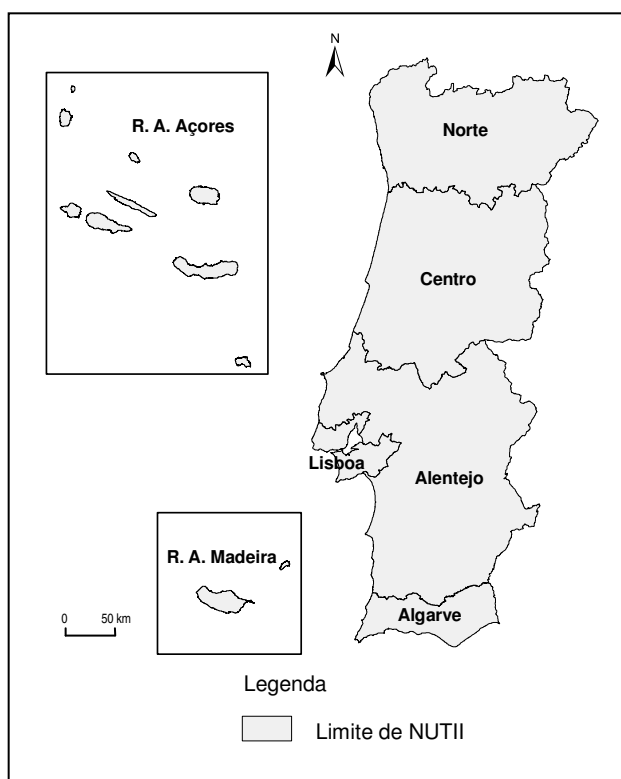


Figura 1. Divisão de Portugal por NUT II.

A metodologia utilizada é facilmente adaptável à previsão de outras variáveis turísticas de uma natureza semelhante como sejam por exemplo a previsão de dormidas num Hotel em particular, ou outra situação em que ocorra uma procura sazonal em que se pretenda conhecer a procura para o mês seguinte ou como no caso dos modelos apresentados nos 12 meses seguintes, com base nos dados actuais.

Para a utilização da metodologia de redes neuronais é necessário ter um conhecimento prévio de um conjunto suficiente de observações do fenómeno que se pretende prever, que seja suficientemente representativo e caracterizador da realidade desse fenómeno.

## 2. A SÉRIE TEMPORAL

A série temporal Dormidas Mensais Registadas na Região Norte de Portugal apresentada na Fig.2, mostra uma sazonalidade anual, atingindo o seu pico máximo, tipicamente em Agosto, e o seu pico mínimo em Dezembro ou Janeiro e apresenta um pico intermédio que pode surgir entre Abril e Junho. A origem deste pico está relacionada com a Páscoa que tradicionalmente atrai muitos turistas à região.

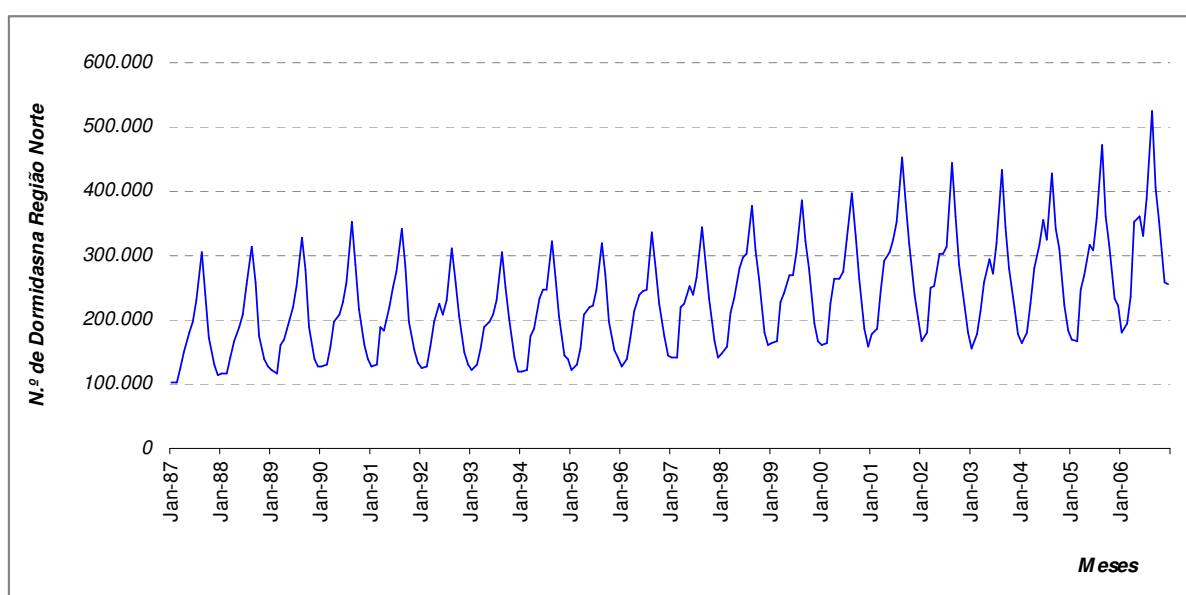


Figura 2. Cronograma da série para o período de 1987:01 a 2006:12. Série original.

A série em estudo apresenta um crescimento acentuado a partir de 1998 até 2001, com um decréscimo até 2004, mas em valores superiores aos anteriores a 1998, e um crescimento muito acentuado em 2005 e 2006. Estas variações podem dificultar a previsão realizada pela RNA, uma vez que ocorre uma evolução temporal fazendo com que o objecto a prever tenha valores superiores daqueles para que foi treinada a rede neuronal.

Para tentar atenuar os efeitos nefastos deste crescimento para a previsão com esta metodologia, foi considerada uma nova série obtida a partir da série original pela aplicação do logaritmo, a que chamaremos série logarítmica (Fig.3).

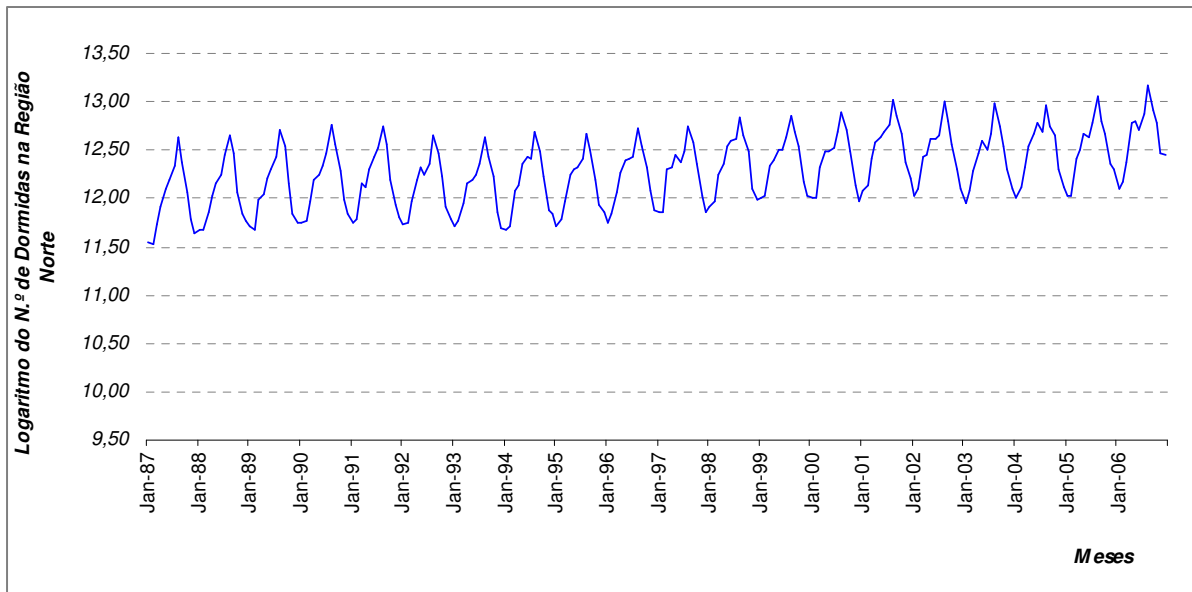


Figura 3. Cronograma da série para o período de 1987:01 a 2006:12. Série logarítmica.

Para atenuar o efeito da sazonalidade foram também processados os dados de entrada pela aplicação das diferenças simples e sazonal, de acordo com a equação às diferenças [1].

$$y(n) = x(n) - x(n-1) + x(n-13) - x(n-12) \quad [1]$$

Em que:

$x(n)$  é a série a limpar;

$y(n)$  é a série limpa.

Estas diferenças foram aplicadas à série original e à série logarítmica, tendo-se obtido as novas séries denominadas, respectivamente série original limpa (Fig. 4) e série logarítmica limpa (Fig. 5). Nestas figuras, os valores apresentados não correspondem directamente ao mês indicado no gráfico, uma vez que esse valor corresponde à aplicação da eq. [1]. Nos casos destas séries, os valores previstos pelas RNA são posteriormente processados para inverter o processo de aplicação das diferenças simples e sazonais.

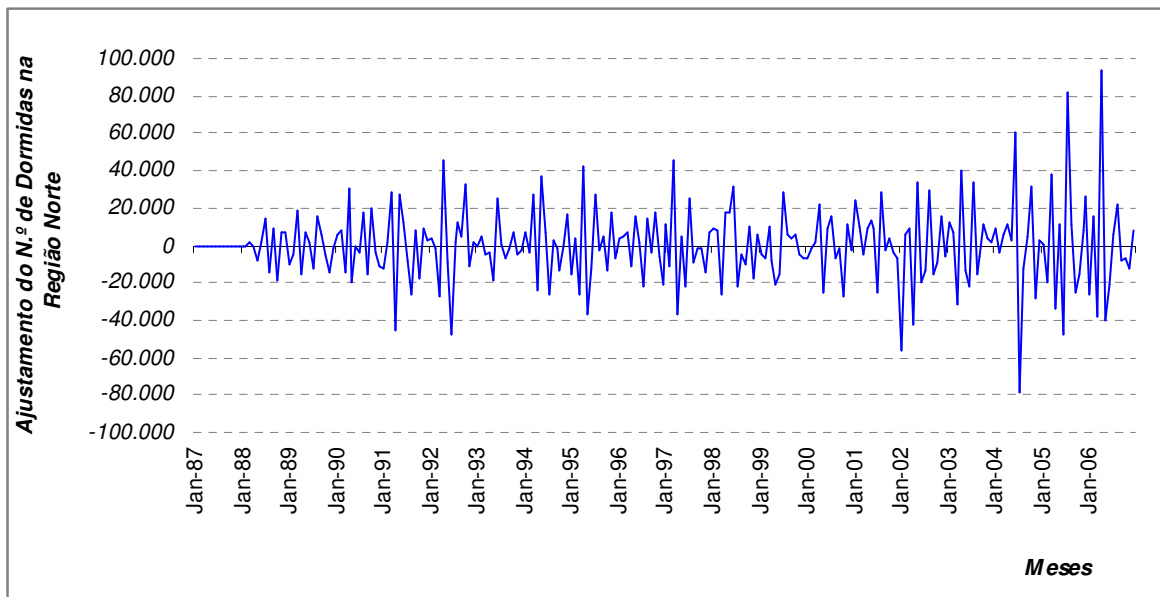


Figura 4. Cronograma da série para o período de 1987:01 a 2006:12. Série original limpa.

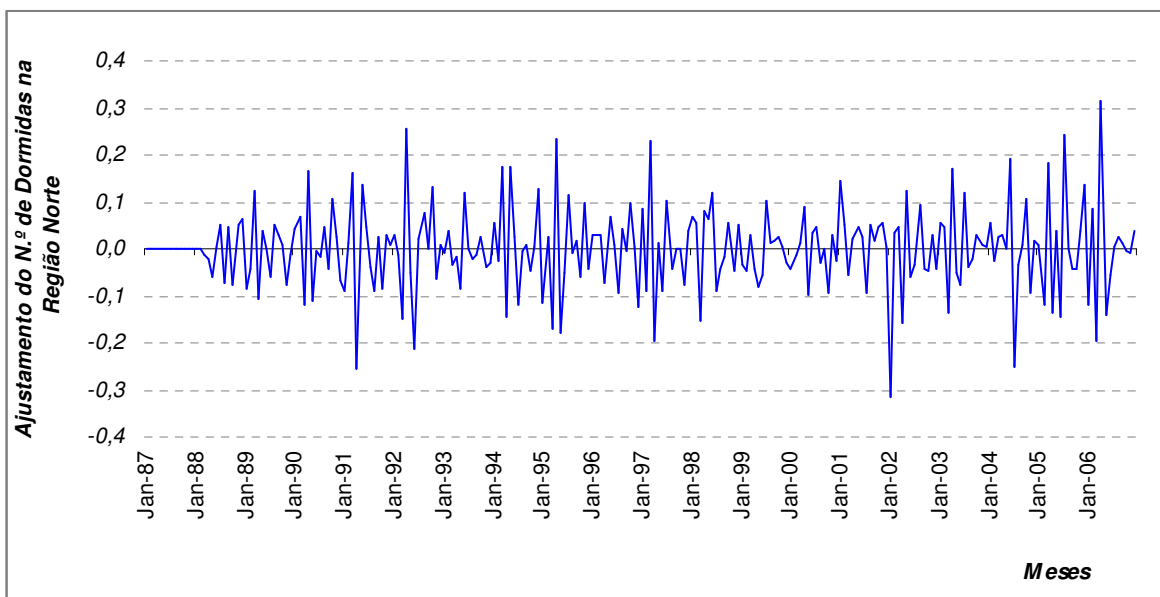


Figura 5. Cronograma da série para o período de 1987:01 a 2006:12. Série limpa logarítmica.

A metodologia de redes neuronais impõe um conjunto de treino, eventualmente um conjunto de validação e um conjunto de teste do modelo. Cada conjunto deve ser uma boa representação do universo do problema a modelar. O conjunto de treino é usado para ‘ensinar’ a rede e normalmente corresponde à maior parte dos dados disponíveis. O conjunto de validação é usado como uma técnica alternativa de paragem do treino da rede para evitar ‘*overfitting*’. O conjunto de teste não é visto na fase de construção do modelo e é usado apenas na fase de medição dos resultados conseguidos.

Neste trabalho foi usado um ano, a que correspondem 12 amostras (12 meses), para o conjunto de validação e um ano, dois ou três para o conjunto de teste, consoante os casos. Os últimos anos foram sempre reservados para o teste, o ano anterior para validação e os restantes para treino.

### 3. AS REDES NEURONAIS ARTIFICIAIS UTILIZADAS

As RNA são modelos que se destacam dentro de uma ampla área do conhecimento da inteligência artificial. Estes têm como base modelos matemáticos de arquitectura semelhante à do cérebro humano. Uma rede neuronal encontra-se composta por um conjunto de neurónios artificiais, nós ou um conjunto de unidades de processamento, ligadas entre si. As conexões entre as unidades de processamento denominam-se por *sinapses*, que processam e transmitem informação através de funções de activação, sendo as mais utilizadas as funções lineares e as sigmoidais - logística e tangente hiperbólica - (Rodrigues, 2000; Fernandes, 2005). Ainda de referir que os neurónios de uma rede estão estruturados em distintas camadas, conhecidas por camada de entrada, intermédias ou oculta e de saída, sendo as RNA mais comuns, para a previsão de séries temporais, as de multi-camadas ou também designadas por MLP<sup>1</sup> (Bishop, 1995), por forma que um neurónio de uma camada encontra-se ligado com os neurónios da camada seguinte às que pode enviar informação, Fig. 6, (Fernandes, 2005). As redes, dependendo da forma como estão ligadas entre as diferentes camadas podem classificar-se em redes recorrentes<sup>2</sup> (*feedback*) e redes não recorrentes<sup>3</sup> (*feedforward*).

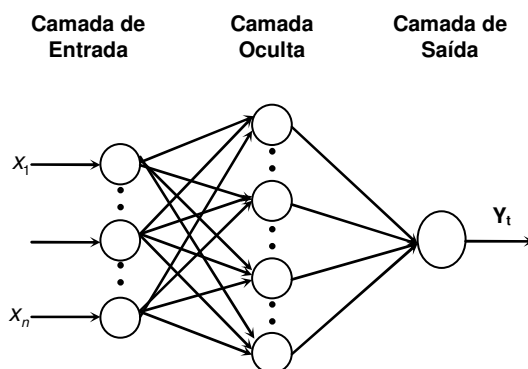


Figura 6. Estrutura de uma Rede Neuronal Artificial *Feedforward*.

Da especificação da rede neuronal também faz parte uma função de erro e um algoritmo para determinar o valor dos parâmetros que minimizam a função de erro. Desta forma, existem dois conceitos centrais: a parte física da rede, ou seja, a sua arquitectura, e o procedimento algorítmico que determina o seu funcionamento, ou seja, a forma como a rede se modifica consoante os dados fornecidos pelo ambiente (Haykin, 1999).

<sup>1</sup> Do inglês *Multi-Layer Perceptron*.

<sup>2</sup> As conexões permitem que a informação regresse a lugares por onde já passou e admite-se também ligações intra-camada (laterais) (Fernandes, 2005).

<sup>3</sup> A informação flui de forma unidireccional de uma camada para outra, desde a camada de entrada para a camada oculta e posteriormente para a camada de saída (Fernandes, 2005).

É importante ainda referir que para as RNA aprenderem com a experiência, têm que ser submetidas a um processo designado de treino, em que a rede é ‘ensinada’ com base num conjunto de exemplos da chamada base de dados de treino. O treino pode ser (i) supervisionados ou (ii) não supervisionado (Haykin, 1999). Para o primeiro caso, durante o processo de treino existe um ‘professor’ que fornece um conjunto de casos de treino. Um caso de treino consiste num vector de entrada  $X$  e no correspondente vector de saída  $Y$ , passando a aprendizagem pela minimização do erro da saída, o que é feito através do ajustamento dos pesos das conexões de acordo com uma determinada regra. Para efectuar este processo de treino existem diferentes algoritmos, sendo um dos mais utilizados na previsão de séries temporais o algoritmo *Backpropagation*<sup>4</sup> nas suas diversas variantes, em que o erro é propagado da retaguarda para o início. No segundo caso, existe um conjunto de entradas que o algoritmo de treino tenta agrupar a partir de padrões por estes apresentados, seguindo assim uma regra de auto-organização (Haykin, 1999; Fernandes, 2005).

O valor produzido por uma rede *feedforward*, com uma camada oculta, pode ser expresso da seguinte forma (Fernandes & Teixeira, 2007):

$$Y_t = b_{2,1} + \sum_{j=1}^n \alpha_j f \left( \sum_{i=1}^m \beta_{ij} y_{t-i} + b_{1,j} \right) \quad [2]$$

Onde,

$m$ , número de nós na camada de entrada;

$n$ , número de nós na camada oculta;

$f$ , função de activação sigmoideal;

$\{\alpha_j, j = 0, 1, \dots, n\}$ , vector de pesos que liga os nós da camada oculta aos da camada de saída;

$\{\beta_{ij}, i = 0, 1, \dots, m; j = 1, 2, \dots, n\}$ , pesos que relacionam os nós da camada de entrada aos da camada oculta;

$b_{2,1}$  y  $b_{1,j}$ , indicam os pesos dos termos independentes (*bias*) associados a cada nó da camada de saída e da camada oculta, respectivamente.

A equação indica também a utilização duma função de activação linear na camada de saída.

O vector de entradas do modelo, corresponde a uma representação de um conjunto de características associadas ao acontecimento de entrada. Uma correcta definição deste vector de características é decisivo para um bom desempenho da RNA. Normalmente, devem ser escolhidas características que estejam relacionadas com a saída, e depois codificadas de forma compacta, mas sem perda de informação. No caso das séries temporais, este vector de características corresponde aos valores anteriores da série temporal.

---

<sup>4</sup> Este algoritmo procura o mínimo da função de erro no espaço de procura dos pesos das ligações entre os neurónios, baseando-se em métodos de gradiente descendente. A combinação de pesos que minimiza a função de erro é considerada a solução para o problema de aprendizagem. A descrição do algoritmo pode ser analisada em Rumelhart and McClelland (1986) e em Haykin (1999).

Também a saída (um único elemento ou um vector de elementos) deve ser uma codificação do resultado correspondente a cada acontecimento caracterizado no vector de entrada. No caso das séries temporais a saída corresponde ao valor da série temporal (eventualmente codificado) para um determinado momento.

A arquitectura da rede corresponde à caracterização da rede quanto ao número de camadas escondidas (normalmente uma é suficiente), quanto ao número de nós em cada camada e quanto à função de activação dos nós (sinapses) de cada camada. Quanto maior o número de nós e de camadas, maior é o número de pesos associados ao modelo. Um problema mais complexo pode necessitar de uma RNA de maior dimensão em termos do número de nós na camada escondida. Uma boa selecção da arquitectura da RNA para um determinado problema é importante, para que a rede tenha a capacidade de dar uma boa resposta ao problema e não ser demasiado exigente em termos do seu número de pesos, já que mais pesos obriga a uma maior quantidade de situações de treino.

#### 4. MODELOS EXPERIMENTADOS

Para o problema em estudo de previsão de dormidas, considerou-se que o modelo deve ter a possibilidade de prever até 12 meses em diante. As entradas da rede são o número de dormidas nos meses anteriores. E a saída será a previsão de dormidas para o(s) mês(es) futuro(s).

Depois foram consideradas várias possibilidades em relação às seguintes alternativas:

- Série usada - foram testadas as 4 séries apresentadas atrás como representação do número de dormidas de cada mês;
- Vector de entrada - foram consideradas várias combinações de meses anteriores: 4 meses anteriores (1-4), 12 meses anteriores (1-12), 13 meses anteriores (1-13), os meses anteriores 1 e 12 (1,12), os meses anteriores 1, 2 e 12 (1, 2, 12); os meses anteriores 1, 12 e 13 (1, 12, 13), os meses anteriores 1, 2, 12 e 13 (1, 2, 12, 13);
- Saída - foi testada a previsão para o mês seguinte (uma saída) e para os 12 meses seguintes (12 saídas);
- Algoritmo de treino - foram testados dois algoritmos de treino com bons resultados: *Resilient Backpropagation* (Riedmiller & Braun, 1993) e *Levenberg-Marquardt* (Marquardt, 1963);
- Arquitectura - foram consideradas RNA apenas com uma camada escondida, com duas funções sigmoidais de activação na camada escondida (tangente hiperbólica [*Tangsig*] e logística [*Logsig*]) e com diversos nós na camada escondida que variaram entre 1 e 10.

Uma comparação exaustiva dos modelos experimentados pode ser encontrada em (Fernandes, 2005), usando dados até 2003. A escolha do melhor modelo em cada situação foi baseada nos resultados (raiz do erro quadrático médio - RMSE) obtidos no conjunto de teste.

A condição do modelo fazer a previsão para os 12 meses seguintes de uma vez só não apresentou resultados competitivos com os modelos que fazem a previsão para apenas o mês seguinte. Por essa razão não se apresentam resultados para esta situação. Os modelos com uma saída fazem a previsão para um ano em 12 previsões.

Apresenta-se na tabela seguinte uma comparação das melhores alternativas com base nos resultados obtidos. Nesta tabela, é apresentado na primeira coluna a arquitectura indicando os meses anteriores que estão presentes na entrada seguido do número de nós na camada escondida e do número de nós da saída. Por exemplo a RNA (1-12;6;1) tem os meses 1 a 12 na sua entrada, e portanto 12 entradas, 6 nós na camada escondida e uma saída, no caso da RNA (1,12;10,1) tem os meses 1 e 12 nas entradas, e portanto 2 entradas, 6 nós na camada oculta e um na camada de saída. Na segunda coluna é indicado o algoritmo de treino (LM, *Levenberg-Marquardt* ou RP, *Resilient Backpropagation*). A terceira coluna indica a função de activação da camada escondida. Na camada de saída a função de activação usada foi sempre uma função puramente linear. Nas restantes 4 colunas são indicados os melhores valores conseguidos para cada série testada, medidos pela RMSE no conjunto de teste (últimos 12 meses).

Tabela 1. Avaliação dos Resultados para as melhores Arquitecturas.

Arquitectura	Algoritmo de Treino	Função de Activação	Avaliação da <i>Performance</i> (RMSE)			
			Série Original	Série Logarítmica	Série Original Limpa	Série Logarítmica Limpa
(1-4;2;1)	LM	Tansig	26.067	28.734	10.253	10.381
	LM	Logsig	26.431	28.277	<b>9.392</b>	<b>9.210</b>
	RP	Tansig	39.437	42.048	10.032	20.215
	RP	Logsig	41.602	39.604	11.808	17.770
(1-12;6;1)	LM	Tansig	13.329	14.355	<b>8.008</b>	<b>8.059</b>
	LM	Logsig	14.748	11.057	<b>9.147</b>	<b>8.422</b>
	RP	Tansig	20.378	16.860	<b>7.678</b>	<b>9.500</b>
	RP	Logsig	16.184	17.046	<b>5.600</b>	<b>8.145</b>
(1-13;10;1)	LM	Tansig	14.793	15.828	<b>9.448</b>	<b>7.508</b>
	LM	Logsig	12.805	11.813	<b>8.022</b>	<b>7.550</b>
	RP	Tansig	12.850	10.913	<b>9.125</b>	<b>7.714</b>
	RP	Logsig	12.851	17.663	<b>6.825</b>	<b>9.487</b>
(1,12;10;1)	LM	Tansig	14.375	12.507	10.164	<b>9.601</b>
	LM	Logsig	13.144	14.426	12.655	<b>9.967</b>
	RP	Tansig	17.307	13.381	12.372	10.396
	RP	Logsig	15.665	14.843	10.489	10.546
(1,2,12;2;1)	LM	Tansig	14.614	13.436	<b>9.566</b>	<b>8.102</b>
	LM	Logsig	16.548	13.001	10.072	<b>8.007</b>
	RP	Tansig	15.984	16.524	<b>8.481</b>	<b>7.621</b>
	RP	Logsig	16.685	13.582	11.443	<b>8.702</b>
(1,12,13;4;1)	LM	Tansig	13.988	13.681	<b>9.767</b>	<b>9.782</b>
	LM	Logsig	13.432	14.100	<b>9.920</b>	<b>8.498</b>
	RP	Tansig	14.279	14.679	10.829	<b>9.874</b>
	RP	Logsig	14.006	18.710	<b>9.975</b>	10.894
(1,2,12,13;4;1)	LM	Tansig	11.590	12.122	10.274	<b>8.366</b>
	LM	Logsig	10.777	13.287	<b>9.581</b>	<b>9.118</b>
	RP	Tansig	11.298	13.638	<b>8.681</b>	<b>9.492</b>
	RP	Logsig	11.405	15.772	10.758	<b>8.902</b>

As RNA foram treinadas com valores iniciais aleatórios diferentes em diversas sessões. O facto de iniciarem de um ponto de partida diferente em cada sessão leva a um valor final que pode ser significativamente diferente entre sessões. Desta forma, cada rede foi treinada algumas centenas de sessões e foram escolhidas as redes com melhores resultados de entre essas sessões no conjunto de validação. O treino era interrompido pelo conjunto de validação.

Pela análise da Tabela 1 verifica-se que os melhores resultados ocorrem para as séries limpas, e em particular para a série original limpa, para a generalidade das arquitecturas e algoritmos de treino. Os resultados para estas séries foram medidos depois de inverter o processo das diferenças simples e sazonais.

O modelo seleccionado de entre os melhores foi assim a RNA com os últimos 12 meses na sua entrada, com 6 nós na camada escondida, e um nó na camada de saída, com a função de activação logística nos nós da camada escondida. A sua arquitectura é apresentada na Fig. 7. Esta rede foi treinada pelo algoritmo *Resilient Backpropagation* e foi aplicado à série original limpa.

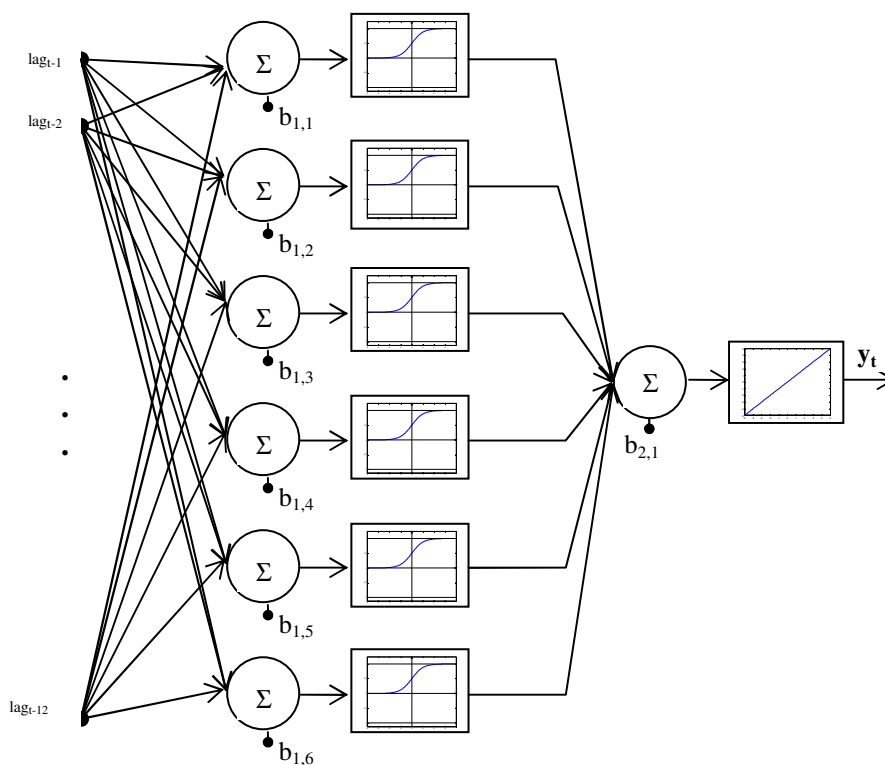


Figura 7. Esquema da Rede Neural Artificial utilizando a arquitectura (1-12; 6; 1) (Fernandes, 2005).

## 5. ANÁLISE DOS RESULTADOS

Faz-se agora uma análise aos resultados obtidos pelo modelo seleccionado.

Para este modelo o conjunto de teste consistiu no ano de 2003. Apresenta-se na Tabela 2 os valores originais e previstos pelo modelo para esse ano, bem como uma medida do erro

percentual absoluto. Verifica-se um erro percentual absoluto relativamente baixo, e nunca superior a 6,3%. O erro percentual absoluto médio (EPAM) é de 1,6 %. O RMSE foi de 5.600 e o coeficiente de correlação  $r$  de 0,999.

Apresenta-se na Fig. 8 a evolução dos valores previstos pelo modelo e os valores reais.

Tabela 2. Previsão para o ano de 2003 (teste).

Meses	Valores Originais	Previsão	Erro Percentual Absoluto
Janeiro	155.527	156.113	0,4%
Fevereiro	177.818	175.726	1,2%
Março	214.106	227.638	6,3%
Abril	258.519	262.722	1,6%
Maió	293.531	297.082	1,2%
Junho	271.454	279.798	3,1%
Julho	318.706	320.947	0,7%
Agosto	433.211	441.706	2,0%
Setembro	343.534	345.760	0,6%
Outubro	281.472	280.144	0,5%
Novembro	219.463	219.109	0,2%
Dezembro	178.439	180.588	1,2%
<b>EPAM</b>	-	-	<b>1,6%</b>

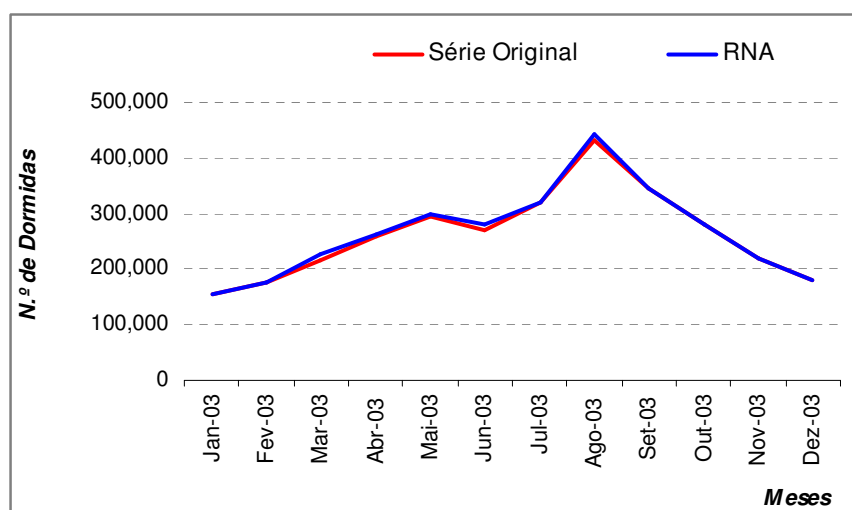


Figura 8. Previsão do modelo e valores alvo para o ano de 2003 (conjunto de teste).

No sentido de avaliar a capacidade de previsão do modelo, realizaram-se previsões para o ano de 2004 (dados ainda não usados neste modelo), que podem ser observadas na Tabela 3 e Figura 9. Assim, pela leitura do gráfico podemos observar que embora os valores da série previstos acompanhem o comportamento da série original, verifica-se algum afastamento entre os valores previstos e os realmente observados, mormente para o mês de Junho, o que permite dizer que o modelo não conseguiu incorporar alguns factos decorridos nesse mês do ano de 2004.

A Tabela 3, mostra-nos, além dos valores observados e previstos, os erros das previsões para o ano de 2004. Verifica-se que os valores previstos revelam-se geralmente ‘subavaliados’ no ano de 2004, sendo de destacar os meses de Maio, Junho, Outubro, Novembro e Dezembro, apresentando erros percentuais absolutos superiores a 10%.

Tabela 3. Previsão para o ano de 2004.

Meses	Valores Originais	Previsão	Erro Percentual Absoluto
Janeiro	162.900	175.211	7.6%
Fevereiro	181.900	188.395	3.6%
Março	224.600	234.134	4.2%
Abril	279.800	267.660	4.3%
Maio	317.300	299.400	5.6%
Junho	355.300	276.511	22.2%
Julho	324.400	319.172	1.6%
Agosto	426.900	416.402	2.5%
Setembro	342.100	330.884	3.3%
Outubro	311.500	259.396	16.7%
Novembro	221.200	196.503	11.2%
Dezembro	182.800	143.979	21.2%
<b>EPAM</b>	-	-	<b>8.7%</b>

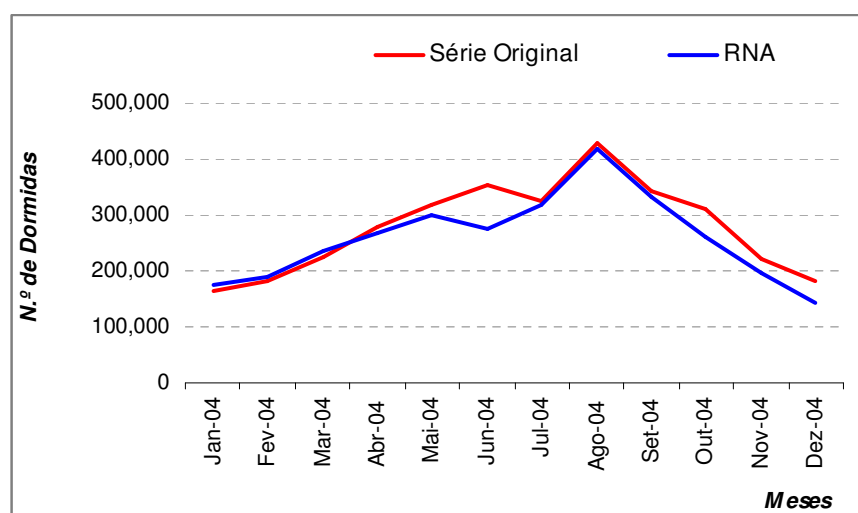


Figura 9. Previsão do modelo e valores alvo para o ano de 2004.

Observando ainda, os dados apresentados na Tabela 3 e Figura 9 será de evidenciar que no ano de 2004 ocorreu um pico de dormidas nos meses de Maio e Junho, relativamente a anos anteriores, coincidente com o evento desportivo Campeonato Europeu de Futebol - EURO2004 - realizado em Portugal nesse período. Nesse evento 5 dos 10 estádios de futebol encontram-se na região em análise. Os piores resultados do modelo ocorreram no período de Junho de 2004, em que ocorreu a fase final do EURO2004 com o consequente aumento no número de dormidas e que obviamente o modelo não previu.

Para este ano de 2004, o RMSE foi de 31.696 e o coeficiente de correlação,  $r$ , foi de 0,946.

Tendo por base o critério de Lewis (1982), podemos dizer que as previsões continuam a ser realizadas com uma precisão elevada, dado que o EPAM (8,7%) registou um valor inferior a 10%.

## 6. CONCLUSÕES

Neste trabalho apresentou-se um estudo de modelação da série temporal relativa ao número de dormidas nos Hotéis da região Norte de Portugal, para efeitos de previsão futura. Foi utilizada uma metodologia de redes neuronais, em que foram testadas diversas variações ao modelo no que concerne às entradas, à sua arquitectura e algoritmo de treino usado. Foram experimentadas 4 séries temporais. A série original, a série obtida pela aplicação do logaritmo à série original e mais duas séries obtidas pela aplicação das diferenças simples e sazonais às duas séries anteriores.

Foi seleccionado um modelo com base na comparação dos resultados dos diversos modelos combinados com as diversas séries e algoritmos de treino. Esse modelo tem na entrada os últimos 12 meses, tem 6 nós na camada escondida, e um nó na camada de saída, com a função de activação logística nos nós da camada escondida e a função puramente linear no nó de saída.

Foi apresentada uma análise dos resultados obtidos por este modelo para o ano de 2003, que foi usado no conjunto de teste e no ano de 2004, cujos dados nunca tinham sido ‘vistos’ pela rede neuronal. Os resultados foram melhores para o ano de 2003. No ano de 2004 o modelo registou um erro percentual absoluto médio de 8,7% que segundo a classificação do EPAM proposto por Lewis ainda se pode classificar como um modelo com precisão elevada, ou seja, poder-se-á dizer que o modelo apresentou qualidades estatísticas e de ajustamento satisfatórias evidenciando ser adequado para a modelação e previsão da série de referência.

A metodologia apresentada neste trabalho pode ser aplicada a outro tipo de séries temporais, sem alterações significativas para fazer previsão de outras variáveis ligadas ao turismo, como sejam a previsão da procura turística, medida por outras variáveis, ou mesmo em determinados sectores da actividade turística.

## REFERÊNCIAS

Bishop, C. M.; (1995); “*Neural Networks for pattern recognition*”; Oxford University Press; Oxford. London.

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 4th International Finance Conference (CD-ROM)*; Université de Cergy; Hammamet, Medina, Tunísia.

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

Lewis, C.D.; (1982); “*Industrial and Business Forecasting Method*”; Butterworth Scientific; London.

Marquardt, Donald; (1963); “An Algorithm for Least-Squares Estimation of Nonlinear Parameters”; *SIAM Journal on Applied Mathematics*; N.º11, pp.431/441.

Riedmiller, M. and Braun, H.; (1993); “A direct adaptive method for faster backpropagation learning: The RPROP algorithm”; *Proceedings of the IEEE International Conference on Neural Networks*.

Rodrigues, Pedro João S.; (2000); “*Redes neuronais aplicadas à segmentação e classificação de leucócitos em imagens*”; Dissertação de Mestrado em Engenharia Electrónica e Telecomunicações. Universidade de Aveiro.

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.

Thawornwong, S. and Enke, D.; (2004); “The adaptive selection of financial and economic variables for use with artificial neural networks”; *Neurocomputing*; N.º6, pp. 205/232.

Witt, Stephen F. and Witt, Christine A.; (1995); “Forecasting tourism demand: a review of empirical research”; *International Journal of Forecasting*; N.º 11, pp.447/475.

Wong, K. F.; (2002); “Introduction: Tourism Forecasting State of the Art”; *Journal of Travel and Tourism Marketing*; N.º 13 (1/2), pp.1/3.

Yu, Gongmei and Schwartz, Zvi; (2006); “Forecasting Short Time-Series Tourism Demand with Artificial Intelligence Models”; *Journal of Travel Research*; N.º 45, pp. 194/203.