

Modelação da **Procura Turística em Portugal**: Regressão Linear *versus* Redes Neurais Artificiais

TELMO NUNO MARTINS MACHADO * [telmo.nm.machado@alunos.ipb.pt]

JOÃO PAULO RAMOS TEIXEIRA ** [joaopt@ipb.pt]

PAULA ODETE FERNANDES *** [pof@ipb.pt]

Resumo | Nos últimos anos, dada a relevância do sector do turismo para a economia de Portugal, a modelação e previsão de séries económicas relacionadas com o turismo têm tido um crescente interesse. Assim, o principal objectivo do presente trabalho assenta no estudo comparativo entre o modelo baseado na regressão linear e o modelo baseado na tecnologia de redes neurais artificiais. A inclusão destas duas metodologias de natureza diferente, tem como finalidade, perceber a respectiva potencialidade aplicada às características peculiares que as séries de turismo evidenciam, tais como, a sazonalidade e a tendência. Recorreu-se para o efeito à série mensal que mede a procura turística: “Dormidas Mensais Registadas nos Estabelecimentos Hoteleiros em Portugal”, para o período compreendido entre Janeiro de 1990 e Dezembro de 2008.

Os modelos desenvolvidos apresentaram qualidades estatísticas e de ajustamento de precisão elevada, pelo que se procedeu à sua utilização para efeitos previsionais. Neste contexto, efectuou-se uma comparação dos resultados previstos com os reais, para os anos de 2007 e 2008, apresentando, para o período em análise, um EPAM de 4.2% e 4.1%, respectivamente para o método de regressão linear (RL) e Redes Neurais Artificiais (RNA).

Palavras-chave | Modelação de Dados, Procura Turística, Séries Temporais, Redes Neurais Artificiais, Previsão.

Abstract | The modulation and forecast of economic time series related with tourism showed an increase interesting, in last years, due to the relevance of the tourism sector for the Portuguese economy. Hence, the central aim of the present paper consists in the comparative study between the linear regression based model and the Artificial Neural Network (ANN) based model. The inclusion of these two different models has the purpose of understand their potentiality to deal with the peculiar characteristics of the tourism time series such as seasonality and trend. The monthly series that measure the tourism demand “Monthly Guest Nights in Hotels” between January 1990 and December 2008 was used.

* Licenciado em Engenharia Informática e Mestre em Sistemas de Informação pelo Instituto Politécnico de Bragança.

** Doutorado em Engenharia Electrotécnica pela Faculdade de Engenharia da Universidade do Porto e Professor Adjunto da Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Bragança.

*** Doutorada em Economia Aplicada e Análise Regional pela Universidade de Valladolid (Espanha) e Professora Adjunta da Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Bragança.

The developed models achieved a high level of statistical quality of adjustment, and therefore they were used for forecast purposes. A comparison between forecast values and original data for the years of 2007 and 2008 were made. The error, measured by the average of the percentage absolute error (EPAM), for the forecast in that period was 4.2% and 4.1% for the linear regression model and ANN model respectively.

Keywords | Data Modelling, Tourism Demand, Time Series, Artificial Neural Networks, Forecasting.

1. Introdução

O turismo, sendo um factor estruturante da dinâmica da economia de qualquer país, se for implementado de uma forma sustentada pode vir a proporcionar benefícios a longo prazo, reflectindo-se na criação de postos de trabalho e na captação de divisas para os diferentes sectores da actividade económica a que está ligado.

À semelhança de outros países, no decurso das últimas décadas, também em Portugal o turismo se transformou num fenómeno singular apesar da situação económica e política lhe serem, muitas vezes, desfavoráveis.

Neste sentido e dado o crescimento substancial deste sector de actividade em Portugal, será de todo proveitoso desenvolver modelos susceptíveis de serem utilizados para efectuar previsões fiáveis da procura turística, dado que esta assume um papel relevante no processo de planeamento e de tomada de decisões tanto no contexto do sector público como do privado. Actualmente, na área da previsão, encontra-se disponível uma grande multiplicidade de métodos que têm vindo a emergir para fazer face às mais variadas situações, apresentando características e metodologias diferentes, que vão desde as abordagens mais simples às mais complexas (Thawornwong e Enke, 2004; Fernandes, 2005; Yu e Schwartz, 2006).

Assim sendo, o propósito do presente trabalho prende-se com a descrição e comparação entre dois modelos desenvolvidos. Um modelo de regressão linear univariado mensal, tendo por base o método

dos mínimos quadrados ordinários e o outro modelo baseado em redes neuronais artificiais (RNA), que tira partido da sua capacidade para modelar problemas não lineares, com frequência associados à previsão de séries temporais.

Por um lado, a regressão linear, apesar de eficiente, reduz-se a modelar problemas de natureza linear, mostrando ser uma metodologia muito limitada, em relação ao elenco de problemas que resolve. As redes neuronais artificiais acarretam um processo mais pesado em termos de cálculo, no entanto mostram ser eficazes na modelação de problemas não lineares, com relação matemática desconhecida.

A série temporal, que serviu de referência ao estudo foi: "Dormidas Mensais Registadas nos Estabelecimentos Hoteleiros em Portugal". Estes dados estão disponíveis, no EUROSAT (2009). Os dados observados reportam-se ao período compreendido entre Janeiro de 1990 e Dezembro de 2008, correspondendo assim a um total 228 observações mensais, 19 para cada mês, 12 para cada ano.

O presente estudo encontra-se estruturado da seguinte forma: depois desta introdução apresenta-se uma secção que aborda as duas metodologias utilizadas na modelação neste trabalho. Segue-se a apresentação da série temporal "Dormidas Mensais Registadas em Portugal". Depois descreve-se a implementação prática do modelo de regressão linear e do modelo de RNA, definindo as suas variáveis e procedimentos associados à utilização destas metodologias. Apresenta-se ainda a previsão da

procura turística, com base nos modelos desenhados previamente, para os anos 2007 e 2008 e analisa-se a *performance* dos mesmos. Por último, serão apresentadas as conclusões do estudo.

2. Metodologias

2.1. Regressão linear

A regressão linear é uma ferramenta estatística utilizada para traçar a tendência de funções com variação linear. Este método, apesar de simples, quando é convenientemente aplicado, pode produzir resultados satisfatórios na previsão de séries temporais quando estas são lineares.

Os coeficientes de regressão linear simples, são estimados utilizando o método dos mínimos quadrados ordinários. Pretende-se assim minimizar a soma do quadrado dos erros observados de forma a obter as equações do tipo (Johnston e Dinardo, 2000):

$$Y_j = \beta_0 + \beta_1 X_j + \varepsilon_j \quad [1]$$

Onde, β_0 , representa a ordenada na origem e β_1 , o declive da recta. Sendo, X_j , a variável independente para cada mês j , a variar no intervalo $[1:j]$ e o termo ε , os erros ou resíduos do modelo.

À semelhança do que acontece com outros modelos, também o modelo de regressão tem por base alguns pressupostos, sabendo-se que os principais pressupostos¹ recaem sobre a componente do erro do modelo, ε_j , e a validação é realizada através dos resíduos (Gujarati, 2003).

2.2. Redes neuronais artificiais (RNA)

Sob a alçada das ciências que se dedicam ao estudo dos processos de aprendizagem, e inspirada no funcionamento do cérebro humano, a metodologia de redes neuronais artificiais é uma ferramenta que se coaduna bem com os problemas relacionados com a previsão de séries temporais. Hill *et al.* (1996) e Hansen *et al.* (1999) referem que as RNA apresentam capacidade para melhorar a previsão de séries temporais através da análise de informação adicional, diminuindo a sua dimensão e reduzindo a sua complexidade.

O conceito subjacente a esta tecnologia assenta no facto do conhecimento ser adquirido pela rede a partir dos dados de entrada (ambiente), através de um processo de aprendizagem. Os pesos das conexões entre os neurónios (sinapses) são usados para armazenar o conhecimento adquirido (Haykin, 1999; Basheer e Hajmeer, 2000).

O processo de aprendizagem pode ser supervisionado ou não supervisionado, dependendo da forma como o processo de treino é assistido ou não. Em redes supervisionadas, de particular interesse para o nosso estudo, existe um conjunto de dados de treino e de valores alvo que é fornecido à rede, numa primeira fase, de modo que o ajuste do peso de cada ligação, seja actualizado de forma a produzir a saída desejada. Este processo de ajuste, normalmente está associado a um determinado algoritmo de treino.

Os tipos de redes neuronais mais habitualmente usadas, por terem melhores resultados, dividem-se entre redes lineares, redes recorrentes, redes radiais e redes *feed-forward* multi-camada. Estas últimas são mesmo as mais utilizadas devido ao sucesso da sua utilização nas mais diversas áreas desde a Engenharia, Economia, Medicina, etc, desde que em 1986 Rumelhart e McClelland (1986) apresentaram um algoritmo capaz de as treinar denominado de algoritmo de *Backpropagation*. Neste trabalho foi usada uma rede deste tipo, pelo que se descreve com mais detalhe.

¹ [i] Assume-se que o erro tem média nula e variância constante; [ii] Assume-se que os elementos do erro são independentes entre si; [iii] O erro deve ter uma distribuição Normal para que se possa usar a inferência estatística e generalizar à população os resultados obtidos a partir da amostra (Gujarati, 2003).

Basicamente, em termos de arquitectura, uma rede neuronal *feed-forward* multi-camada é constituída por um conjunto de elementos, designados neurónios, dispostos em várias camadas (MLP²), concretamente, uma camada de entrada, uma ou mais camadas escondidas e uma de saída (Tsauro et al., 2002), como disposto na Figura 1. O fluxo de informação flui da entrada para a saída, pelo que se denomina de *feed-forward*. As funções de activação ou transferência mais utilizadas são as lineares e as sigmoidais³, dependendo do que se pretende à saída, apresentadas na Figura 2.

A aprendizagem da rede é realizada num processo iterativo em que os pesos das sinapses, w_{ij} , e desvios dos neurónios, b_k , são ajustados de forma a minimizar o erro entre a saída da rede e o alvo dado para um determinado conjunto de treino, como se ilustra na Figura 3.

Em redes *feed-forward* com uma ou mais camada escondida, o algoritmo de treino mais vulgar é baseado na retro-propagação do erro (algoritmo

Backpropagation). Este algoritmo permite, após o cálculo do erro, propagar esse mesmo erro para as camadas predecessoras, de modo a dividi-lo pelos pesos que mais contribuíram para ele, recorrendo ao gradiente da função de erro.

O processo de treino é interrompido quando se atingem níveis de *performance* desejáveis na função de erro. Habitualmente, a função de erro é a função erro quadrático médio MSE⁴. O treino pode ainda ser interrompido quando se atinge um limite de iterações previamente definido. Existem outros critérios de paragem, nomeadamente, a utilização de um conjunto de validação, que permite impedir que o processo de treino imponha uma perda do poder de generalização na classificação do conjunto de validação. Denomina-se habitualmente por validação cruzada.

² Do inglês *Multi-Layer Perceptron*.

³ Podem ser logísticas ou tangentes hiperbólicas.

⁴ Do inglês *Mean Squared Error*.

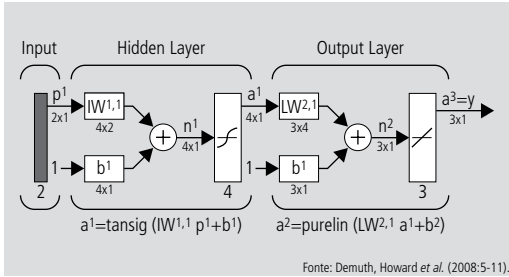


Figura 1 | Rede neuronal do tipo *feed-forward* genérica.

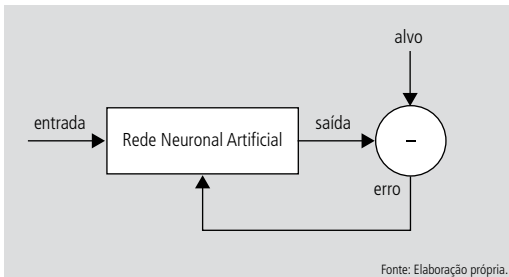


Figura 3 | Processo de ajuste dos parâmetros da RNA em aprendizagem supervisionada.

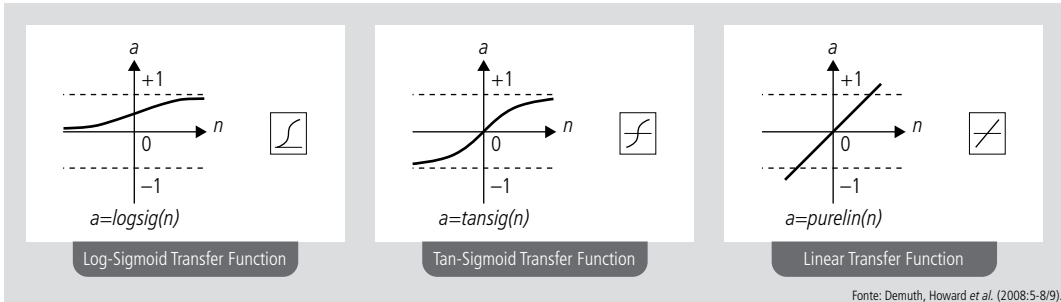


Figura 2 | Funções de activação.

3. Apresentação e análise da série

Na Figura 4, pode observar-se a série temporal “Dormidas Mensais Registadas nos Estabelecimentos Hoteleiros em Portugal”. Os dados recolhidos consideram o período compreendido entre Janeiro de 1990 e Dezembro de 2008, correspondendo a 228 dados mensais ao longo de 19 anos.

Optou-se por estudar esta série, uma vez que tem sido considerada como significativa da actividade turística devido às suas características e contempla os visitantes que usufruíram das capacidades turísticas de Portugal, nesse período.

Analisando o seu comportamento verifica-se a existência de oscilações periódicas em torno de um valor médio crescente, sugerindo a presença de sazonalidade (valores máximos nos meses de Verão e mínimos nos meses de Inverno), isto é, os valores das dormidas vêm em função da altura do ano.

A Figura 4 mostra que o turismo em Portugal, em particular a procura de estabelecimentos hoteleiros, segue um comportamento bastante típico, evidenciando para cada ano, um pico no mês de Agosto, seguido de uma queda abrupta até Janeiro e novamente uma subida na direcção dos meses de Verão. Esta característica, conhecida como sazonalidade, resulta da associação que existe entre a procura turística e as condições

climáticas que a favorecem. Segundo Fúster (1989), a sazonalidade é uma das características do fenómeno turístico. Butler (2001) refere que a sazonalidade pode apresentar-se sob duas formas: a natural e a institucional. Relativamente à primeira, pode resultar de alterações nas condições climáticas; já a segunda, pode resultar de aspectos relacionados com as instituições sociais, religiosas, éticas e culturais, sendo menos previsível que a sazonalidade natural.

Também se pode constatar, que existe uma tendência crescente ao longo do intervalo de tempo abrangido neste estudo, sendo esta manifestamente mais evidente em meses com maior valor absoluto de dormidas. A variação é mais acentuada para os valores da série, relativos ao mês de Agosto, se compararmos com a evolução do mês de Janeiro ao longo dos anos.

A tendência registada é resultado do crescimento económico e dos investimentos no sector turístico, que se têm verificado em Portugal nos últimos anos. Verifica-se, no entanto, que esta tendência não é tão linear quanto isso. O início da década de noventa apresenta uma tendência decrescente, que se prolonga até o ano de 1994. O mesmo aconteceu na transição do ano 1998 para 2000. Estas tendências decrescentes coincidem com períodos de recessão ou abrandamento económicos.

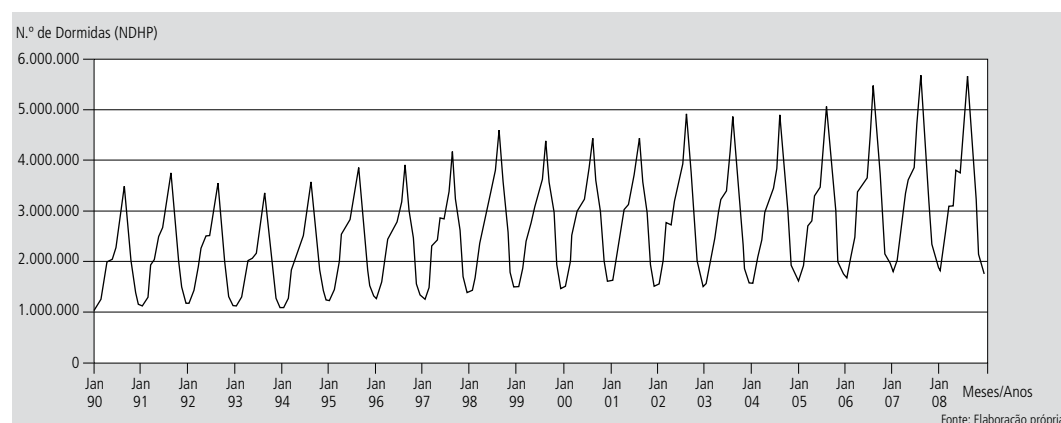


Figura 4 | Dormidas dos turistas nas unidades de alojamento, em Portugal, no intervalo [Jan-90:Dez-08].

4. Aplicação das metodologias

4.1 Modelo de regressão linear (RL)

A série apresentada na secção anterior mostra uma sazonalidade difícil de modelar com uma função linear, contudo Machado *et al.* (2009), decimaram a série em 12 séries resultando num conjunto de 12 séries passíveis de modelação linear. Cada uma destas doze séries corresponde aos valores de cada um dos doze meses, em que os seus valores são o valor de cada mês ao longo dos anos entre Janeiro de 1990 e Dezembro de 2008, num total de 19 amostras anuais para cada mês.

Nas Figuras 5 a 8, apresentam-se as novas séries e as rectas sobre a dispersão que define a procura turística de cada mês, ao longo de 17 anos. Esta representação permite criar uma perspectiva diferente de abordar o problema, simplificando a sua análise. Numa primeira abordagem ao problema apresentado, consegue-se perceber que existe uma relação directa positiva entre a variável independente, o tempo, e a série temporal, que sugere a aplicação de métodos lineares.

O modelo de regressão apresentado atrás, será aplicado ao intervalo [Jan-90:Dez-06], de forma a reservar os dois últimos anos para efectuar a previsão e validar o modelo. Desta forma as rectas de regressão Y_j , serão definidas para X_j a variar no intervalo [1:17]. Apresenta-se no quadro 1 os valores dos coeficientes das rectas de regressão para cada mês.

Uma vez definidas as rectas de regressão, procede-se à reconstrução do restante intervalo, calculando a projecção de Y_j para $X_j=[18:19]$. Efectuada a projecção dos períodos referentes aos anos de 2007 e 2008, procedeu-se à inversão da transformação inicial definindo-se novamente a variável independente no intervalo [Jan-07:Dez-08], obtendo-se assim Y'_j com a previsão da série.

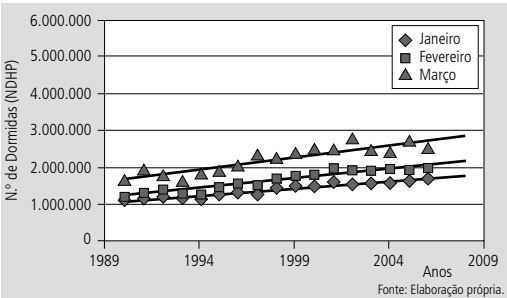


Figura 5 | Representação das rectas de regressão para os meses de Janeiro, Fevereiro e Março, no intervalo [1990:2006].

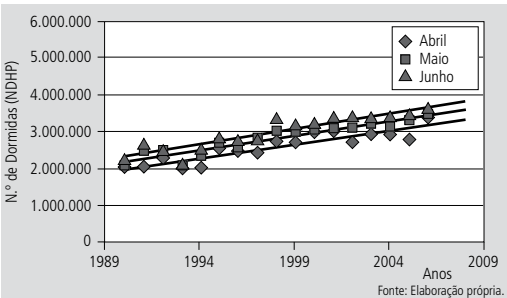


Figura 6 | Representação das rectas de regressão para os meses de Abril, Maio e Junho, no intervalo [1990:2006].

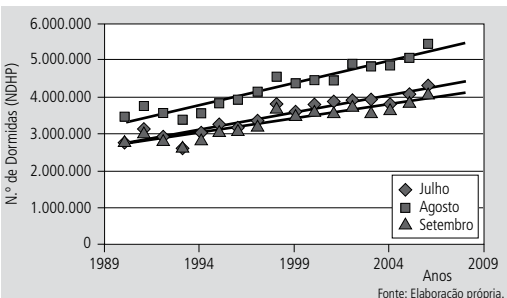


Figura 7 | Representação das rectas de regressão para os meses de Julho, Agosto e Setembro, no intervalo [1990:2006].

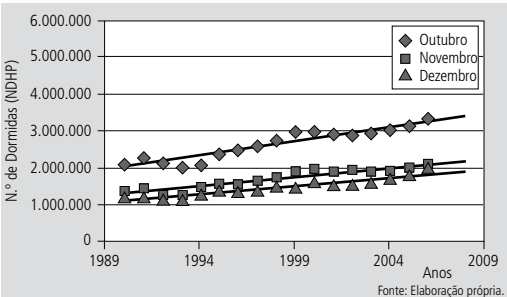


Figura 8 | Representação das rectas de regressão para os meses de Outubro, Novembro e Dezembro, no intervalo [1990:2006].

Quadro 1 | Equações das rectas de regressão linear, mensais; valor da estatística *t-Student* e respectivo valor de prova, para cada coeficiente; coeficiente de correlação

Y_j	β_0 [<i>t-Student</i> ; valor de prova]	β_1 [<i>t-Student</i> ; valor de prova]	$\beta_0 + \beta_1 X_j$	r
Y _{Jan}	1 008 736 [31,679; 0,0000]	40 017 [12,877; 0,0000]	1 008 736+40 017 X	0,96975
Y _{Fev}	1 165 255 [24,584; 0,0000]	51 619 [11,159; 0,0000]	1 165 255+51 619 X	0,95642
Y _{Mar}	1 607 897 [19,029; 0,0000]	64 758 [7,854; 0,0000]	1 607 897+64 758 X	0,92493
Y _{Abr}	1 901 811 [20,729; 0,0000]	74 502 [8,321; 0,0000]	1 901 811+74 502 X	0,91674
Y _{Mai}	2 106 436 [25,259; 0,0000]	77 885 [9,5701; 0,0000]	2 106 436+77 885 X	0,94753
Y _{Jun}	2 217 542 [25,311; 0,0000]	84 188 [9,847; 0,0000]	2 217 542+84 188 X	0,94679
Y _{Jul}	2 641 517 [28,347; 0,0000]	93 323 [10,262; 0,0000]	2 641 517+93 323 X	0,95264
Y _{Ago}	3 193 696 [33,259; 0,0000]	118 627 [12,659; 0,0000]	3 193 696+118 627 X	0,96705
Y _{Set}	2 630 092 [30,520; 0,0000]	78 564 [9,342; 0,0000]	2 630 092+78 564 X	0,94403
Y _{Out}	1 939 979 [26,554; 0,0000]	76 480 [10,727; 0,0000]	1 939 979+76 480 X	0,95486
Y _{Nov}	1 225 761 [22,966; 0,0000]	50 934 [9,779; 0,0000]	1 225 761+50 934 X	0,93976
Y _{Dez}	1 019 961 [24,912; 0,0000]	44 875 [11,231; 0,0000]	1 019 961+50 934 X	0,94826

Fonte: Elaboração própria.

4.2. Modelo com redes neuronais artificiais (RNA)

A abordagem por esta metodologia consiste em considerar a série original, da Figura 4, constituída por 228 amostras distribuídas no intervalo que decorre desde Janeiro de 1990 e Dezembro de 2008.

A organização dos dados foi ajustada ao método de previsão mensal, método este designado, *one-step-ahead forecast*. Desta forma, para prever o mês i da série, inclui-se na entrada da rede o conjunto de valores $[i-p : i-1]$, em que p define a janela temporal, ou seja, o número de amostras do passado a incluir na previsão de um determinado mês i .

A janela temporal p , foi definida com exactamente 12 amostras, ou seja, são utilizados os valores de um ano para prever o mês imediatamente a seguir, razão pela qual, não existe previsão para o primeiro ano. A dimensão da janela temporal definida, tem produzido bons resultados em trabalhos semelhantes (Fernandes *et al.*, 2008).

A segmentação do conjunto de dados em subconjuntos de treino, validação e teste, foi efectuada segundo uma percentagem para cada subconjunto. O conjunto de treino deve reunir o maior número possível de amostras, sem prejudicar obviamente os outros conjuntos, de forma a representar convenientemente o universo das observações. Para o efeito seleccionou-se 78% das amostras, portanto, 168 observações. O conjunto de validação, utilizado como critério de paragem durante o processo de treino, de forma a evitar *overfitting*⁵, inclui 11% das amostras, portanto, 24 observações, correspondentes a uma selecção aleatória⁶ no intervalo que decorre entre, Janeiro de 1990 e Dezembro de 2006. Desta forma, o conjunto de validação reúne amostras de diferentes anos, garantindo uma representação transversal da série. O conjunto de dados de teste fica com o restante 11%, constituindo o alvo da previsão. As últimas 24 observações correspondem portanto aos valores da série para os anos de 2007 e 2008.

A arquitectura é definida por uma rede neuronal *feed-forward*, com três camadas, sendo o número de neurónios da camada de entrada definido pela variável p , optou-se por um número variável de neurónios na camada escondida, e escolhido em função da *performance* do modelo. Após algumas experiências fixou-se em 10 neurónios. A camada de saída consiste em apenas 1 neurónio com o resultado da previsão de um determinado mês. As funções de activação são *logsig*⁷ e *purelin*⁸, respectivamente para a camada escondida e para a de saída.

O algoritmo de treino utilizado foi o *Resilient Backpropagation*⁹, baseado na retro-propagação do erro (Riedmiller e Braun, 1993). Este algoritmo conhecido como Rpro, baseia-se no sinal do gradiente

⁵ Perda de capacidade de generalização aquando da classificação do conjunto de teste, devido ao facto de estar demasiado adaptado ao conjunto de treino.

⁶ Função *rand* da *toolbox* do Matlab.

⁷ Função de transferência sigmoidal logarítmica.

⁸ Função de transferência linear.

⁹ *Toolbox*, *Neural Networks* do Matlab.

descendente da função *performance*, em vez do próprio gradiente descendente, como é habitual. Este aspecto permite que continue a haver uma evolução no ajuste dos parâmetros da rede em zonas em que o gradiente da função tenha uma amplitude muito baixa, como sucede nas extremidades das funções sigmoidais usadas na camada escondida. Este algoritmo tem sido utilizado por vários autores (Fernandes et al., 2008) e foi a solução com melhor *performance* de entre uma selecção de algoritmos de treino experimentados (*Levenberg-Marquardt* e *Gradient Descent with Momentum*), também estes variantes do algoritmo *Backpropagation*.

O processo de treino da rede é um processo iterativo que é interrompido por um dos critérios de paragem: atingir o erro mínimo desejado, atingir o número de iterações de treino ou validação cruzada. No presente estudo foi utilizada a função de *performance* MSE, que mede o erro quadrático médio e interrompe o processo sempre que é atingido o valor desejado. Utilizou-se um conjunto de validação, já descrito, para paragem antecipada com a finalidade de evitar a perda de generalização por parte da rede, aquando da classificação dos dados de teste. O MSE foi inicializado com um valor de 0.002, o que significa que o critério de paragem accionado será principalmente baseado no conjunto de validação. Refira-se que o conjunto de teste não é usado no processo de treino.

O processo de treino foi repetido várias vezes, com diferentes inicializações dos pesos, uma vez que diferentes valores iniciais dos parâmetros da rede levam a diferentes mínimos locais. Das várias repetições foi seleccionada a rede com menor erro relativo (Erro Percentual Absoluto Médio - EPAM), apresentado na equação 2.

5. Análise e discussão dos resultados

De forma a validar a qualidade de ajustamento dos resultados obtidos foi determinado o erro

percentual absoluto médio definido como:

EPAM=1/n∑i=1n|Yi′−Yi|/Yi [2]

Onde, *Y_i'* representa o conjunto de valores previstos e *Y_i* os valores reais.

No presente caso, o conjunto de teste contém *n*=24, o erro obtido foi de 4.16% e 4.13% para as metodologias baseadas na regressão linear e redes neuronais artificiais respectivamente (Quadro 2).

Ainda, tendo por base o quadro 2, verifica-se que o erro foi mais acentuado em amostras que não seguem o padrão da maioria, como é o caso das variações identificadas nos meses de Março, Abril e Maio de 2008, que registam um erro absoluto superior a 7%. Este erro deve-se ao comportamento irregular, de difícil modelação, da série para estes meses.

De forma geral, os resultados não mostram uma grande distinção no desempenho entre os modelos estudados. Como se pode observar na Figura 9, ambos os modelos acompanham o comportamento da série original, sendo neste caso, o modelo baseado em redes neuronais artificiais a conseguir, mesmo que de forma ligeira, a melhor aproximação.

Quadro 2 | Erro percentual mensal e EPAM no intervalo [Jan-2007:Dez-2008] para ambos os modelos, regressão linear (RL) e redes neuronais artificiais (RNA)

Meses	2007		2008		EPAM	
	RL	RNA	RL	RNA	RL	RNA
Jan	1.94%	1.60%	2.94%	2.51%	4.16%	4.13%
Fev	3.43%	3.90%	4.16%	9.12%		
Mar	1.53%	4.12%	7.81%	4.05%		
Abr	1.71%	0.69%	8.27%	9.64%		
Mai	1.88%	3.58%	5.20%	7.30%		
Jun	1.92%	2.66%	2.56%	0.78%		
Jul	6.04%	2.38%	2.55%	8.07%		
Ago	5.58%	2.73%	3.05%	1.00%		
Set	6.11%	0.32%	2.40%	6.65%		
Out	3.98%	0.47%	1.61%	6.38%		
Nov	8.07%	1.67%	3.82%	7.90%		
Dez	6.62%	7.82%	6.73%	3.78%		

Fonte: Elaboração própria.

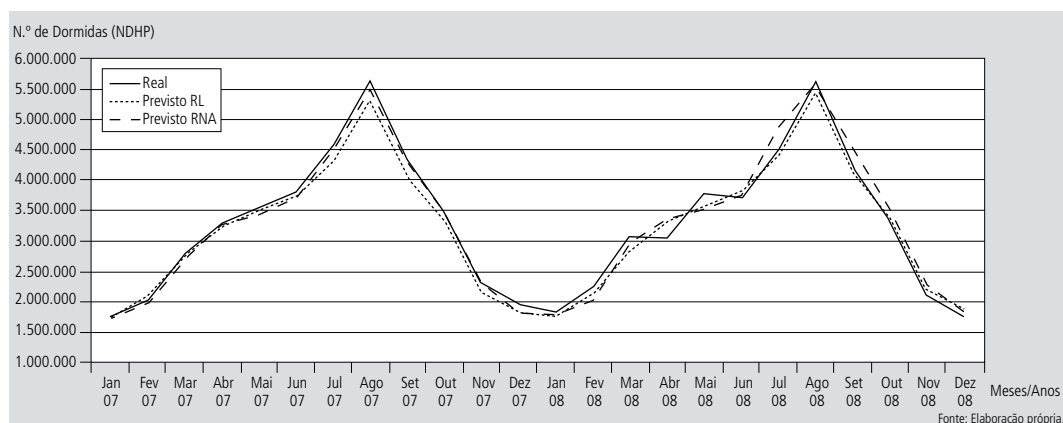


Figura 9 | Representação da série e a respectiva previsão para os anos de 2007 e 2008, pelos modelos RL e RNA.

Projectando os resultados da previsão em todo o domínio da série consegue ter-se uma ideia mais abrangente dos resultados obtidos por ambos os métodos (Figuras 10 e 11). Refira-se que na Figura 11 só se apresentam os valores estimados pela RNA a partir de 1991, uma vez que a rede necessita dos valores do ano anterior na sua entrada.

Como se pode observar, na Figura 10, o erro é mais acentuado no ano de 1994, com o modelo RL, com um EPAM para esse ano de 12.2%, sendo o período que registou a estimativa menos satisfatória.

Observa-se também na Figura 11, que o modelo baseado em RNA, acompanha de perto o comportamento da série, nomeadamente, para os

anos de 1993/94 e 2000/01/02, onde a tendência da série se inverteu ligeiramente. Isto deve-se fundamentalmente ao facto de se estar perante uma modelação sobre dados de treino, e portanto a aproximação é tanto melhor quanto maior o número de iterações do algoritmo de treino. No limite, se o processo de treino não fosse interrompido, a aproximação seria ainda maior.

Também se pode verificar da mesma observação, que existem meses que seguem uma tendência mais próxima do linear, como é o caso do mês de Janeiro, e outros que descrevem um comportamento próximo do linear. Março é o mês que ao longo de todo o intervalo, apresenta um maior desvio em relação à tendência calculada.

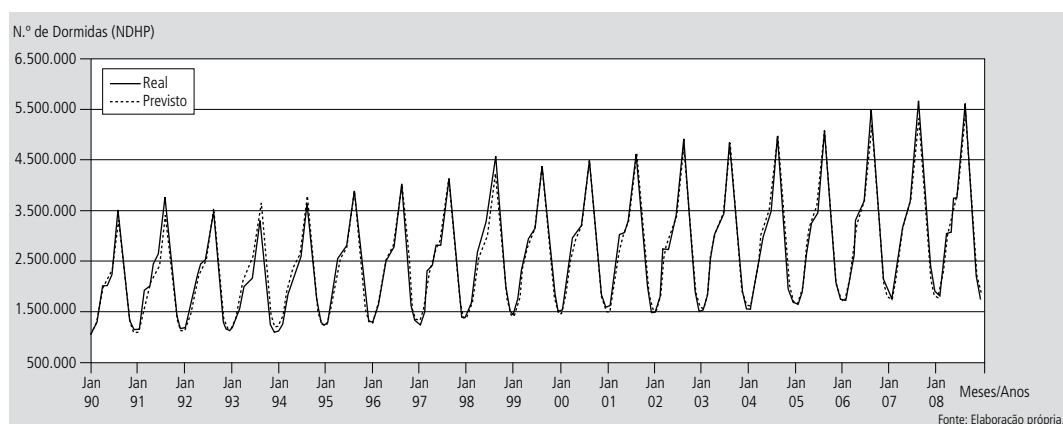


Figura 10 | Valores previstos sobre a totalidade da série com o modelo RL.

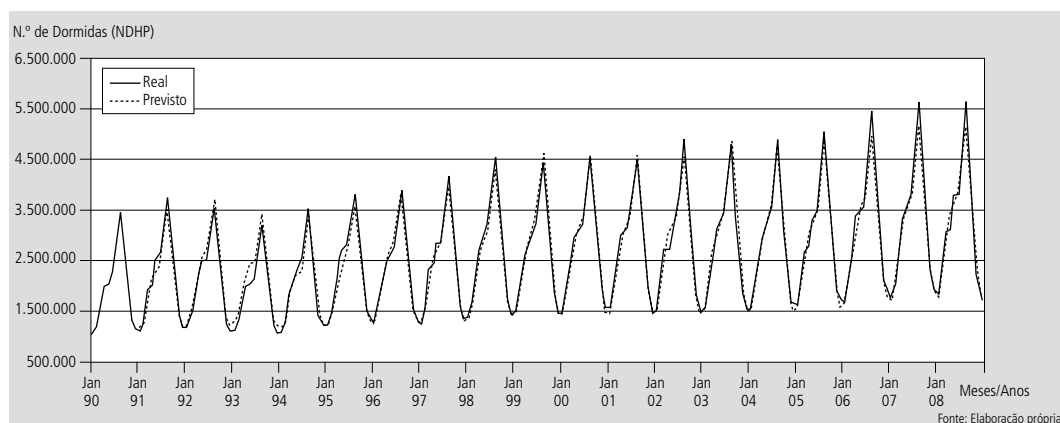


Figura 11 | Valores previstos sobre a totalidade da série com o modelo RNA.

Os valores obtidos correspondem a um resultado satisfatório em ambos os modelos. Atendendo à classificação do MAPE¹⁰, ou EPAM, proposto por (Lewis, 1982), pode mesmo dizer-se que ambos os modelos produziram previsões de precisão elevada, uma vez que para o EPAM se obteve um valor inferior a 10%.

6. Conclusões e desenvolvimentos futuros

A série temporal aqui estudada apresenta um comportamento linear quando analisado por mês. Assim, pode concluir-se com este estudo, que os métodos de modelação linear, mais concretamente a regressão linear, permitem estimar com uma precisão bastante satisfatória, a procura turística de estabelecimentos hoteleiros em Portugal.

O facto de se ter analisado a evolução de cada mês de forma isolada, permite lidar com a especificidade característica de cada mês e assim aumentar a qualidade da estimativa.

No entanto, o modelo baseado na regressão linear apresenta alguns problemas característicos dos modelos lineares, nomeadamente a impossibilidade de estimar determinados padrões que seguem

um comportamento menos determinístico, ou simplesmente não linear. Este problema costuma ser resolvido com recurso a métodos não lineares, como as redes neuronais.

Convém também referir que os resultados obtidos através da metodologia de regressão linear, seriam bem diferentes, se a série descrevesse um comportamento menos linear, como é caso das séries da procura turística regional (Fernandes e Teixeira, 2008).

Atendendo também ao conceito de eficiência, pode concluir-se deste estudo que o modelo de regressão linear é mais eficiente que o modelo de redes neuronais, uma vez que a sua aplicação envolve um número de variáveis reduzido e um esforço computacional inferior, quando comparado com os parâmetros necessários para o modelo baseado em redes neuronais artificiais.

A utilização do modelo RNA permitirá a inclusão de outras variáveis nesta série ou outras séries de procura turística por regiões (com menores linearidades que a série apresentada), que possam ser causa de algumas variações além do crescimento registado pela tendência. Compreende-se que

¹⁰ Do inglês *Mean Absolute Percentage Error*.

factores, macroeconómicos, como o investimento no turismo de determinadas regiões ou do próprio país, bem como o crescimento económico, a taxa de desemprego, o índice de preços ao consumidor, do próprio país (neste caso Portugal), bem como do país de origem dos turistas que se deslocam a Portugal, podem influenciar significativamente a procura turística. Assim, a motivação para um modelo baseado em RNA reside nesta possibilidade adicional, de se poderem incluir estes ou outros factores como variáveis de entradas, tornando o modelo mais robusto.

Atendendo à filosofia por detrás da metodologia de redes neuronais, pode realçar-se ainda a ideia de que estas têm mostrado ser uma alternativa bastante viável na área da modelação de séries temporais. A sua capacidade para modelar padrões complexos é mais um trunfo no campo da ciência previsional.

Referências

- Basheer, I.A., Hajmeer, M., 2000, Artificial Neural Networks: fundamentals, computing, design and application, *Journal of Microbiological Methods*, Vol. 43, pp. 3-31.
- Butler, R.W., 2001, *Seasonality in Tourism: Issues and Implications*, Edited by Toum Baum, Pergamon, Oxford, UK.
- Demuth, H., Beale, M., Hagan, M., 2008, *Neural Network, Toolbox™ 6: User's Guide*, United States.
- EUROSTAT, *Statistical Office of the European Communities*, [<http://epp.eurostat.ec.europa.eu/>], (Site accessed 10 March 2008).
- Fernandes, P.O., Teixeira, J.P., 2008, Modelação da Procura Turística: Um estudo comparativo entre Redes Neuronais Artificiais e a Metodologia de Box-Jenkins, *Actas do XII Congreso Internacional de Investigación en Ciencias Administrativas (ACACIA)*, Universidad de Baja California, México.
- Fernandes, P.O., Teixeira, J.P., Ferreira, J.M., Azevedo, S.G., 2008, Modelling Tourism Demand: A Comparative Study between Artificial Neural Networks and the Box-Jenkins Methodology, *Romanian Journal of Economic Forecasting*, Vol. 5(3), pp. 30-50.
- Fernandes, P.O., 2005, *Modelling, Prediction and Behaviour Analysis of Tourism Demand in the North of Portugal*, PhD Thesis, University of Valladolid, Valladolid, Spain.
- Fúster, L.F., 1989, *Introducción a la Teoría y Técnica del Turismo*, Alianza, Madrid.
- Gujarati, D.N., 2003, *Basic Econometrics*, 4th Edition, McGraw-Hill.
- Hansen, J.V., McDonald, J.B., Nelson, R.D., 1999, Time series prediction with genetic-algorithm designed neural networks: an empirical comparison with modern statistical model, *ComputIntell*, Vol. 15, pp. 171-184.
- Haykin, S., 1999, *Neural Networks: A comprehensive Foundation, Second Edition*, Upper Saddle River, Prentice-Hall, NJ.
- Hill, T., O'Connor, M., Remus, W., 1996, Neural network models for time series forecasts, *Management Science*, Vol. 42(7), pp. 1082-1092.
- Johnston, J., Dinardo, J., 2000, *Métodos Económicos*, 4ª Edição, McGraw-Hill.
- Lewis, C.D., 1982, *Industrial and Business Forecasting Method*, Butterworth Scientific, London, UK.
- Machado, T., Teixeira, J., Fernandes, P., 2009, Modelação Linear da Procura Turística em Portugal, *Actas do XXXIII Congresso Internacional de Economia Aplicada, ASEPELT'2009*, Covilhã, Portugal.
- Riedmiller, M., Braun, H., 1993, A direct adaptive method for faster backpropagation learning: The RPROP algorithm, *Proceedings of the IEEE International Conference on Neural Networks*.
- Rumelhard, D.E., McClelland, J.L., 1986, *Parallel Distributed Processing - Explorations in the Microstructure of Cognition*, Vol. 1, Foundations, The Massachusetts Institute of Technology Press.
- Thawornwong, S., Enke, D., 2004, The adaptive selection of financial and economic variables for use with artificial neural networks, *Neurocomputing*, Vol. 6, pp. 205-232.
- Tsaur, S.-H., Chiu, Y.-C., Huang, C.-H., 2002, Determinants of guest loyalty to international tourist hotels-a neural network approach, *Tourism Management*, Vol. 23, pp. 397-405.
- Yu, G., Schwartz, Z., 2006, Forecasting Short Time-Series Tourism Demand with Artificial Intelligence Models, *Journal of Travel Research*, Vol. 45, pp. 194-203.