

O Impacto da Variável Páscoa na Previsão da Procura Turística

Paula Odete Fernandes; Professora Adjunta (pof@ipb.pt)

João Paulo Teixeira; Professor Adjunto (joaopt@ipb.pt)

Instituto Politécnico de Bragança

João José Ferreira; Professor Auxiliar (jjmf@ubi.pt)

Susana Garrido Azevedo; Professora Auxiliar (sazevedo@ubi.pt)

Universidade da Beira Interior

Resumo

Com este estudo pretende-se analisar o impacto do feriado móvel da Páscoa na previsão da procura turística, para as regiões Norte e Centro de Portugal. De salientar que a série temporal “Dormidas Mensais registadas nos estabelecimentos hoteleiros”, considerada como significativa da actividade turística, devido às suas características, denota que os fenómenos influenciam de forma distinta a procura turística, nas regiões em estudo.

Assim, tendo por base modelos não lineares, sustentados pela metodologia das Redes Neurais Artificiais (RNA), vai-se verificar se os resultados sofreram alterações significativas antes e após a utilização da variável *dummy* Páscoa. A inclusão desta nova variável no modelo prende-se com o facto de se ter detectado, em estudos anteriores, alguns valores atípicos na série temporal, dormidas mensais nos estabelecimentos hoteleiros nas regiões em estudo, pelo que se tenciona captar esse efeito.

Palavras-chave: Procura Turística; Redes Neurais Artificiais; Modelação Não Linear; Previsão.

Área temática: 12. Modelação Económica e Financeira.

Abstract

This study aims to analyze the impact of the mobile holiday of Easter on the tourist demand' forecast, for the North and Centre regions of Portugal. To point out that the time series “Monthly Guest Nights in Hotels”, considered as significant of the tourist activity due to its characteristics, it denotes that the phenomena influence in a different way the tourist demand, in the focused regions.

Thus, based on non-linear models, supported by the Artificial Neural Networks methodology, we intent to compare the results before and after the use of the dummy variable Easter. The inclusion of this new variable in the model is due to the fact that in previous studies on the same region, some atypical values in the time series “Monthly Guest Nights in Hotels” were detected. Being so, this paper intends also to catch this effect.

Key words: Tourism Demand; Artificial Neuronal Network; Non-Linear Modelling; Forecasting.

Thematic Area: 12. Economic and Financial Modelling.

1. INTRODUÇÃO

O estudo da procura turística tem despertado o interesse da comunidade científica e empresarial pela sua importância para a economia mundial. Esta temática foi alvo da atenção da comunidade científica ainda na década de 70 do século XX no trabalho de Archer (1976). Duas décadas mais tarde Theobald (1998) apresentou o turismo como o fenómeno do século XX. A World Tourism Organisation (2000) também descreve o turismo como um dos fenómenos económicos e sociais mais marcantes do último século. Efectivamente, o turismo nas últimas décadas tem crescido consideravelmente no contexto internacional, pois este tem sido reconhecido como o maior bem/serviço transaccionado no comércio internacional (Goeldner, *et al.*, 1999).

Numa perspectiva macro económica este sector tem emergido como um dos mais dinâmicos e líderes das economias de muitos países. Este foi mesmo considerado na década de 90 como a principal fonte de receitas de economias como a Italiana e a Espanhola (Goeldner *et al.*, 1999) e como o principal sector empregador nas economias do Canadá, Japão, Reino Unido (Hawkins *et al.*, 1992). Ao nível local, o turismo também fornece oportunidades económicas para as comunidades locais beneficiarem de negócios complementares a esta actividade.

Para capitalização de uma forma bem sucedida dos benefícios e minimização dos potenciais impactos negativos do turismo, os decisores associados a este sector devem ter um conhecimento mais profundo das mudanças e desenvolvimentos que afectam as tendências no turismo. Frenchling (1996), Faulkner e Valério (2000) e Gee e Fayos-Solá (1997) realçam que mais do que em qualquer outro sector de actividade, no turismo deve haver a preocupação com a redução do risco associado a tomadas de decisão. Logo, as actividades de monitorização e previsão da procura são particularmente importantes na gestão turística devido sobretudo ao carácter perecível do produto/serviço turístico e ao facto de a fase de produção e consumo do produto/serviço turístico se verificarem no mesmo momento.

De forma a ir ao encontro destas características únicas dos produto/serviço turísticos, uma resposta rápida deve ser dada pelos decisores mediante a observação dos níveis de oferta e procura para que consigam reagir em sintonia com a situação do mercado. De facto, uma previsão efectiva fornece informação credível e atempada para que os gestores turísticos possam fazer esse balanceamento entre a procura do mercado e a oferta turística. A previsão da procura turística apresenta-se assim estratégica ao permitir obter informação importante para que os decisores turísticos consigam antecipar potenciais crescimentos e riscos num futuro próximo para um país, região ou economia específica (Hu, 2002).

Com o presente estudo pretende-se assim explorar o campo de estudo da previsão da procura turística e evidenciar a utilidade da metodologia das Redes Neurais Artificiais (RNA) no estudo específico do impacto do feriado móvel da Páscoa na previsão da procura turística, para as regiões Norte e Centro de Portugal.

A presente investigação está estruturada da seguinte forma: depois da presente introdução, são apresentadas, dentro da metodologia, as variáveis utilizadas neste estudo bem como as fontes onde estas foram recolhidas; seguidamente é feita a descrição do Modelo das Redes Neurais Artificiais. Finalmente após a apresentação e discussão dos resultados obtidos algumas considerações finais são tecidas.

2. METODOLOGIA

2.1. Apresentação das Variáveis

A informação de suporte ao trabalho foi recolhida junto do Instituto Nacional de Estatísticas de Portugal (INE). Optou-se por considerar as séries Dormidas Mensais Registadas na Região Norte [DRN] e as Dormidas Mensais Registadas na Região Centro [DRC], por dois motivos: (i) por ser uma série que mede a procura turística efectiva; (ii) para comparar com resultados de estudos anteriores, produzidos pelos mesmos autores.

Assim, os dados recolhidos reportam-se ao período compreendido entre Janeiro de 1987 e Dezembro de 2006, correspondendo assim a um total de 240 observações mensais ao longo de 20 anos (ver Anexo A, Tabelas A.1 e A.2).

Pela análise à Figura 1, esta sugere a existência de oscilações irregulares indicando uma não estabilização da média e variância e a presença de sazonalidade (valores máximos nos meses de Verão e mínimos nos meses de Inverno). Ainda, para ambas as séries verifica-se:

- uma tendência crescente nos dois últimos anos, registando-se uma variação positiva, de 2005 para 2006, de 11,8% e de 6,4%, para a região Norte e para a região Centro, respectivamente;
- entre os meses de Março e Maio um pico intermédio que pode estar relacionado com a Páscoa que tradicionalmente atrai muitos turistas para as regiões.

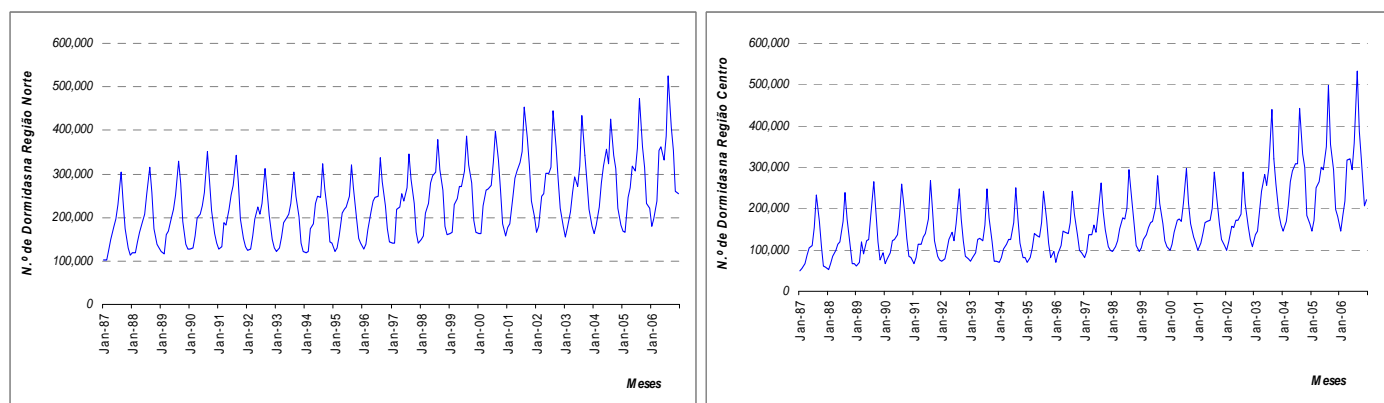


Figura 1. Cronograma das séries, para o período de 1987:01 a 2006:12.

O crescimento registado em 2005 e 2006 para valores nunca antes atingidos acarreta uma dificuldade muito grande para os modelos de previsão que se baseiam apenas no passado, como são os modelos aqui usados, uma vez que o modelo terá de prever um valor que nunca aconteceu.

Para tentar atenuar os efeitos nefastos para a previsão com RNA observados nas séries e no sentido de estabilizar a variância das séries, estas transformaram-se aplicando, a cada uma, o logaritmo natural: LRN e LRC, respectivamente para a região Norte e para a região Centro (Figura 2).

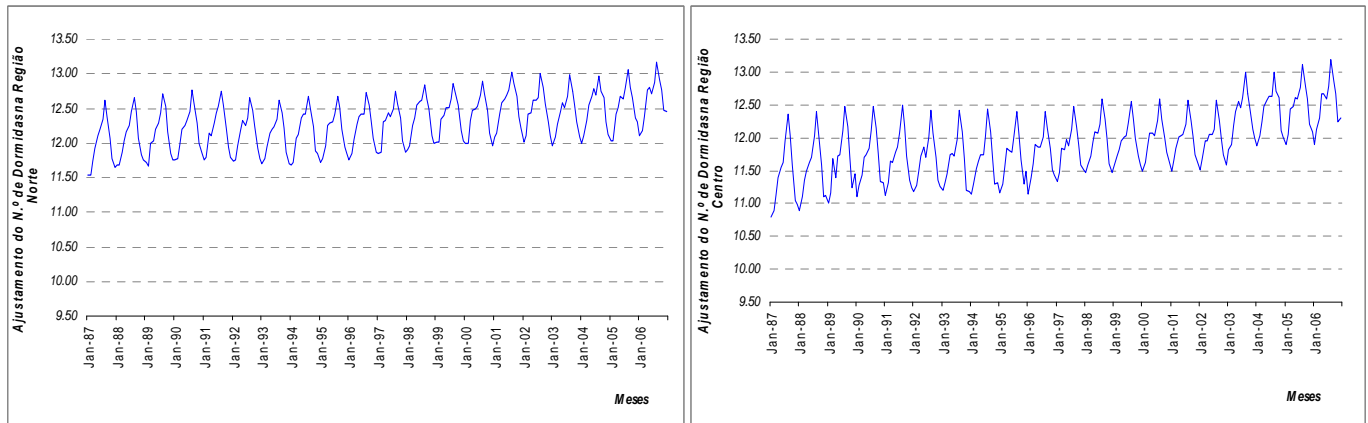


Figura 2. Cronograma das séries transformadas para o período de 1987:01 a 2006:12.

Pela análise da figura anterior pode observar-se que as séries continuam com uma certa estacionaridade que consiste num movimento cíclico, quase periódico, em torno de um valor médio crescente. Contudo, no domínio logarítmico as variações impostas pelo crescimento dos últimos anos aparecem agora menos pronunciadas.

Como se pode observar os resultados tornam-se mais satisfatórios para a região Norte [LRN] do que para a região Centro [LRC]. Assim e no sentido de atenuar o efeito da sazonalidade, ainda patente, foram também processados os dados de entrada pela aplicação das diferenças simples e sazonal, de acordo com a equação às diferenças [1], doravante designada de [LLRC], série ‘limpa’ (Fernandes & Teixeira, 2008).

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

Onde:

$x(n)$ é a série a limpar, já no domínio logarítmico;

$y(n)$ é a série limpa.

De referir que para a série LLRC, os valores previstos com a aplicação da metodologia que se propõe neste estudo, Redes Neurais Artificiais (RNA), são posteriormente processados para inverter o processo de aplicação das diferenças simples e sazonais.

Assim, as novas séries que serviram de base a todo o estudo foram: LRN no domínio logarítmico e LLRC no domínio logarítmico com a aplicação de diferenças, uma simples e outra sazonal.

Ainda, tal como já foi referido, a análise gráfica (Figuras 1 e 2) das séries sugere alguns valores atípicos, ao longo das séries, que podem estar relacionados com época festiva da Páscoa que tradicionalmente gera em simultâneo um fluxo de turismo religioso e um mini período de férias associado às férias escolares. Tenciona-se captar esse efeito, com a inclusão no modelo de uma variável *dummy* [DP_t]. Como se conhece, este fluxo turístico e o período de mini férias ocorre habitualmente na semana anterior ao dia de Páscoa (semana Santa), pelo que se procurou que esta nova variável reflectisse essa realidade usando um peso proporcional aos dias dessa semana em cada mês (normalmente entre os meses de Março e Abril). Assim, de acordo com o Dia da Páscoa (DP) terá o valor apresentado na Tabela 1.

Tabela 1. Valor de DP.

<i>Dia de Páscoa</i>	<i>Valor de DP no mês da</i>	<i>Valor de DP no mês</i>
----------------------	------------------------------	---------------------------

	<i>Páscoa</i>	<i>anterior</i>
1	0.1	0.9
2	0.3	0.7
3	0.5	0.5
4	0.6	0.4
5	0.8	0.2
6	0.9	0.1
>6	1	0

2.2. Redes Neurais Artificiais

A modelação das séries em estudo vai ter como metodologia de base a utilização das Redes Neurais Artificiais (RNA). As RNA são modelos que se destacam dentro de uma ampla área do conhecimento da inteligência artificial e têm como suporte modelos matemáticos de arquitectura semelhante à do cérebro humano.

Uma rede neuronal encontra-se composta por um conjunto de neurónios artificiais, nós ligados entre si - as conexões entre os nós 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; Cortez, 2002; Fernandes, 2005). Os neurónios de uma rede encontram-se estruturados em distintas camadas (mais conhecidas por camada de entrada, intermédia ou oculta e de saída), sendo as mais comuns, para a previsão de séries temporais, as multi-camadas ou também designadas por MLP¹ (Bishop, 1995; Thawornwong & Enke, 2004), por forma que um neurónio de uma camada encontra-se ligado com os neurónios da camada seguinte às que pode enviar informação, Figura 3. As redes, dependendo da forma como estão ligadas entre as diferentes camadas podem classificar-se em redes não recorrentes² (*feedforward*) e redes recorrentes³ (*feedback*) (Faraway & Chatfield, 1998; Fernandes, 2005).

¹ Do inglês *Multi-Layer Perceptron*.

² 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).

³ As conexões permitem que a informação regresse a lugares por onde já passou e admite-se também ligações intra-camada (laterais) (Fernandes, 2005).

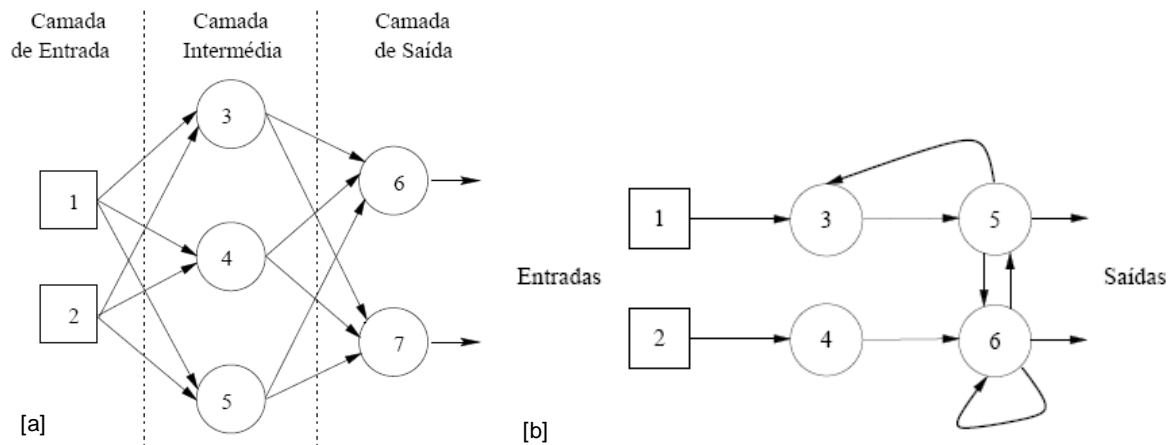


Figura 3. Estrutura de uma Rede Neuronal Artificial: [a] *Feedforward*; [b] *Feedback* (Cortez, 2002).

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 arquitetura, 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).

Assim, para as RNA aprenderem com a experiência têm que ser submetidas a um processo designado de treino, existindo diferentes algoritmos de treino, sendo um dos mais populares e utilizados na previsão de séries temporais o algoritmo *Backpropagation*⁴ ou seus variantes, distribuindo-se os mesmos por duas classes (Gallant, 1993; Haykin, 1999; Fernandes, 2005):

- (i) supervisionados, onde no processo de treino existe um ‘professor’ que fornece um conjunto de casos de treino, e 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;
- (ii) não supervisionados, onde existe um conjunto de entradas, pelo que o algoritmo de treino tenta agrupar os dados a partir de padrões por estes apresentados, seguindo assim uma regra de auto-organização.

Resumidamente um valor produzido por uma rede *feedforward*, com uma camada oculta, pode ser expressa 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 sigmoidal;

⁴ 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).

$\{\alpha_j, j = 0, 1, \dots, n\}$, 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.

2.3. Modelo de Redes Neurais Artificiais

Para este estudo seleccionou-se um modelo de redes neuronais artificiais, já utilizado em estudos anteriores pelos mesmos autores (Fernandes *et al.*, 2008), sendo o mesmo do tipo multi-camada, onde se utilizaram três camadas: de entrada, oculta e de saída, e com uma estrutura tipo *feedforward*. Na camada oculta foi utilizada a função de activação sigmoidal logística [*Logsig*] e na camada de saída utilizou-se a função de activação linear. No treino da rede utilizou-se o algoritmo *Resilient Backpropagation*, uma variante do algoritmo de treino *backpropagation*. A rede, utilizada neste estudo, para cada uma das séries, tem a seguinte arquitectura: 12 nós na camada de entrada, correspondentes aos últimos 12 valores da série, 4 nós na camada oculta e 1 na camada de saída, correspondente à previsão do valor do mês seguinte, ou seja, (1-12;4;1). A estimação/previsão produzida foi efectuada mensalmente. Contudo, para a previsão das observações relativas aos meses de 2005 e 2006 apenas se usaram valores conhecidos até Dezembro de 2004. Para as 12 entradas do modelo correspondentes aos 12 meses anteriores àquele para o qual se está a efectuar a previsão, posteriores a Dezembro de 2004, foram usados os valores previstos em iterações anteriores. Desta forma, a previsão para 2005 e 2006 é realizada num processo iterativo, usando em cada iteração valores previstos em iterações anteriores. O processo de treino para actualização dos pesos fez-se utilizando o método *batch training*. Foi usado um conjunto de validação para interromper as iterações de aprendizagem quando a *performance* nesse conjunto não melhorava ao fim de 5 iterações sucessivas. Para cada rede realizaram-se 500 sessões de treino, seleccionando os resultados da melhor sessão de treino e escolhendo a RNA com melhores resultados no conjunto de validação. A realização de várias sessões de treinos justifica-se porque em cada sessão de treino os valores iniciais dos pesos são diferentes, chegando-se a soluções também diferentes, podendo ter *performances* significativamente diferentes. O critério que esteve na base para escolher o melhor modelo, para cada uma das séries em análise, foi a raiz do erro quadrático médio

(REQM⁵) na comparação dos resultados obtidos pela rede com os valores observados, para o conjunto de validação.

De salientar que para ambas as séries, foi realizado um pré-processamento dos dados de entrada e dos dados de saída correspondente apenas a uma normalização entre -1 e 1.

As séries temporais, obtidas no domínio logarítmico, foram divididas em três conjuntos distintos:

1º - conjunto de teste, para ambas as séries, utilizaram-se as últimas 24 observações, correspondentes aos anos de 2005 e 2006;

2º - conjunto de validação, para a série LRN as 12 observações, correspondentes ao ano de 2004; para a série LLRC as observações: Janeiro de 1999, Fevereiro de 2004, Março de 2002, Abril de 1996, Maio de 2003, Junho de 2000, Julho de 1998, Agosto de 2004, Setembro de 1997, Outubro de 2001, Novembro de 1994 e Dezembro de 2003;

Por observação da série LLRC da Figura 2, verifica-se, apesar da passagem para o domínio logarítmico, que os valores dos últimos 4 ciclos, correspondentes aos anos de 2003 a 2006, assumem valores que se situam a um nível mais elevado. Sendo um dos preceitos para a utilização das RNA que os conjuntos de treino, validação e teste sejam uma boa representação do fenómeno a prever, estaríamos a desviar-nos muito deste preceito se os conjuntos de validação fossem as observações correspondentes ao ano de 2004, já que o teste têm obrigatoriamente que ser 2005 e 2006, restando para o treino as observações até Dezembro de 2003. Desta forma o conjunto de treino só disporia de 1 ano com os valores elevados e cerca de 15 anos com valores bastante inferiores. Por outro lado, o conjunto de validação só disporia de observações correspondentes aos anos com valores mais elevados. Com a distribuição do conjunto de validação usando os 12 meses do ano distribuídos por vários anos, estamos a usar mais observações com valores elevados no conjunto de treino equilibrando melhor estes dois conjuntos.

O conjunto de validação é usado como uma técnica alternativa de paragem do treino da rede para evitar uma sobre adequação ao conjunto de treino - '*overfitting*';

3º - conjunto de treino:

- (i) Para a série LRN corresponde aos meses de Janeiro de 1988 a Dezembro de 2003, num total de 192 observações. Note-se que a primeira observação que este modelo pode usar é a de Janeiro de 1988, porque necessita das 12 observações anteriores na sua entrada;
- (ii) Para LLRC são usadas 179 observações correspondentes aos meses entre Fevereiro de 1989 e Dezembro de 2004 exceptuando os meses do conjunto de validação. Note-se que esta série dispõe de menos 13 observações que a anterior. Estas 13 observações são 'consumidas' no processo de limpar a série descrito na equação [1], uma vez que n começa em 14 para que $x(n-13)$ seja o primeiro mês (Janeiro de 1987).

5

$$REQM = \sqrt{\frac{\sum_{t=1}^n (A_t - P_t)^2}{n}}; \text{ com } : A_t, \text{ valores reais no período } t; P_t, \text{ valores previstos no período } t; n, \text{ total de observações utilizadas.}$$

3. APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Nesta secção pretende-se verificar se os resultados sofrem alterações antes e após a inclusão da variável *dummy* Páscoa [DP_i] no modelo, para ambas as regiões em estudo.

Para tal vão ser analisados os resultados para o conjunto de teste (anos 2005 e 2006), comparando os valores observados com os valores previstos, para as duas séries e utilizando a metodologia das redes neuronais artificiais, e não esquecendo o processo inverso das transformações efectuadas para produzir os valores previstos (tabela 2).

Como critério de análise da qualidade de ajustamento dos valores previstos para cada uma das situações teve-se por base a classificação do MAPE⁶, ou EPAM, proposto por Lewis (1982), dado pela equação:

$$EPAM = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - P_t}{Y_t} \right|; Y_t, \text{valor observado e } P_t, \text{valor previsto.} \quad [3]$$

Tabela 2. Avaliação da *Performance* com os modelos das RNA.

Destino Turístico	Conjunto de Teste	<i>r</i>		<i>EPAM</i>	
		Sem DP_i	Com DP_i	Sem DP_i	Com DP_i
LRN _{RNA(1-12;4;1)}	2005	0.967	0.979	6.89%	5.30%
	2006	0.962	0.953	8.70%	12.30%
LLRC _{RNA(1-12;4;1)}	2005	0.963	0.979	7.44%	5.27%
	2006	0.956	0.979	8.17%	5.71%

Fazendo uma leitura à Tabela 2, pode observar-se para as duas regiões que os modelos, antes e após a inclusão da variável *dummy*, apresentaram qualidades estatísticas e de ajustamento satisfatórias, evidenciando serem adequados para modelar e prever a procura turística das regiões em estudo. Ainda, observa-se que os modelos conseguem produzir resultados mais satisfatórios no curto prazo, neste caso para um ano, como seria de esperar, uma vez que para prever um mês à frente entra-se em linha de conta com os meses anteriores e na previsão dos meses de 2006, o modelo tem na entrada apenas valores previstos em iterações anteriores para os 12 meses precedentes, que incorporam um desvio em relação à realidade e por conseguinte podem levar a um desvio significativo em relação às previsões para 2006.

Para a região Norte verifica-se, para o ano de 2005, uma melhoria de aproximadamente 2 pontos percentuais quando se utiliza a variável *dummy*, reflectindo-se também essa melhoria quando analisado o coeficiente de correlação. O mesmo já não se verifica para o ano de 2006, em que a inclusão da variável *dummy* penaliza os resultados em quase 4 pontos percentuais. Esta situação pode ter a ver com o facto de terem ocorrido outros factores que influenciam a série e que o modelo não considera nas suas entradas.

Relativamente à região Centro, para ambos os anos, incluir a variável *dummy* no modelo permite produzir melhorias significativas que rondam os 2 pontos percentuais, o que se traduz em resultados satisfatórios.

⁶ Mean Absolute Percentage Error.

Contudo pela análise efectuada anteriormente e tendo por base os critérios estabelecidos por Lewis (1982), poder-se-á dizer que para as regiões em estudo, os modelos conseguiram produzir previsões de precisão elevada, sem a variável *dummy*, uma vez que o EPAM apresenta valores inferiores a 10%. Com a variável *dummy* introduzida no modelo, apenas para 2006 e para a região Norte o modelo conseguiu fazer previsões de boa precisão, uma vez que apresentou um EPAM entre 10 e 20%, tendo obtido previsões de elevada precisão para o ano de 2005, para ambas as regiões.

Na Tabela 3 apresenta-se o valor do erro percentual absoluto (EPA) para os meses da Páscoa nos anos de 2005 e 2006, respectivamente Março e Abril, para os modelos com e sem a variável DP, nas regiões Norte e Centro.

Tabela 3. EPA nos meses da Páscoa.

Conjunto de Teste	Região Norte		Região Centro	
	Sem DP_i	Com DP_i	Sem DP_i	Com DP_i
2005	10.7%	0.6%	9.9%	18.4%
2006	18.9%	23.5%	1.9%	3.2%

Pela análise da Tabela 3 constata-se que a inclusão da variável *dummy* também não trouxe melhorias de previsão para os meses particulares em que se esperaria que esta variável actuasse.

4. CONSIDERAÇÕES FINAIS

O turismo tem sido apresentada como uma actividade geradora de riqueza e cada dia que passa tem assumido um papel significativo na economia de um país, Portugal também não fica alheio a este facto. Deste modo, também as organizações, públicas ou privadas, particularmente ligadas ao sector turístico, implementadas nas regiões em estudo (Regiões Norte e Centro de Portugal) devem preocupar-se em construir mecanismos que lhes permitam obter uma visão antecipada da evolução da procura turística, no sentido de criar condições favoráveis aos visitantes que se deslocam a esses destinos turísticos.

Pretendia-se com esta investigação analisar o impacto que o feriado móvel da Páscoa tem na previsão da procura turística, tendo por base modelos não-lineares, sustentados pelas metodologias das RNA, para as respectivas regiões Norte e Centro de Portugal. Pretendeu-se assim verificar se os resultados sofreram alterações significativas antes e após a utilização da variável *dummy*.

Em conclusão, os resultados apresentados pelo modelo com a variável *dummy* Páscoa, não resultam numa melhoria significativa em relação ao modelo sem esta variável. A pequena variação observada (ver tabela 2) encontra-se dentro de uma pequena variação existente entre a escolha do melhor modelo (diferentes sessões de treino), pelo que a eventual melhoria observada não se pode considerar significativa. Além do mais os resultados para os meses da Páscoa do modelo com e sem a variável *dummy* mostram que a inclusão desta nova variável também não traz melhorias para os meses em que esta intervém. Assim, pode concluir-se a irrelevância desta variável para o modelo.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- Bishop, C. M. (1995). *“Neural Networks for pattern recognition”*. Oxford University Press. Oxford. London.
- Cortez, P. (2002). *“Modelos Inspirados na Natureza para a Previsão de Séries Temporais”*. Tese de Doutoramento, Universidade do Minho, Portugal.
- Faraway, J. e Chatfield, C. (1998). “Time series forecasting with neural networks: a comparative study using the airline data”. *Applied Statistics*. N.º47, pp.231/250.
- Faulkner, B. and Valerio, P. (2000). *“An integrative approach to tourism demand forecasting”*. In C. Ryan and S. Page (Eds), *“Tourism Management: Toward the new millennium”*. Oxford, UK: Pergamon, Elsevier Science, Ltd.
- Fernandes, P. O. (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, P. O. e Teixeira, J. P. (2007). “A new approach to modelling and forecasting monthly overnights in the Northern Region of Portugal”. *Proceedings of the 4th International Finance Conference*. Université de Cergy; Hammamet, Medina, Tunísia.
- Fernandes, P. O. e Teixeira, João P. (2008). “Previsão da Série Temporal Turismo com Redes Neurais Artificiais”; 5º Congresso Internacional Luso-Moçambicano de Engenharia - CLME'2008- *“A Engenharia no Combate à Pobreza, pelo Desenvolvimento e Competitividade”*. Maputo-Moçambique.
- Fernandes, P. O.; Teixeira, J. P.; Ferreira, J. e Azevedo, S. (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.
- Frechtling, D. (2001). *“Forecasting tourism demand: Methods and strategies”*. Oxford, UK: Butterworth-Heinemann.
- Gallant, S. (1993). *“Neuronal Network Learning and Expert Systems”*. MIT Press; Cambridge; USA.
- Gee, C. and Fayos-Solá, E. (1997). *“International Tourism: A Global Perspective”*. Madrid, Spain: World Tourism Organization.
- Goeldner, C., Ritchie, B. and McIntosh, R. (1999). *“Tourism: Principles, practices, philosophies”*. 8th. New York, NY: John Wiley & Sons.
- Hawkins, D. and Ritchie, B. (1992). *“World travel and tourism review: indicators, trends and forecasts”*. Wallingford, Oxon, UK: C.A.B. International.
- Haykin, S. (1999). *“Neural Networks. A comprehensive foundation”*. New Jersey, Prentice Hall.
- Hu, C. (2002). *“Advanced tourism demand forecasting: artificial Neural Network and Box-Jenkins Modeling”*, Ph.D Thesis, Purdue University.
- INE. *“Anuários Estatísticos do Turismo de 1987 e 2006”*. Lisboa.
- Lewis, C.D. (1982). *“Industrial and Business Forecasting Method”*. Butterworth Scientific. London.
- Rodrigues, S. (2000). *“Redes neuronais aplicadas à segmentação e classificação de leucócitos em imagens”*. Dissertação de Mestrado em Engenharia Electrónica e Telecomunicações. Universidade de Aveiro.
- Rumelhart, D. E. e McClelland, J. L. (1986). *“Parallel Distributed Processing: Explorations in the Microstructure of Cognition”*. Volume 1: Foundations. The Massachusetts Institute of Technology Press, Cambridge.

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

Theobald, W. (1998). “*Global tourism*”. 2nd ed. Oxford. UK:Butterworth-Heinemann.

World Tourism Organization. (1978-1981). “*World tourism statistics: Statistiques du tourisme*”. Madrid, Spain.

ANEXO

Tabela A.1. Valores da Série Original, para o período entre 1987:01 e 1996:12, Região Norte (cont.).

ANOS MESES	ANOS									
	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
Janeiro	102.447	118.011	122.217	126.671	126.826	124.194	121.469	118.606	122.480	126.910
Fevereiro	102.123	117.547	116.837	129.802	131.653	127.474	129.284	122.988	130.393	139.403
Março	125.401	142.687	160.658	158.701	188.999	157.536	154.734	175.261	156.645	172.393
Abril	150.042	167.118	169.326	197.757	182.290	196.087	189.142	185.525	209.263	213.973
Mai	180.430	189.823	199.158	207.876	219.187	223.918	198.402	232.075	218.666	239.142
Junho	197.113	207.729	218.595	227.159	251.295	207.907	207.216	248.237	222.720	245.264
Julho	229.293	254.523	252.634	257.633	273.927	231.801	231.453	246.274	247.589	248.398
Agosto	304.847	315.113	329.014	351.500	341.490	312.026	304.576	322.366	320.750	336.086
Setembro	238.542	258.287	278.074	284.867	283.378	259.023	249.583	266.094	269.433	280.769
Outubro	173.503	174.359	189.664	216.286	197.241	205.400	202.792	206.256	196.466	225.734
Novembro	130.187	137.933	138.683	162.062	152.554	149.289	141.976	144.803	152.340	175.438
Dezembro	114.229	128.774	127.730	139.683	132.802	130.963	120.748	139.706	140.643	143.163
TOTAL	2.048.157	2.211.904	2.302.590	2.459.997	2.481.642	2.325.618	2.251.375	2.408.191	2.387.388	2.546.673

Fonte: Elaboração própria com base nos dados facultados pelo INE.

Tabela A.1. Valores da Série Original, para o período entre 1997:01 e 2006:12, Região Norte.

ANOS MESES	ANOS									
	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Janeiro	140.430	148.218	163.696	162.389	176.690	165.653	155.527	162.900	168.100	180.700
Fevereiro	141.183	157.415	165.988	162.637	186.586	181.005	177.818	181.900	166.800	195.100
Março	219.465	209.929	228.149	226.010	245.261	249.214	214.106	224.600	247.000	237.200
Abril	224.382	232.767	242.744	262.865	291.395	253.274	258.519	279.800	268.500	352.600
Mai	253.833	280.326	269.854	264.497	306.743	302.028	293.531	317.300	316.900	361.200
Junho	238.334	296.612	270.126	273.881	325.568	301.465	271.454	355.300	307.700	331.500
Julho	266.993	303.866	306.031	324.962	351.955	314.560	318.706	324.400	358.500	388.400
Agosto	345.672	377.645	385.868	397.405	452.581	444.991	433.211	426.900	472.400	524.500
Setembro	288.409	309.700	321.248	331.155	383.793	361.181	343.534	342.100	362.200	406.500
Outubro	232.052	263.522	280.597	263.217	319.417	287.383	281.472	311.500	315.900	353.300
Novembro	166.835	180.796	193.062	186.445	238.925	221.910	219.463	221.200	233.400	258.800
Dezembro	141.349	161.273	166.990	157.210	202.351	179.766	178.439	182.800	221.300	254.700
TOTAL	2.658.937	2.922.069	2.994.353	3.012.673	3.481.265	3.262.430	3.145.780	3.330.700	3.438.700	3.844.500

Fonte: Elaboração própria com base nos dados facultados pelo INE.

Tabela A.2. Valores da Série Original, para o período entre 1987:01 e 1996:12, Região Centro (cont.).

ANOS MESES	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
	Janeiro	48.413	53.251	60.593	66.389	67.712	72.006	73.457	69.142	70.798
Fevereiro	53.932	66.257	70.923	78.898	81.963	78.873	82.466	80.463	81.326	89.418
Março	67.949	84.982	118.949	91.836	114.931	98.200	93.210	101.582	104.727	110.697
Abril	88.730	97.751	88.999	121.039	112.756	124.425	125.441	113.765	139.292	145.682
Mai	103.595	112.881	122.323	125.580	130.316	141.334	127.772	125.687	133.419	142.172
Junho	111.331	120.029	126.325	138.110	140.715	121.020	122.687	125.656	130.530	141.044
Julho	154.594	167.631	182.117	183.161	175.843	163.168	158.791	166.728	164.749	166.283
Agosto	233.117	240.183	263.974	259.879	267.754	247.192	247.527	250.555	242.433	241.940
Setembro	168.602	176.127	190.951	190.030	193.701	175.842	176.980	177.707	171.988	187.513
Outubro	106.730	107.174	118.864	127.891	123.425	121.295	118.980	116.944	116.247	137.972
Novembro	62.249	67.058	75.367	83.646	85.675	84.867	72.739	80.985	80.925	100.324
Dezembro	58.618	67.540	94.352	82.305	76.662	78.134	72.227	81.664	97.189	93.096
TOTAL	1.257.860	1.360.864	1.513.737	1.548.764	1.571.453	1.506.356	1.472.277	1.490.878	1.533.623	1.625.327

Fonte: Elaboração própria com base nos dados facultados pelo INE.

Tabela A.2. Valores da Série Original, para o período entre 1997:01 e 2006:12, Região Centro.

ANOS MESES	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
	Janeiro	82.964	95.078	105.697	97.748	97.835	99.913	136.669	144.481	146.800
Fevereiro	95.439	106.779	123.941	112.210	117.057	118.807	146.512	169.494	172.000	184.500
Março	137.757	122.126	136.214	141.973	138.851	156.803	196.309	206.316	251.100	219.800
Abril	136.194	151.959	155.533	173.166	164.615	154.440	240.487	263.603	264.200	317.200
Mai	159.817	176.390	165.865	173.781	168.582	172.775	282.940	290.185	299.900	320.500
Junho	144.019	173.863	169.182	167.906	171.690	172.701	256.314	308.510	293.000	294.000
Julho	185.696	200.270	203.694	211.569	200.343	185.184	297.678	308.175	348.200	358.000
Agosto	262.815	294.081	280.780	296.264	287.122	288.336	439.293	442.413	496.700	534.200
Setembro	193.321	216.871	214.071	213.978	211.241	211.734	319.576	331.474	353.900	388.300
Outubro	147.357	162.655	161.856	162.932	163.283	158.020	257.783	300.534	294.200	316.100
Novembro	107.827	109.382	122.468	131.786	125.344	125.915	183.431	182.155	198.200	208.200
Dezembro	100.364	96.465	108.546	116.821	110.652	108.691	161.020	163.759	179.200	221.000
TOTAL	1.753.570	1.905.919	1.947.847	2.000.134	1.956.615	1.953.319	2.918.012	3.111.099	3.297.400	3.508.100

Fonte: Elaboração própria com base nos dados facultados pelo INE.