

ESTUDO COMPARATIVO ENTRE OS MODELOS DE REDES NEURONAIAS ARTIFICIAIS E REGRESSÃO LINEAR MÚLTIPLA NA MODELAÇÃO E PREVISÃO DA PROCURA TURÍSTICA EM MOÇAMBIQUE

A COMPARATIVE STUDY BETWEEN ARTIFICIAL NEURAL NETWORK AND MULTIPLE LINEAR REGRESSION TO MODELLING AND FORECAST THE TOURISM DEMAND IN MOZAMBIQUE

Hortêncio Constantino¹, Paula Odete Fernandes² e João Paulo Teixeira³

Resumo

O principal objetivo do presente trabalho assenta no estudo comparativo entre Redes Neuronais Artificiais e o modelo de Regressão Linear Múltipla, para prever a procura turística em Moçambique. Utilizou-se para tal, o número de dormidas mensais registadas nos estabelecimentos hoteleiros, para o período de Janeiro de 2004 a Dezembro de 2013 como variável dependente. Neste contexto foram selecionadas as variáveis explicativas: Índice Harmonizado de Preços ao Consumidor, Produto Interno Bruto e Taxa de Câmbio para os principais mercados emissores: África Sul, Estados Unidos da América, Moçambique, Portugal e Reino Unido. O modelo de redes neuronais artificiais tem uma arquitetura feedforward com uma camada escondida. Experimentaram-se diversas combinações das variáveis de entrada, diferente número de nós na camada escondida, e diferentes funções de ativação. Experimentaram-se ainda diferentes formas de apresentar a variável dependente, nomeadamente na sua forma natural (o seu valor absoluto), na forma exponencial e na forma de diferenças logarítmicas. Das diversas experiências apresentam-se aqui as melhores soluções. Destacam-se, com melhores resultados, a utilização das variáveis de entrada: índice de preços ao consumidor de Moçambique, e as taxas de câmbio do euro, do rand e do dólar face ao metical. Em alguns casos as variáveis do produto interno bruto dos Estados Unidos da América e de Portugal adicionalmente às anteriores melhoram as previsões. A partir dos resultados obtidos pôde-se observar que o modelo de Redes Neuronais Artificiais produziu melhores previsões, ou seja, apresentou uma qualidade estatística de ajuste bastante satisfatória. Assim, o mesmo permitiu efetuar previsões para a procura turística em Moçambique, apresentando um coeficiente de correlação de Pearson de 0,696 e um Erro Percentual Absoluto Médio (MAPE) de 6,5%, num conjunto de teste nunca visto no processo de ajuste/treino do modelo.

PALAVRAS-CHAVE: Modelação, Previsão, Procura Turística, Moçambique, Regressão Linear Múltipla, Redes Neuronais Artificiais.

¹ hcvai@ipb.pt, Instituto Superior Politécnico de Gaza; Instituto Politécnico de Bragança.

² pof@ipb.pt, Instituto Politécnico de Bragança; UNIAG; NECE (UBI).

³ joaopt@ipb.pt, Instituto Politécnico de Bragança; UNIAG.

Abstract

The main objective of this work is based on a comparative study of Artificial Neural Networks (ANN) and Multiple Linear Regression models to modelling and forecast the tourism demand in Mozambique. It was used, for that purpose, the number of registered monthly overnight stays in hotels within the period from January 2004 to December 2013, as the dependent variable. Aiming to reach that objective, the explanatory variables were selected: Harmonized Index of Consumer Prices, Gross Domestic Product and Exchange Rates for the main markets: South Africa, United States of America, Mozambique, Portugal and the United Kingdom. The ANN model has a feedforward architecture with one hidden layer. Several experiments were conducted with the combination of the input variables, different number of nodes in the hidden layer and different activation functions. Also several experiments with different forms of presenting the dependent variable (output), namely the natural form (its absolute value), the exponential form and the form of logarithmic differences. From several experiments the best solutions are presented here. With better results the following input variables are highlight: harmonized index prices of Mozambique, the exchange rates of the euro, rand and dollar to the local metical. Some cases, the gross domestic product of USA and Portugal, added to the previous variables improve the prediction ability. From the results it was observed that the artificial neural networks model produced better results, ie presented a quite satisfactory statistical quality adjustment. Thus, it allowed make forecasts for tourism demand in Mozambique, with a Pearson correlation coefficient of 0,696 and a mean absolute percentage error of 6.5%, in a test set never seen during the adjustment/training process of the model.

KEYWORDS: Modelling, Forecasting, Tourism Demand, Mozambique, Multiple Linear Regression, Artificial Neural Networks.

1. INTRODUÇÃO

A natureza precível do produto turístico faz com que a precisão exata desempenhe um papel fundamental na planificação e previsão de eventos futuros relacionados com a procura de produtos turísticos (e.g., Witt & Witt, 1995; Goh & Law, 2002; Preez & Witt, 2003; Gunter & Önder, 2015). Nesta ordem de ideias, estimativas da procura futura são de grande importância para evitar: empate de capital em mercadorias e em construção de novas unidades hoteleiras; viagem de avião com lugares vagos; hotéis sem clientes; contratação de mão-de-obra sem necessidade ou insuficiente.

Com intuito de medir a procura turística de uma região ou país, vários autores tem recorrido ao número de entradas durante um certo período de referência (e.g., Witt & Witt, 1995; Lim, 1997; Song & Li, 2008; Song, Witt, Fei & Li, 2010). A segunda medida mais usada é o ‘gasto’ efetuado pelos turistas no local de destino ou a ‘receita turística’ resultante da venda de bens e serviços. A outra medida, não menos importante, para medir a procura turística é o ‘número de dormidas de turistas no local de destino registado nos estabelecimentos hoteleiros e similares’ (e.g., Witt & Witt, 1995; Lim, 1997; Santos & Fernandes, 2010; Coshall & Charlesworth, 2011; Cunha & Abrantes, 2013; Claveria & Torra, 2014).

Para o presente estudo, esta variável ‘número de dormidas’ foi selecionada como variável dependente ou explicativa. Esta escolha fundamenta-se pelo fato desta variável ser expressiva quando o assunto em questão é modelar a procura turística, nacional e internacional, em conjunto, ou seja, esta variável tem a vantagem de captar fluxos turísticos de turistas nacionais e estrangeiros (Cunha & Abrantes, 2013). Segundo estes autores, a variável ‘número de dormidas’ engloba dormidas referentes a nacionais e estrangeiros para um certo período em análise o que permite uma avaliação mais criteriosa e realística sobre os movimentos turísticos numa dada região, o que pode ajudar os

gestores turísticos a tomar certas medidas preventivas com relação a épocas de maior ou menor concentração turística (Cunha & Abrantes, 2013).

O principal objetivo deste trabalho passa por construir um modelo econométrico para explicar o comportamento da procura turística em Moçambique representada pelo número de dormidas nos estabelecimentos hoteleiros e similares. Para tal, recorreu-se ao modelo de Redes Neurais Artificiais (ANN) e ao modelo de Regressão Linear Múltipla, (RLM). Para explicar o comportamento da variável dependente foi selecionado um conjunto de variáveis independentes ou explicativas de acordo com a teoria económica (Índice Harmonizado de Preços ao Consumidor, Produto Interno Bruto, e Taxa de Câmbio) referentes aos cinco maiores mercados emissores turísticos de Moçambique, que se presume influenciarem o número de dormidas.

O presente artigo encontra-se estruturado da seguinte forma: após a presente introdução, segue-se no primeiro ponto o enquadramento da temática; no segundo ponto apresenta o desenvolvimento metodológico; no terceiro ponto apresenta-se a parte empírica e a resposta ao principal objetivo do trabalho, ou seja, a criação de um modelo que permita modelar a procura turística em Moçambique; e por fim, apresentam-se as principais conclusões do estudo.

2. ENQUADRAMENTO TEÓRICO

Vários estudos que tratam a previsão e modelação da procura turística têm sido publicados nos últimos anos (e.g., Uysal & Roubi, 1999; Law & Au, 1999; Law, 2000; Song, Witt & Li, 2003; Fernandes, 2005; Li, Song & Witt, 2005; Athanasopoulos & Hyndman, 2008; Dwyer, Forsyth & Dwyer, 2010; Machado, Fernandes & Teixeira, 2007; Fernandes & Teixeira, 2008; Fernandes & Teixeira, 2009; Santos & Fernandes, 2011; Tribe & Xiao, 2011; Song & Witt, 2012; Cankurt & Subasi, 2012; Peng, Song & Crouch, 2014). Os autores apresentaram também uma primeira versão desta comparação em Constantino, Fernandes e Teixeira (2016).

Por exemplo, Uysal e Roubi (1999) introduziram as redes neuronais artificiais (ANN) como alternativa ao modelo de regressão linear múltipla para modelar os gastos dos turistas Canadenses nos Estados Unidos da América a partir do produto interno bruto, índices de preço ao consumidor e taxas de câmbio e quatro variáveis *dummies*. Resultados obtidos permitiram concluir que o modelo ANN modela convenientemente quando comparado com o modelo de Regressão Linear Múltipla RLM.

Law e Au (1999) previram a procura turística no Japão através dos modelo de redes neuronais artificiais (ANN) e os resultados obtidos permitiram concluir que o modelo ANN teve melhores resultados quando comparado com os modelos de Regressão Múltipla, Naïve, Médias Moveis e Alisamento Exponencial. Law (2000) para melhorar a precisão do modelo ANN introduziu a componente de aprendizagem *Back-Propagation* e os resultados obtidos permitiram concluir que o modelo teve melhores resultados quando comparado com os modelos de Regressão Múltipla, Séries Temporais e Rede Neuronal *Feed-forward*.

Burger, Dohnal, Kathrada e Law (2001) introduziram os modelos Naïve, Regressão Múltipla, Redes Neurais, ARIMA, Médias Moveis entre outros para modelar e prever a procura turística em Durban por parte de turistas dos Estados Unidos da América. Resultados previsionais obtidos levaram a concluir que o modelo ANN modela e prevê convenientemente quando comparado com os outros modelos.

Machado, Fernandes e Teixeira (2010) modelaram a procura turística representada pelo número de dormidas na região norte de Portugal para o período de 1990 a 2008 numa base mensal através do modelo de Regressão Linear e Redes Neurais Artificiais. Os resultados obtidos permitiram concluir que a ANN apresenta melhores resultados quando comparado com o modelo de RL.

Cankurt e Subasi (2012) usaram os modelos ANN e RLM para prever a procura turística na Turquia, representada pelo número de chegadas. Os resultados previsionais obtidos permitiram

concluir que o modelo ANN teve melhor desempenho quando comparado com o modelo de RL.

A procura turística para um determinado destino pode definir-se como uma combinação de produtos e serviços turísticos que os consumidores (turistas) estão dispostos a comprar durante um determinado período de tempo específico e sob um dado conjunto de condições (Song & Witt, 2000). Segundo estes autores, as condições que se relacionam com a quantidade de turismo procurado incluem preços de turismo para o destino (custo de vida dos turistas no destino e o custo de viajar para o destino) a disponibilidade de preços turísticos de destinos concorrentes ou substitutos, rendimento de potenciais consumidores, despesas de publicidade, gostos dos consumidores do mercado emissor e outros fatores sociais, culturais, geográficos e políticos.

Seguindo esta linha de raciocínio, e uma vez que se pretende contribuir para a criação de um modelo que melhor se ajuste às séries temporais que permitem modelar a procura turística em Moçambique recorreu-se a modelos causais que têm a vantagem de analisar as relações causais entre uma variável dependente ou explicada e as variáveis ou fatores que a influenciam, ou seja, variáveis explicativas ou dependentes (e.g., Lim & McAleer, 2002; Song & Li, 2008).

3. METODOLOGIA DE INVESTIGAÇÃO

3.1. Recolha e Tratamento dos Dados

Tendo por base o objetivo de estudo utilizaram-se dados mensais relacionados com o número de dormidas registadas nos estabelecimentos hoteleiros em Moçambique. Estes foram utilizados no estudo como variável dependente ou explicada para o período de janeiro de 2004 a Dezembro de 2013. A variável dependente vai ser explicada por um conjunto de variáveis selecionadas de acordo com a revisão da literatura e da teoria económica, nomeadamente: Produto Interno Bruto, Índice Harmonizado de Preços ao Consumidor e Taxa de Câmbio. As referidas variáveis foram selecionadas para os principais cinco mercados emissores turísticos que se presume influenciarem o número de dormidas em Moçambique, nomeadamente: África do Sul, [AS]; Estados Unidos da América, [EUA]; Moçambique, [MOC]; Portugal, [PT]; e Reino Unido, [RU].

De referir que se recorreram às seguintes fontes externas e secundárias para a obtenção dos dados:

- Para os dados referentes a África do Sul acedeu-se ao instituto Statistics South Africa (SSA, 2014);
- Para Moçambique os dados foram recolhidos junto do Instituto Nacional de Estatística de Moçambique (INE, 2004-2014) e no Ministério do Turismo de Moçambique (MITUR, 2004-2014);
- Os dados referentes a Portugal foram recolhidos junto do EUROSTAT;
- Junto do *Office for National Statistics*, recolheram-se os dados do Reino Unido;
- Para os dados referentes aos Estados Unidos da América (EUA) consultou-se o *Federal Reserve Bank*;
- Os dados referentes à taxa de câmbio dos principais cinco mercados emissores turísticos (África do Sul, Estados Unidos da América, Moçambique, Portugal e Reino Unido) foram recolhidos junto do Oanda, (2015).

No presente estudo, todos os valores monetários estão expressos em Meticais, moeda de Moçambique.

Para a materialização do estudo foram selecionados dois modelos de previsão, nomeadamente, o modelo de Redes Neuronais Artificiais e modelo de Regressão Linear Múltipla. Por sua vez, para avaliar os resultados da previsão recorreu-se às medidas de Erro Percentual Absoluto Médio (MAPE) e Coeficiente de Correlação de *Pearson* (r).

3.2. Modelo de Redes Neurais Artificiais

O modelo de redes neuronais artificiais (ANN) foi introduzido na modelação e previsão da procura turística por volta da segunda metade da década 1990 (Chen, Lai & Yeh, 2012). O Modelo ANN é considerado como sendo um modelo promissor e alternativamente válido para a modelação e previsão principalmente quando comparado com modelos tradicionais (modelo regressão linear múltipla e séries temporais univariadas) (Zhang, 2003). As ANN são novas ferramentas computacionais que têm encontrado uma grande utilização na resolução de muitos problemas complexos do mundo real (Basheer & Hajmeer, 2000) e são definidas como sendo um sistema de processamento de informação cuja estrutura e funcionamento é inspirado nas redes neuronais biológicas (Palmer, Montañó & Sesé, 2006).

O modelo de ANN tem despertado interesse a vários investigadores ligados a Engenharia, Medicina, Psicologia, Matemática, Estatística e Biologia (Palmer, *et al.*, 2006; Khashei, Hamadani & Bijari, 2012; Rodrigues & Teixeira, 2010). O crescente interesse no modelo ANN está ligado a benefícios a si adjacentes, tais como, a capacidade de reconhecer e aprender das relações subjacentes entre a camada de entrada e a camada de saída, sem consideração física explícita, independentemente dos problemas de dimensão e sistema de não-linearidade; e a elevada tolerância aos dados contendo erros de ruído e de medição devido a processamento distribuído no interior da rede (Basheer & Hajmeer, 2000; Palmer, *et al.*, 2006).

Na perspetiva de Law e Au (1999) a arquitetura da rede neuronal do tipo *feedforward* é composta por três camadas distintas, ou seja, uma camada de entrada, uma ou mais camadas escondidas (ocultas) e uma camada de saída; cada uma dessas camadas contém nós, e os mesmos encontram-se conectados com os nós na camada adjacente. Para estes autores, cada nó de uma rede neuronal é uma unidade de processamento que contém um peso e uma função soma. Um peso (w) retorna um valor matemático para a força relativa de conexões para transferir dados a partir de uma camada para outra, considerando que uma função soma (y) calcula a soma ponderada de todos os elementos de entrada inseridos numa unidade de processamento. Os nós na camada de entrada representam variáveis independentes do problema, a camada oculta é usada para adicionar uma representação interna de manipulação de dados não-lineares e a saída de uma rede neuronal é a solução para um problema (Law & Au, 1999).

A relação entre a saída (*output*) e as entradas (*input*) $x_{i,1}, x_{i,2}, \dots, x_{i,p}$ segue a seguinte representação matemática definida na Eq. 1 (Zhang & Qi, 2005; Khashei, Hejazi, & Bijari, 2008).

$$Y_k = b_k + \sum_{j=1}^n w_{jk} f \left(\sum_{i=1}^m W_{ji} X_i + B_{0j} \right) \quad [1]$$

Onde: Y_k ($k=1,2,3\dots p$) representa a variável de saída; m corresponde ao número de nós na camada de entrada (número de variáveis de entrada); n corresponde ao número de nós na camada oculta; f corresponde a função de ativação sigmoideal (a equação indica também a utilização duma função de ativação linear na camada de saída); W_{ji} ($i=0,1,2,3,\dots,m; j=1,2,3,\dots,n$) corresponde ao vetor de pesos que liga os nós da camada oculta aos da camada de saída; w_{jk} corresponde aos pesos que relacionam os nós da camada de entrada com os da camada oculta e são parâmetros do modelo, designados por pesos de conexão. O b_k e B_{0j} indicam os desvios dos termos independentes (*bias*) associados à cada nó da camada de saída e da camada oculta, respetivamente.

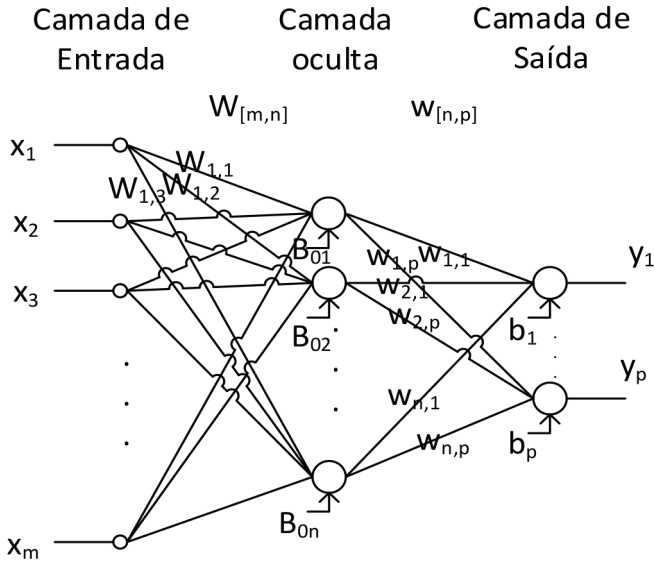


Figura 1. Rede Neuronal Feedforward.

Na construção de Redes Neurais Artificiais, a função ativação mais comumente utilizada na camada oculta é a Função Sigmoidal Logarítmica (Haykin, 1999; Zhang, Patuwo & Hu, 1998). A função de ativação sigmoidal logarítmica é dada pela Eq. 2 e representada na Figura 2.a. A função de ativação sigmoidal logarítmica tem uma variação de 0 a +1 (Haykin, 1999). Esta função é utilizada para transformar a saída de modo que ela caia numa zona aceitável e é definida como uma função estritamente crescente que exibe um balanceamento adequado entre comportamento linear e não linear (Haykin, 1999; Fernandes, 2005). Para Law (2000), a referida transformação é feita antes que a saída atinja o próximo nível e a finalidade desta função é de impedir que o valor de saída seja muito grande, pois o valor de $y(a)$ tem de situar-se entre 0 e +1.

$$f(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

Outra função, não menos importante e que é usada na camada oculta, é a Função Tangente Hiperbólica ou Tangente Sigmoidal, definida pela Eq. 3 e representada na Figura 2.b. É usada quando se pretende que a função de ativação assuma valores negativos, ou seja, uma forma antissimétrica em relação à origem (Haykin, 1999).

$$f(x) = \left(\frac{e^x - e^{-x}}{e^x + e^{-x}} \right) \tag{3}$$

Na camada de saída, a função mais usada é a função linear, dada pela Eq. 4 e representada na Figura 2.c.

$$f(x) = x \tag{4}$$

A Figura 2 ilustra as funções de ativação sigmoidal logarítmica, tangente sigmoidal e linear.

A característica mais importante de uma rede neuronal artificial é a sua capacidade de aprender a partir de seu ambiente e de melhorar o seu desempenho através da aprendizagem (Haykin, 1999).

Para Sivanandam e Paulraj (2003), aprendizagem é um processo no qual a rede ajusta os seus parâmetros (pesos sinápticos) em resposta a estímulos de entrada de modo que a resposta de saída

real convirja para a resposta de saída desejada. A aprendizagem supervisionada ou com professor é de longe a técnica mais usada no campo das redes neurais artificiais. Neste tipo de aprendizagem, a principal condição é a existência de um professor capaz de fornecer respostas exatas para as saídas da rede quando um erro ocorrer, ou de fixar numa relação ao ambiente as unidades de entrada e de saída da rede (Haykin, 1999).

Uma rede neuronal *feedforward* supervisionada aprende a partir de dados de treino para descobrir padrões que representam entrada e variáveis de saída. Normalmente, o processo de aprendizagem envolve as seguintes etapas (Law & Au, 1999): (i) atribuir números aleatórios para os pesos; (ii) para cada elemento no conjunto de treino (um conjunto de observações da amostra utilizada para desenvolver o padrão ou relacionamento entre as observações), ajustar os pesos alfa e beta num processo de retropropagação do erro (algoritmo *Backpropagation*); (iii) comparar a saída com valores observados. O processo é interrompido quando o erro entre a saída e o alvo for inferior a um valor predefinido ou se, durante um número previamente definido de iterações, o erro num outro conjunto de validação não for inferior. Este último processo denomina-se por validação cruzada e impede que o modelo se sobre-ajuste aos dados do conjunto de treino, garantindo à rede neuronal a capacidade de generalização.

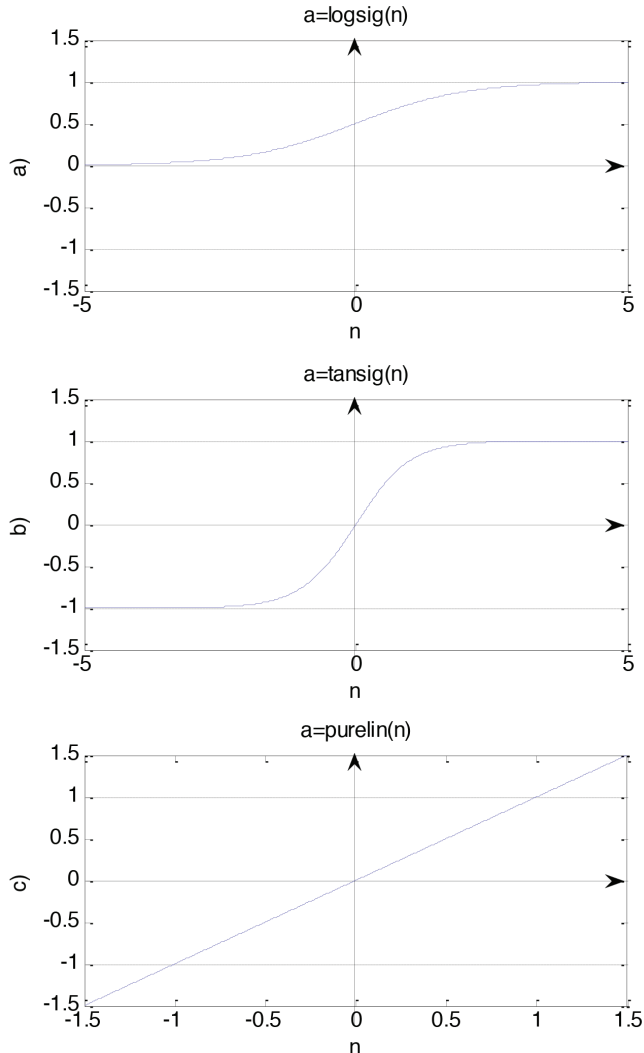


Figura 2. Gráfico de Funções de Ativação.

3.3 Modelo de Regressão Linear Múltipla

O modelo RLM considera-se flexível quando se pressupõe existir uma relação linear entre a variável dependente (explicada) e conjunto de variáveis independentes (explicativas) de natureza quantitativa ou categórica (e.g., Aiken, West & Pitts, 2003; Marôco, 2014). O objetivo deste modelo é explicar e prever o comportamento da variável dependente em função das variáveis independentes (Song & Li, 2008), necessitando para tal da expressão analítica obtida através do Método dos Mínimos Quadrados, tal como referem diferentes autores (e.g., Pestana & Gageiro, 2005; Hair, Black, Babin, Anderson & Tatham, 2009; Marôco, 2014). O método dos mínimos quadrados é utilizado para determinar a equação da regressão, os coeficientes estimados da regressão parciais, medidas de ajustamento global do modelo e a contribuição de cada variável independente no modelo. A expressão do modelo vem dada por Pestana e Gageiro (2005, p. 15):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_j X_{ji} + \dots + \beta_k X_{ki} + \text{com } i = 1, 2, 3, \dots, n; j = 1, 2, \dots, [5]$$

Onde: n - representa a dimensão da amostra; k - representa o número de variáveis exógenas observáveis adicionadas a constante, onde X e Y são variáveis observáveis; ε_i - variável exógena não observável e aleatória, a qual inclui todas as influências em que não são explicados por X ; β_0 - representa a ordenada na origem; $\beta_j (j=1, 2, \dots, k)$ - representa os parâmetros do modelo, isto é, grandezas que nele assumem sempre o mesmo valor, o mesmo é dizer representam os declives parciais.

Segundo Marôco (2014), ao modelo de regressão linear está subjacente um conjunto de pressupostos para a sua aplicação, nomeadamente a linearidade, o valor esperado nulo da variável residual, a homocedasticidade, a ausência de correlação, a independência dos resíduos, e a multico-linearidade.

3.4 Medidas de Avaliação do Desempenho dos Modelos

Existem várias medidas para medir a precisão da previsão. Para o presente trabalho, tendo em vista a medição da precisão dos modelos de regressão linear múltipla e o modelo de redes neurais artificiais vai recorrer-se à medida de precisão sugerido por Burger, Dohnal, Kathrada e Law (2001), ou seja, o erro percentual absoluto médio (MAPE - Mean Absolute Percentage Error) e ao coeficiente de correlação de Pearson (r).

A medida de Erro Percentual Absoluto Médio (MAPE) é de longe a mais frequentemente usada para medir a precisão dos modelos de previsão (Li, Song & Witt, 2005). O MAPE é uma medida baseada no erro da previsão dada pela subtração entre o valor real da série e o valor obtido pela previsão, ou seja, que mede o erro em qualquer período da previsão (Frechtling, 2012). Para este autor, o MAPE é a medida mais simples, mais atrativa e permite comparar o desempenho de diferentes modelos de previsão com diferentes períodos e ainda com diferente número de observações. É obtido através da soma dos erros absolutos para qualquer período, dividido pelo número de observações. Importa referir que, na precisão dos modelos um MAPE baixo é preferível do que um valor elevado, dado que, indica um erro percentual baixo que é o critério para a escolha de um modelo em detrimento doutro modelo.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \quad [6]$$

Na Eq. 6, o n representa o número de observações utilizadas no estudo, no presente estudo são 120 observações; o $Y - \hat{Y}$ designa-se de erro de previsão; Y_i representa o valor atual real da va-

riável, que no presente trabalho é a Procura Turística para Moçambique e \hat{Y} representa o valor da Procura Turística para Moçambique, prevista para o mesmo período; e i representa cada período em análise.

Na Tabela 1, apresenta-se pormenorizadamente o critério proposto por Lewis (1982) para avaliação dos modelos de modelação, RLM e ANN, usando a medida MAPE.

Tabela 1. Critério de Avaliação dos Modelos.

MAPE (Valor em Percentagem)	Critério de avaliação do modelo
$MAPE \leq 10\%$	Previsão com alta precisão
$10 < MAPE < 20\%$	Previsão com boa precisão
$20 \leq MAPE < 50\%$	Previsão com precisão Razoável
$MAPE \geq 50\%$	Previsão com baixa precisão

Fonte: Adaptado de Lewis (1982, p. 40).

O coeficiente de correlação de *Pearson* (r) mede o grau de associação linear entre duas variáveis numéricas (Levine, Berenson & Krehbiel, 2006). Para estes autores, o coeficiente de correlação varia de -1 (correlação negativa perfeita) a 1 (correlação positiva perfeita).

Quando o coeficiente de correlação de *Pearson* (r) entre duas variáveis é negativo, significa que elas variam em sentido inverso (quando uma aumenta a outra diminui). O inverso verifica-se quando o coeficiente de correlação é positivo, ou seja, as variáveis variam no mesmo sentido (quando uma variável aumenta a outra variável também aumenta). A Eq. 10 apresenta a fórmula sugerida por: Burger, Dohnal, Kathrada, (2001:407) para o cálculo de coeficiente correlação de *Pearson* (r).

$$r = \frac{\left[n \sum (Y_i * \hat{Y}_i) - \sum (Y_i) * \sum (\hat{Y}_i) \right]}{\sqrt{\left[n \sum Y_i^2 - (\sum Y_i)^2 \right] * \left[n \sum \hat{Y}_i^2 - (\sum \hat{Y}_i)^2 \right]}}, (i = 1, 2, 3, \dots, n) \quad [7]$$

Onde, o r representa o coeficiente de correlação de Pearson; e os parâmetros Y_i e \hat{Y}_i representam, respetivamente, o valor real e o valor previsto do número de dormidas em Moçambique, para $i=1,2,\dots,n$.

4. APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Em termos das variáveis há a referir que o estudo tem uma variável dependente ou explicada designada número de dormidas e um conjunto de catorze variáveis explicativas referentes aos cinco maiores emissores turísticos de Moçambique, nomeadamente: África do Sul [AS]; Estados Unidos da América [EUA], Moçambique [MOC], Portugal [PT], e Reino Unido [RU].

4.1 Análise do Comportamento das Variáveis

Fazendo agora uma análise descritiva, gráfica, da variável 'número de dormidas', (Figura 3), pode observar-se a evolução para o período de Janeiro de 2004 a Dezembro de 2013. Para o caso de Moçambique, esta variável não apresenta uma tendência de sazonalidade típica e constante ao longo dos anos, mas há a considerar três situações distintas para os meses de Janeiro, Abril e Dezembro. O Janeiro é o mês em que menos turistas são recebidos e este facto deve-se à razão de que os turistas neste período estão a fazer um movimento inverso, isto é, retorno à terra de origem.

Relativamente ao mês de Abril, há a considerar que a série regista um incremento e este deve-se ao facto de ser um mês em que se comemora a Páscoa Cristã, sendo um período de tolerância nos países vizinhos, o que até certo ponto origina um fluxo turístico de entrada relevante para Moçambique e consequentemente uma procura maior pelas estâncias turísticas. Por fim, Dezembro é o mês em que mais turistas são recebidos, facto que se deve às seguintes razões: este é o mês da quadra festiva natalícia e é o período em que se verificam as férias laborais e escolares o que motiva muitos turistas a se deslocarem, quer nacionais ou estrangeiros, com principal destaque para os turistas Sul-Africanos que representa a maior percentagem de entradas com cerca de 32% em 2004 e cerca de 44% em 2013 (INE, 2014). De salientar que no ano de 2011 verificou-se um pico mais elevado devido à preparação e realização dos X Jogos de Africanos. Também já em Junho/Julho de 2010 se verificou um aumento no número de dormidas e isso pode dever-se à realização do Mundial de Futebol realizado na África do Sul, pois a cidade de Maputo era um local de entrada de adeptos do futebol, onde os mesmos tinham a oportunidade de visitar dois países aquando da sua estada ao longo do mundial de futebol.

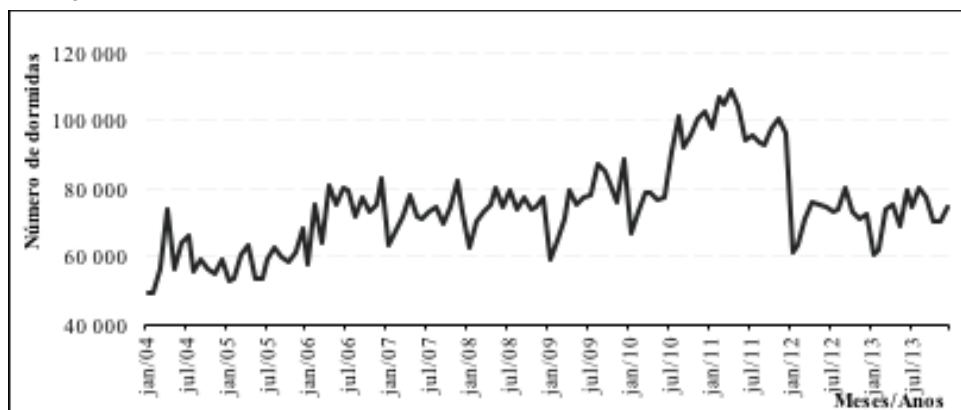


Figura 3. Número de Dormidas em Moçambique.

Fazendo agora uma análise descritiva exploratória, gráfica, da variável 'Índice Harmonizado de Preços ao Consumidor, IPC' (Figura 4), estão evidenciados os IPC dos cinco principais mercados emissores turísticos que se presumem influenciar significativamente o número de dormidas em Moçambique. O IPC representa o custo de vida numa dada economia. Observando para já, os gráficos da figura nota-se uma evolução e tendência positiva ao longo do tempo, ou seja, os preços de bens e serviços têm uma tendência de crescimento, o que até certo ponto penaliza a quantidade dos bens e serviços a adquirir no geral e turísticos, em particular.

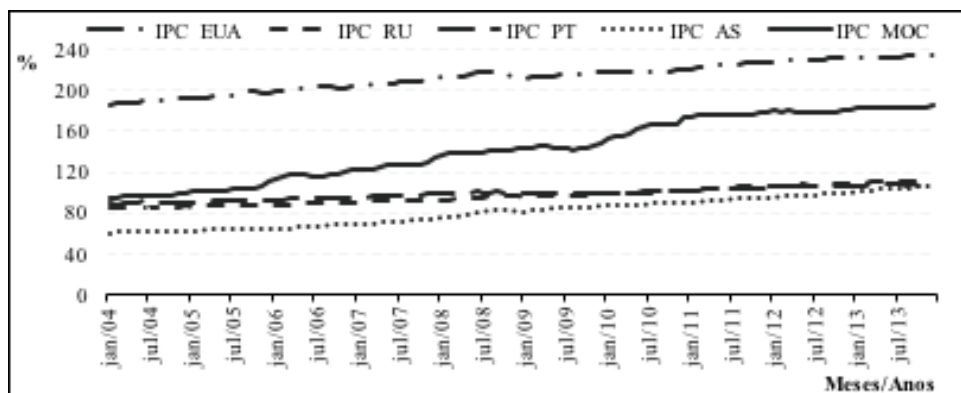


Figura 4. Índice Harmonizado de Preços ao Consumidor (IPC).

Analisando a informação apresentada na Figura 5, que evidencia o ‘Produto Interno Bruto per capita, PIB’, ou seja, rendimento médio por habitante, dos principais mercados emissores turísticos de Moçambique (incluindo Moçambique), constata-se que: o PIB regista uma tendência de crescimento ao longo do tempo. A tendência crescente do PIB pode considerar-se um impulsionador ou catalisador do turismo, dado que, para que os turistas se desloquem, o rendimento constitui um dos elementos chave. Da análise da figura abaixo, constata-se ainda que os turistas de nacionalidade Norte Americana possuem maior PIB (rendimento por habitante), o que desde logo os coloca com maior probabilidade de praticar turismo. Em segundo, estão os turistas do Reino Unido, de Portugal, da África do Sul e por fim, turistas de Moçambique.

Importa referir que o produto interno bruto per capita de cada país foi multiplicado pela taxa de câmbio entre essa moeda e o metical (moeda de Moçambique) de modo a obter o produto interno bruto em moeda de Moçambique.

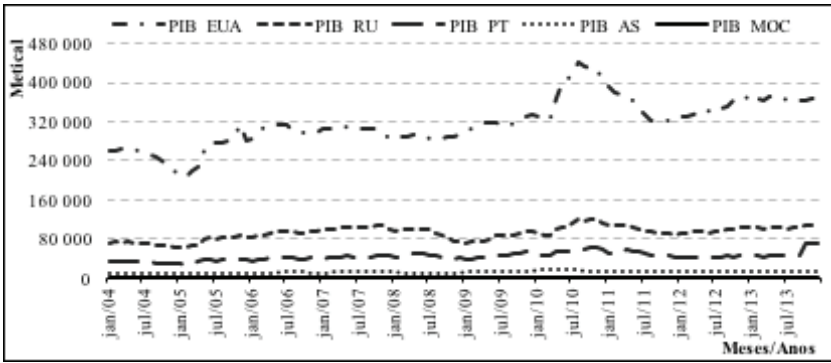


Figura 5. Produto Interno Bruto per capita.

Quanto à variável ‘Taxa de Câmbio, TC’ (Figura 6) dos principais mercados emissores turísticos em Moçambique que se presume influenciar o número de dormidas, pode observar-se a existência de uma evolução ao longo do tempo, pese embora, haja muita oscilação entre os anos 2004 e 2013, com exceção da África do Sul que tem fraca variabilidade ou oscilação.

A taxa de câmbio é representativa do custo de vida, quando há tendência de crescimento, ela serve como catalisador do turismo e quando tende a baixar ela penaliza a propensão ao turismo, dado que, os turistas ficam com menos rendimento para gastar. Na comparação entre as moedas patentes na figura abaixo, constata-se que a LIBRA é a moeda mais forte face ao METICAL (medido em unidade monetárias, u.m), desde modo, mantendo os restantes fatores constantes os turistas do Reino tem maior probabilidade de propensão ao turismo. Segue-se o EURO, o DOLAR e por último o RAND.

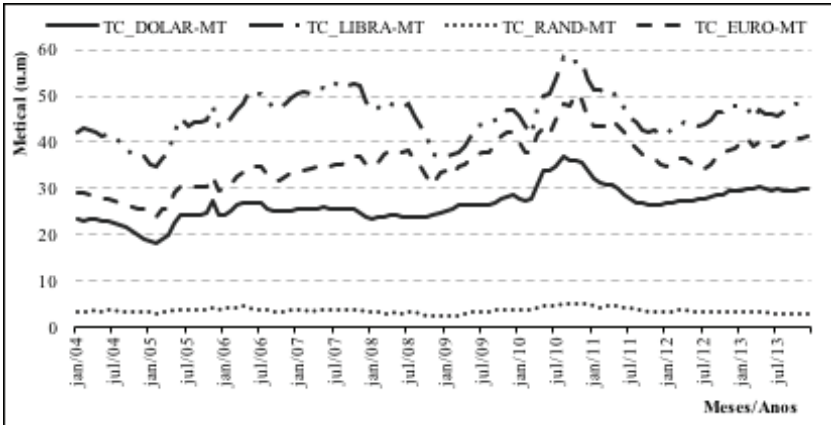


Figura 6. Taxas de Câmbio dos Principais Mercados Emissores face ao METICAL.

4.2. Modelação e Previsão da Procura Turística: Modelo Ann Vs Modelo Rlm

4.2.1. Modelo de Redes Neuronais Artificiais

Para a construção do Modelo de Redes Neuronais Artificiais, começou-se por construir a matriz das correlações de *Pearson*, tendo sido selecionadas e testadas, para a camada de entrada as variáveis que mais estavam correlacionadas com a variável a prever (número de dormidas nos estabelecimentos hoteleiros e similares) e menos entre si. Neste âmbito, foram selecionadas as seguintes variáveis: índice harmonizado de preços ao consumidor de Moçambique (IPC_MOC), taxas de câmbio entre TC_DOLAR-MT, TC_EURO-MT, e TC_RAND-MT. Optou-se por incluir as variáveis PIB_PT e PIB_EUA por terem uma correlação significativa com a variável de saída, e por não serem muito correlacionadas entre elas. Foram treinados e testados vários modelos com a combinação destas variáveis na entrada para identificar o melhor conjunto de variáveis. No total foram construídos 107 modelos. Na Tabela 5 são apresentados os oito melhores resultados a nível do MAPE e do coeficiente de correlação de *Pearson* (r).

O modelo de redes neuronais usado é do tipo multicamadas, tendo-se usado três camadas distintas, ou seja, uma camada de entrada, que corresponde às variáveis anteriormente referenciadas mais 12 meses de atraso correspondentes ao ano anterior; uma camada escondida (oculta); e uma camada de saída (que corresponde ao número de dormidas nos estabelecimentos hoteleiros e similares), com uma estrutura *feedforward*. Na camada escondida usaram-se as funções de ativação Sigmoidais [TanSig] e [LogSig] e na camada de saída utilizou-se a função de ativação linear [PureLin], por ser as que melhores resultados proporcionam para este tipo de arquiteturas. No treino da rede utilizou-se o algoritmo *Resilient Backpropagation* (Fernandes, Teixeira, Ferreira & Azevedo, 2013), uma variante do algoritmo de treino *Backpropagation*.

Os dados disponíveis foram divididos em três conjuntos distintos, ou seja, num conjunto de treino, de validação e de teste (Law & Au, 1999). O conjunto de teste nunca foi visto pelo modelo no processo de treino. O conjunto de teste consiste nos últimos 12 meses do ano 2013. O conjunto de validação foi experimentado com duas dimensões de 6 e 12 meses. Este conjunto de validação consiste nos 6 ou 12 meses anteriores ao conjunto de teste (de janeiro ou julho a dezembro de 2012). Tendo-se verificado que os resultados não diferiam significativamente, optou-se por usar o conjunto de validação de 6 meses (de julho a dezembro de 2012). O conjunto de treino corresponde aos restantes meses disponíveis. Num total que variou, dependendo do modelo, entre Fevereiro de 2005 a Junho de 2012, para o caso dos modelos com as diferenças e usando 12 dados anteriores na entrada; e Maio de 2004 a Junho de 2012, no modelo com apenas 4 meses anteriores na entrada. Assim o conjunto de treino contém entre 89 e 97 pares de entrada saída.

Foram experimentadas variantes do modelo de redes neuronais. Nomeadamente:

- (i) Combinações das variáveis IPC_MOC, TC_EURO-MT, TC_RAND-MT, TC_DOLAR-MT, PIB_PT e PIB_EUA,
- (ii) Número de meses anteriores da variável de saída (Dormidas),
- (iii) Domínio da variável de saída – foi usada a variável dormidas com o seu valor absoluto (d), no domínio logarítmico (ld) e com as diferenças de logaritmos (dld),
- (iv) Funções de ativação dos nos da camada escondida – experimentou-se as funções tangente sigmoideal e sigmoideal logarítmica,
- (v) Número de nos na camada escondida – alguns valores entre 3 e 40 nós.

Para os modelos em que foram usadas as diferenças dos logaritmos, estas foram determinadas pela diferença das dormidas do mês atual para o mês anterior como na Eq. 8.

$$dld(i) = ld(i) - ld(i-1)$$

Neste caso a previsão das dormidas é obtido por um processo de reposição das diferenças como na Eq. 9.

$$ld(i) = dld(i) + ld(i-1) \quad [9]$$

Sendo $ld(i)$ o valor no domínio logarítmico das dormidas previstas para o mês i , $dld(i)$ a diferença das dormidas determinadas na saída da rede neuronal e $ld(i-1)$ o logaritmo do valor real das dormidas no mês anterior.

Das diversas experiências de treino e teste das redes neuronais, contruiu-se uma tabela com 107 modelos, medindo o MAPE e o coeficiente r nos conjuntos de teste e nos três conjuntos (Série=treino+validação+teste). Da análise dos resultados mais satisfatórios no conjunto de teste foram selecionados os modelos com melhores resultados apresentados resumidamente na Tabela 2.

Os resultados foram selecionados com base no Erro Percentual Absoluto Médio (MAPE) e no coeficiente de correlação de *Pearson* do conjunto de teste. Na regra de decisão pelo MAPE um modelo tem melhor previsão quando apresentar o menor valor e com relação ao coeficiente de correlação de *Pearson* modelo tem melhor previsão quando apresenta o maior valor.

Tabela 2. Tabela Resumo do Modelo de Redes Neuronais Artificiais.

Modelo	Modelo				Nós	Série (3)		Test	
	Domínio (1)	Lags	Variáveis	Função de ativação (2)		MAPE	r	MAPE	r
1	Log	12	IPC_MOC; TC_EURO-MT; TC_RAND-MT; TC_DOLAR-MT	TanSig. e PureLin	25	10,54	0,585	7,92	0,574
2	Log	12	IPC_MOC; TC_EURO-MT; TC_RAND-MT; TC_DOLAR-MT; PIB_PT	TanSig. e PureLin	11	5,36	0,93	6,45	0,585
3	Log	12	IPC_MOC; TC_EURO-MT; TC_RAND_MT; TC_DOLAR-MT; PIB_PT; PIB_EUA	TanSig. e PureLin	11	1,13	0,978	8,22	0,712
4	Log	12	IPC_MOC; TC_EURO-MT; TC_RAND_MT; TC_DOLAR-MT; PIB_EUA	TanSig. e PureLin	18	1,13	0,982	6,5	0,696
5	Dif-Log	12	IPC_MOC; TC_EURO-MT; TC_RAND_MT; TC_DOLAR-MT; PIB_PT	TanSig. e PureLin	6	4,99	0,919	7,84	0,506

6	Dif-Log	12	IPC_MOC; TC_EURO-MT; TC_RAND-MT; TC_DOLAR-MT; PIB_PT	TanSig, e PureLin	25	5,92	0,886	7,81	0,612
7	Dif-Log	12	IPC_MOC; TC_EURO-MT; TC_RAND-MT; TC_DOLAR-MT; PIB_EUA	TanSig, e PureLin	7	7,03	0,891	7,82	0,594
8	Dif-Log	12	IPC_MOC; TC_EURO-MT; TC_RAND-MT; TC_DOLAR-MT; PIB_EUA	TanSig, e PureLin	18	9,49	0,829	7,71	0,505

Notas:

- (1) - Domínio: Log = Logarítmico; e, Dif-Log = Diferença de Logaritmo
- (2) - Funções de ativação: TanSig = Tangente Sigmoidal e PureLin = Função Linear
- (3) - Série: conjunto de Treino, Validação e Teste

Em relação ao número de Lags ou meses anteriores na entrada da rede neuronal o modelo apresenta melhores resultados quando são usados 12 meses. Na Tabela 2 (resumo) já se apresentam apenas as variantes do modelo com 12 Lags, todavia foram experimentadas outras variantes com piores resultados.

A Tabela 2 mostra que em relação às variáveis de entrada, 4 delas (IPC_MOC, TC_EURO MT, TC_RAND-MT e TC_DOLAR-MT) são comuns a todos os melhores modelos e as variáveis relativas ao PIB são necessárias, pelo menos uma delas ou mesmo a combinação das duas (PIB_PT e PIB_EUA).

A função de ativação na camada escondida com melhores resultados foi a tangente sigmoidal. Na camada de saída foi sempre usada a função linear. O número de nós na camada escondida varia entre 6 e 25 nós para as diferentes combinações apresentadas na Tabela 2.

Analisando agora os resultados da Tabela 2 constata-se que o valor do MAPE e o valor do coeficiente de r variam no conjunto de teste entre [6,45 e 7,92] e [0,505 e 0,712], respectivamente.

Para a seleção do melhor modelo pode considerar-se aquele que tem um menor MAPE, que corresponde ao modelo cuja sequência de dormidas previstas está mais próxima das dormidas reais, ou pode selecionar-se o modelo com maior r que corresponde aquele cuja sequência de previsões das dormidas segue melhor as variações do comportamento das dormidas reais. Nem sempre um modelo com melhor MAPE (mais baixo) apresenta um melhor r (maior). Assim, uma vez que os valores são muito próximos entre os 8 modelos selecionados e apresentados na Tabela 2 qualquer um deles poderia ser usado para efeitos de previsão das dormidas. Contudo o modelo n.o 4 apresenta um valor de MAPE bastante baixo (6.50), quase o menor valor, e simultaneamente um r elevado (0.696) relativamente aos restantes modelos, no conjunto de teste. Este modelo apresenta ainda resultados muito bons quando considerados todos os dados (conjuntos de treino, validação e teste) com um valor de MAPE de 1.13 e r de 0.982.

O modelo n.o 4 tem 18 nós na camada escondida; usou o domínio logarítmico; possui 12 Lags na camada de entrada; usou a função de ativação tangente sigmoidal na camada escondida e função linear na camada de saída, e usou as seguintes variáveis IPC_MOC, TC_EURO-MT, TC_RAND MT, TC_DOLAR-MT e PIB_EUA conduziu ao melhor resultado em termos de MAPE e Coeficiente r .

4.2.2. Modelo de Regressão Linear Múltipla

Tal como já referido para a modelação da procura turística utilizando o modelo de RLM tomou-se em consideração os dados de Janeiro de 2004 a Dezembro de 2012, somando um total de 108 observações mensais. Seguindo esta linha de pensamento, para a obtenção dos parâmetros de regressão deste modelo, recorreu-se ao método dos mínimos quadrados ordinários OLS⁴. Inicialmente, utilizou-se o modelo com todas as catorze variáveis independentes ou explicativas referentes aos cinco emissores turísticos descritos no ponto 4.1. Com a utilização de todas variáveis explicativas constatou-se que havia violação dos pressupostos associados ao modelo de RLM. O principal destaque vai para o pressuposto da multicolinearidade, dado que, os valores do fator de inflação da variância (VIF) são superiores ao valor de referência 10 (valor máximo) na maioria das variáveis explicativas e ao fato de nem todas variáveis terem poder explicativo (Bicak, Altinay & Jenkins, 2005). Adicionalmente, é possível verificar a violação de pressupostos, tais como, o pressuposto da autocorrelação e o pressuposto da homogeneidade através da estatística de *White*. Havendo a violação de pressupostos, principalmente o da multicolinearidade, a análise da regressão é extremamente confusa e desprovida de significado, ou seja, o modelo de RLM deixa de ser válido (Marôco, 2014).

Neste contexto, tendo em vista ultrapassar a violação dos pressupostos associados ao modelo RLM, trabalharam-se várias combinações diferentes com as catorze variáveis dependentes. As variáveis que não tinham grande contribuição no modelo foram removidas, deixando apenas aquelas que tinham contribuição significativa. Neste contexto, as variáveis seguintes: TC_EURO-MT, TC_RAND-MT, TC_DOLAR_MT e IPC_MOC conduziram ao melhor modelo em termos de qualidades estatísticas e que não violava os pressupostos associados ao modelo de RLM.

A Eq. 10 apresenta a equação da procura turística para Moçambique, tendo por base as variáveis referidas anteriormente.

$$PTM_i = \beta_0 + \beta_1 IPC_MOC + \beta_2 TC_DOLAR_MT + \beta_3 TC_RAND_MT + \beta_4 TC_EURO_MT + \varepsilon_i \quad [10]$$

Posteriormente a escolha do modelo que mais se ajustava aos dados, tal como já foi referenciado no início deste ponto, para a obtenção dos parâmetros do modelo recorreu-se novamente ao método dos mínimos quadrados ordinários (OLS).

Na Tabela 3 apresentam-se os resultados do modelo estimado.

Tabela 3. Parâmetros do Modelo RLM e Respetivos Testes/Indicadores Estatísticos.

	Coeficiente	Erro Padrão	Rácio-t	Valor de prova	Significância
CONSTANTE	2939,04	5635,82	0,5215	0,6031	
IPC_MOC	186,526	49,1424	3,796	0,0002	*
TC_DOLAR-MT	-1623,41	608,96	-2,666	0,0089	*
TC_RAND-MT	7107,18	2286,09	3,109	0,0024	*
TC_EURO-MT	1777,24	344,115	5,165	<0,001	*
r^2	0,698586		r^2 Ajustado		0,686880
$F(4, 103)$	59,68058		Valor de prova (F)		<0,001
<i>Durbin-Watson (d)</i>	1,046063		Valor de prova (d)		<0,001
<i>Teste de White (TR²)</i>	29,20223		Valor de prova (White)		0,142514
<i>Normalidade: χ^2</i>	0,482		Valor de prova χ^2		0,7858
<i>Cochrane-Orcutt</i>	2,04353		Fator de Inflação da variância		VIF < 10

Nota: *, Significância de 1%.

⁴ Do inglês Ordinary Least Square.

Da análise dos resultados que constam da Tabela 3, referentes ao modelo RLM estimado constata-se que:

- (i) O coeficiente de determinação r^2 que mede a qualidade do ajustamento do modelo de regressão aos dados é de 0,699 e o ajustado é de 0,687. Logo, pode afirmar-se que cerca de 68,7% da variabilidade total da variável dependente é explicada pelas variáveis independente que formam parte do modelo de regressão linear ajustado;
- (ii) Da estatística de F tendo em vista analisar o ajustamento do modelo na sua globalidade, o modelo registou $F(4,103) = 59,68058$ e um valor de prova inferior a 0,001, logo verifica-se que o modelo, na sua globalidade apresenta qualidades de ajuste satisfatórias;
- (iii) Através da estatística de t para adequabilidade das variáveis independentes de forma individual no modelo, verifica-se que as variáveis IPC_MOC, TC_EURO-MT, TC_DOLAR-MT e TC_RAND MT tem qualidades estatísticas significativas aceitáveis e satisfatórias, dado que, possuem um valor de prova inferior a 1%. De salientar ainda que no modelo o coeficiente das variáveis é positivo com exceção da variável TC_DOLAR-MT que apresenta uma tendência de crescimento inverso à variável número de dormidas;
- (iv) Para testar a autocorrelação dos resíduos, utilizou-se a estatística de *Durbin-Watson* (d). Verifica-se que o valor de teste de *Durbin-Watson* é de $d=1,05$, valor de prova inferior a 0,001, sendo que o valor de limite inferior $d_L=1,63$ e limite superior $d_U=1,74$, deste modo o valor de (d) encontra-se entre $[0;d_L]$, ou seja, na zona de autocorrelação positiva dos resíduos, pelo que, rejeita-se a hipótese nula, isto é, há evidências estatísticas para afirmar que existe autocorrelação entre os resíduos. Na correção da autocorrelação entre os resíduos através do teste de *Cochrane-Orcutt*, constata-se que o $d=2,04$, ou seja, o valor de *Durbin-Watson* encontra-se no intervalo de $[1,74;2,27]$ correspondente a $[d_U;4-d_U]$, pelo que pode concluir-se pela ausência da autocorrelação, ou seja, há evidências estatísticas suficientes para não se rejeitar a hipótese nula; Para testar o pressuposto da homogeneidade dos resíduos, utilizou-se a estatística de *White* através do teste de TR^2 , o valor da estatística obtido foi de $TR^2=29,202$ e valor de prova de 0,146. Como o valor de prova é superior a 10%, logo constata-se que existem evidências estatísticas para não rejeitar a hipótese nula, isto é, pode-se afirmar que não existe violação do pressuposto de homoscedasticidade e as características dos estimadores são BLUE⁵.
- (v) Do teste de Multicolinearidade, tendo em vista a independência das variáveis explicativas, os valores do fator de inflação da variância (VIF) são inferiores ao valor de referência de 10, tal como apresentado na Tabela 4. Deste modo, pode concluir-se pela ausência de multicolinearidade, ou seja, não há relação entre as variáveis explicativas ou por outras palavras há independência entre as variáveis explicativas.

Tabela 4. Fator de Inflação da Variância.

Variáveis independentes do modelo	Valor de inflação da variância (VIF)
IPC_MOC	3,552
TC_DOLAR-MT	8,770
TC_RAND-MT	3,061
TC_EURO-MT	6,641

- (vi) Do teste da normalidade, na Figura 7, constata-se que o modelo segue distribuição normal dado que a estatística de $\chi^2(2)=0,482$ com valor de prova de 0,786, ou seja,

⁵ Do inglês Best Linear Unbiased Estimators.

maior que o nível de significância de 5%. Pode dizer-se que a hipótese dos erros seguem a distribuição normal de média 0 e desvio padrão constante $\varepsilon_i \sim N(0, \sigma)$ está salvaguardado o pressuposto, ou seja, há evidências estatísticas para afirmar que a hipótese nula não é violada porque a média é aproximadamente igual a zero e o desvio padrão é constante.

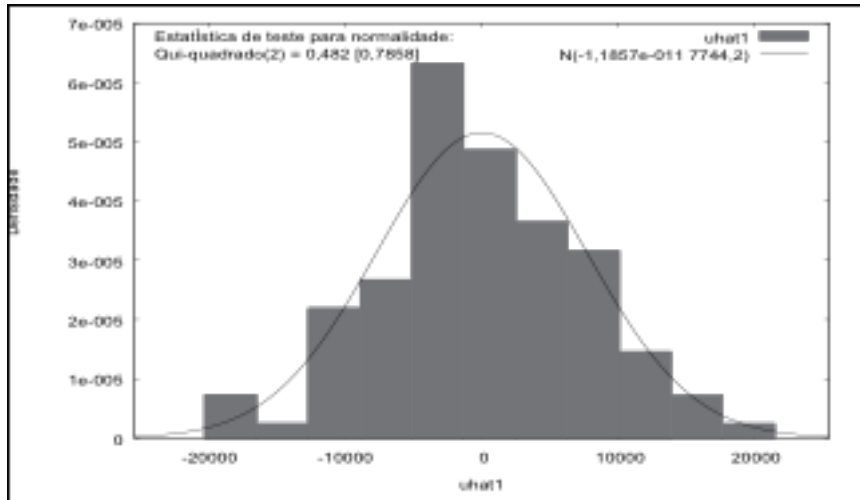


Figura 7. Normalidade dos Resíduos.

Tendo por base a informação anterior pode dizer-se que o modelo de Regressão Linear Múltipla construído para explicar a procura turística para Moçambique é o seguinte:

$$\begin{aligned} PTM_i &= 2939,04 + 186,526IPC_MOC - 1623,42TC_DOLAR - MT + 7107,18 \\ TC_RAND - MT + 1777,24TC_EURO - MT + \varepsilon_i \end{aligned} \quad [11]$$

Após a validação de todos os pressupostos associados ao modelo de Regressão Linear Múltipla (RLM), o passo seguinte foi calcular o Erro Percentual Absoluto Médio (MAPE) e o Coeficiente de correlação de Pearson. Para o modelo de RLM, o cálculo dos indicadores, acima referenciados, foi efetuado com base nos dados constantes na Tabela 4.

4.2.3. Desempenho do Modelo Rlm

Para o cálculo do MAPE e coeficiente de correlação de *Pearson*, recorreu-se a metodologia descrita no ponto 3.4, respetivamente às fórmulas da Eq. 6 e da Eq. 7. Assim sendo, para o cálculo do MAPE, começou-se por calcular a previsão referente ao ano de 2013. Após este processo, o segundo passo foi calcular o erro absoluto através da subtração do módulo entre os dados da variável real número de dormidas do ano 2013 e os dados da previsão de dormidas para o mesmo período. O terceiro passo foi calcular o erro percentual através da divisão do erro pelos dados da variável real número de dormidas, e por fim, o quarto passo foi calcular o erro percentual absoluto médio (MAPE) através da soma dos erros percentuais absolutos divididos por 12 observações referentes aos 12 meses do ano de 2013.

Para a obtenção dos valores para o cálculo do r , primeiro, calculou-se a previsão do número de dormidas através do método descrito anteriormente para o indicador MAPE. Em segundo lugar, multiplicou-se os valores da variável número de dormidas com os valores da variável previsão de dormidas. Em terceiro lugar, calculou-se o quadrado da variável número de dormidas. Em quarto lugar, calculou o quadrado do número da variável previsão de dormidas. Por fim em quinto lugar,

recorreu-se a fórmula proposta por (Law & Au, 1999), para obter o coeficiente de correlação de *Pearson*. De referir que, o valor do coeficiente r obtido é para o período de teste, ou seja, para o ano de 2013.

Na Tabela 5, apresenta-se resumidamente os resultados obtidos através da medida de Erro Percentual Absoluto Médio (MAPE) e do Coeficiente de Correlação de *Pearson* (r).

Tabela 5. Valores do MAPE e do Coeficiente de Correlação de Pearson para o Modelo de RLM.

Meses/Ano	Número Dormidas	Previsão de Dormidas	MAPE	r
Jan-13	60.230	83.879	14,65%	-0,783
Fev-13	61.804	84.946		
Mar-13	74.121	81.578		
Abr-13	75.256	82.834		
Mai-13	68.976	81.555		
Jun-13	79.797	80.249		
Jul-13	74.884	80.160		
Ago-13	79.938	81.098		
Set-13	77.162	81.337		
Out-13	70.167	83.046		
Nov-13	70.118	82.115		
Dez-13	74.895	82.967		
Total	867.348	985.763		

A partir dos resultados apresentados na Tabela 5, constata-se que o valor do MAPE associado ao modelo de RLM é de 14,65%. Tal como foi referido, este valor de MAPE foi calculado para o período de Janeiro a Dezembro de 2013, ou seja, um conjunto de 12 observações mensais. O resultado obtido para o critério sugerido por Lewis (1982) é considerado como sendo bom (boa precisão) dado que se situa no intervalo de $10\% < \text{MAPE} < 20\%$.

A partir dos resultados da Tabela 5, constata-se que o valor do r associado ao modelo é negativo, ou seja, $r=-0,7829$. Tal como já foi referenciado, quando o valor do r é negativo, isto é, correlação negativa, o valor real da série número de dormidas e o valor da previsão das dormidas têm comportamentos diferentes.

4.3. Desempenho dos Modelos Ann Vs Rlm

Tal como já foi referenciado anteriormente, para avaliação dos resultados precisão da previsão, recorreu-se aos indicadores Erro Percentual Absoluto Médio (MAPE) e coeficiente de correlação de *Pearson* (r). Assim, os resultados obtidos para os dois modelos selecionados com as duas metodologias utilizadas e para as respetivas medidas estatísticas/indicadores apresentam-se na Tabela 6.

Tabela 6. Comparação da Precisão dos Modelos ANN e RLM.

Modelos de previsão	Modelação		Previsão (Conjunto de teste_ ano 2013)	
	MAPE	r	MAPE	r
Modelo de Redes Neurais Artificiais	1,13%	0,982	6,50%	0,696
Modelo de Regressão Linear Múltipla	8,23%	0,836	14,65%	-0,783

A partir da Tabela 6 pode constatar-se o seguinte:

- (i) Para o modelo RLM, na modelação, o valor de MAPE correspondente é de 8,23% e o coeficiente de correlação de *Pearson* é de 0,836. Para a previsão o valor de MAPE é de 14,65% e o coeficiente de correlação de *Pearson* é de -0,783.
- (ii) Em contrapartida, para o método ANN, na modelação, o valor de MAPE correspondente é de 1,13% e o valor do coeficiente r é de 0,982. Para a previsão o valor de MAPE é de 6,50% e o valor do coeficiente r é de 0,696.
- (iii) Para avaliar os resultados da precisão dos modelos, com a medida MAPE usou-se o critério sugerido por Lewis, (1982), ou seja, valores inferiores a 10% são indicadores de alta precisão do modelo; valores inferiores a 20% são indicadores de boa precisão; e, valores inferiores a 50% são indicadores de precisão razoável. Analisando os resultados do MAPE descritos na tabela constata-se que o MAPE do modelo ANN está abaixo dos 10% o que revela alta precisão do modelo e o MAPE do modelo RLM é 14,65% o que revela boa precisão do modelo.
- (iv) Para avaliar os resultados da precisão, usando método de coeficiente de correlação de *Pearson* recorreu-se ao critério sugerido pelos autores Levine, Berenson e Krehbiel (2006), onde os valores do coeficiente r positivos são indicadores de correlação positiva entre a variável número de dormidas nos estabelecimentos hoteleiros e similares e a variável que resulta da previsão e valores negativos são indicadores da correlação negativa entre as variáveis anteriormente referenciadas. Observando para os resultados descritos na Tabela 6 constata-se que o coeficiente r do modelo ANN é positivo, o que pressupõe uma correlação positiva entre as variáveis, ou seja, a variável número de dormidas e previsão de dormidas têm o mesmo comportamento ao longo do tempo (quando uma cresce a outra também cresce), e o inverso, verifica-se no modelo RLM, que possui coeficiente r negativo (correlação negativa), ou seja, quando cresce a outra toma um comportamento contrário.
- (v) Em suma, os resultados apresentados anteriormente, para a previsão durante o período de Janeiro a Dezembro do ano 2013, indicam que o modelo ANN apresenta resultados melhores quando comparado com o modelo RLM, resultados similares foram obtidos por mais autores (e.g., Law e Au, 1999, Uysal, Sherif, & Roubi, 1999; Burger, Dohnal, Kathrada & Law, 2001).

A Figura 8 mostra a precisão dos modelos de previsão usados no trabalho, nomeadamente: modelo de Regressão Linear Múltipla e modelo de Redes Neurais Artificiais.

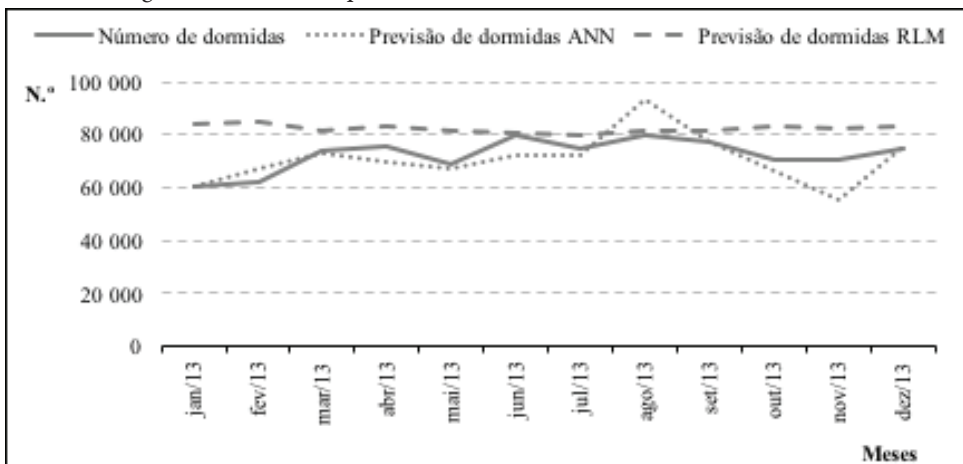


Figura 8. Número de dormidas, previsão de dormidas com os modelos ANN e RLM.

Observando a Figura 8 constata-se uma vez mais que o modelo ANN apresenta dados que tentam acompanhar a evolução da variável dormidas reais e o modelo de RLM não acompanha a evolução da variável número de dormidas.

5. CONCLUSÃO E FUTURAS LINHAS DE INVESTIGAÇÃO

O presente trabalho tinha como principal objetivo a comparação de dois modelos de previsão, nomeadamente: Modelo de Redes Neuronais Artificiais e Modelo de Regressão Linear Múltipla e aplicá-los para prever a procura turística em Moçambique a partir dos seguintes mercados emissores turísticos que se presume influenciarem o número de dormidas: África do Sul, Estados Unidos da América, Moçambique, Portugal e Reino Unido.

Relativamente ao modelo de Redes Neuronais Artificiais, o modelo inicial tinha catorze variáveis, mas devido a violação do pressuposto da multicolinearidade recorreu-se a matriz das correlações bivariadas, onde foram selecionadas as variáveis que estavam fracamente correlacionadas entre e si e fortemente correlacionadas com a variável dependente 'número de dormidas'. Nesta ordem de ideias, as variáveis IPC_MOC, TC_EURO-MT; TC_RAND-MT e TC_DOLAR-MT foram selecionadas à priori e as variáveis PIB_PT e PIB_EUA por melhorarem os resultados foram consideradas na construção do modelo.

Para o modelo de Regressão Linear Múltipla, o modelo inicial tinha catorze variáveis explicativas mas devido a violação de pressupostos associados a este modelo, recorreu-se a várias combinações com as variáveis até se encontrar o modelo ideal. Nesta ordem de ideias, as variáveis IPC_MOC; TC_EURO-MT; TC_RAND-MT e TC_DOLAR-MT constituem o modelo ideal e foram selecionadas para explicar a procura turística em Moçambique.

Para a avaliação do desempenho dos modelos recorreu-se às seguintes medidas: Erro Percentual Absoluto Médio (MAPE) e ao Coeficiente de Correlação de *Pearson* (r). De referir que a previsão foi efetuada para o período de Janeiro a Dezembro de 2013, correspondendo a 12 observações mensais.

Com os resultados empíricos obtidos nos dois modelos de previsão, chega-se a conclusão que o modelo de redes neuronais artificiais teve melhor desempenho a nível do MAPE e do Coeficiente de Correlação de *Pearson* quando comparado com o modelo de regressão Linear Múltipla (e.g., Uysal & Roubi, 1999; Law & Au, 1999; Law, 2000; Burger, Dohnal, Kathrada & Law, 2001; Machado, Fernandes & Teixeira, 2010; Cankurt & Subasi, 2012). Deste modo, sugere-se a aplicação do modelo de Redes Neuronais Artificiais para a previsão da procura turística em Moçambique, dado que, minimiza o MAPE e maximiza o coeficiente r .

Seguindo esta linha de raciocínio e como principais linhas de investigação futuras, tendo por base a investigação realizada e pelas principais conclusões obtidas, sugere-se a utilização de uma variável *dummy* para explicar o crescimento anormal da variável número de dormidas nos estabelecimentos hoteleiros e similares em Moçambique para o ano de 2011. O referido crescimento foi motivado pelo acolhimento dos Jogos Africanos em Moçambique. Ainda, sugere-se a utilização de outras variáveis determinantes da procura turística, como os gastos dos visitantes a Moçambique, as despesas com as deslocações e respetiva distância entre o mercado emissor e mercado receptor. Seria importante também conhecer a competitividade regional em Moçambique e utilizar estas variáveis como determinantes da procura turística para Moçambique.

REFERÊNCIAS BIBLIOGRÁFICAS

Aiken, L., West, S., & Pitts, S. (2003). Multiple regression analysis. In John A. Schinka & Wayne F. Velicer (Eds.), *Comprehensive handbook of psychology*. Volume 2. *Research methods in psychology* (pp. 483-507). New York: Wiley.

- Athanasopoulos, G., & Hyndman, R. (2008): Modelling and forecasting Australian domestic tourism. *Tourism Management*, 29(1), 19-31.
- Basheer, I., & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. *Journal of microbiological methods*, 43(1), 3-31.
- Burger, C., Dohnal, M., Kathrada, M., & Law, R. (2001). A practitioners guide to time-series methods for tourism demand forecasting: A case study of Durban, South Africa. *Tourism Management*, 22(4), 403-409.
- Cankurt, S., & Subasi, A. (2012). Comparison of linear regression and neural network models forecasting tourist arrivals to Turkey. In: *3rd International Symposium on Sustainable Development*, Sarajevo.
- Constantino H.; Fernandes, P. & Teixeira, J. P. (2016). Modelação e Previsão da Procura Turística em Moçambique: Um Estudo Comparativo Entre os Modelos de Redes Neurais Artificiais e Regressão Linear Múltipla. *Atas das XXVI Jornadas Luso-Espanholas - Gestão Científica*, Idanha-a-Nova. Portugal.
- Chen, C., Lai, M., & Yeh, C. (2012). Forecasting tourism demand based on empirical mode decomposition and neural network. *Knowledge-Based Systems*, 26, 281-287.
- Claveria, O., & Torra, S. (2014). Forecasting tourism demand to Catalonia: Neural networks vs. time series models. *Economic Modelling*, 36, 220-228.
- Coshall, J., & Charlesworth, R. (2011). A management oriented approach to combination forecasting of tourism demand. *Tourism Management*, 32(4), 759-769.
- Cunha, L., & Abrantes, A. (2013). *Introdução ao Turismo*, (5.a Ed.), Lisboa.
- Dwyer, L., Forsyth, P., & Dwyer, W. (2010). *Tourism Economics and Policy*. Bristol: Channel View Publications.
- Eurostat (2015). *Eurostat*. Acedido em Março de 2015, disponível em <http://ec.europa.eu/eurostat>.
- Federal Reserve Bank of St. Louis* (2015), acedido em Março de 2015, disponível On-line em: <https://resea-rch.stlouisfed.org/fred2/series/CPIAUCSL>
- Fernandes, P. (2005). *Modelling, Prediction and Behaviour Analysis of Tourism Demand in the North of Portugal*. Tese de Doutoramento, Universidade de Valladolid, Espanha.
- Fernandes, P. O., & Teixeira, J. P. (2007). A new approach to modelling and forecasting monthly overnights in the Northern Region of Portugal. In proceedings of *4th International Finance Conference*; Tunisia.
- Fernandes, P., & Teixeira, J. (2008). Previsão da Série Temporal Turismo com Redes Neurais Artificiais. *5.º Congresso Luso-Moçambicano de Engenharia - CLME' 2008 - A Engenharia no Combate à Pobreza, pelo Desenvolvimento e Competitividade*, Maputo-Moçambique.
- Fernandes, P. Odete, & Teixeira, J. P. (2009). New Approach of the ANN Methodology for Forecasting Time Series: Use of Time Index. *International Conference on Tourism Development and Management*. Kos - Greece.
- Fernandes, P., Teixeira, J., Ferreira, J., & Azevedo, S. (2013). Training neural networks by resilient backpropagation algorithm for tourism forecasting. In *Management Intelligent Systems* (pp. 41-49). Springer International Publishing.
- Frechtling, D. (2012). *Forecasting tourism demand*. Routledge.
- Goh, C., & Law, R. (2002). Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. *Tourism Management*, 23(5), 499-510.
- Gunter, U., & Önder, I. (2015). Forecasting international city tourism demand for Paris: Accuracy of uni-and multivariate models employing monthly data. *Tourism Management*, 46, 123-135.
- Hair, J., Black, W., Babin, B., Anderson, R., & Tatham, R. (2009). *Análise multivariada de dados*. Bookman.
- Haykin, S. (1999). *Neural Networks A Comprehensive Introduction*.
- INE (2014): *Instituto Nacional de Estatística de Moçambique*. On-line em www.ine.gov.mz.

- Khashei, M., Hamadani, A., & Bijari, M. (2012). A novel hybrid classification model of artificial neural networks and multiple linear regression models. *Expert Systems with Applications*, 39(3), 2606-2620.
- Khashei, M., Hejazi, S., & Bijari, M. (2008). A new hybrid artificial neural networks and fuzzy regression model for time series forecasting. *Fuzzy Sets and Systems*, 159(7), 769-786.
- Law, R. (2000). Backpropagation learning in improving the accuracy of neural network-based tourism demand forecasting. *Tourism Management*, 21, 331-340.
- Law, R., & Au, N. (1999). A neural network model to forecast Japanese demand for travel to Hong Kong. *Tourism Management*, 20(1), 89-97.
- Levine, D., Berenson, M., & Krehbiel, T. (2006). *Estatística para administração*. Pearson Educação.
- Lewis, C. (1982). *Industrial and business forecasting methods*. Butterworths. London.
- Li, G., Song, H., & Witt, S. (2005). Recent developments in econometric modeling and forecasting. *Journal of Travel Research*, 44(1), 82-99.
- Lim, C. (1997). Review of international tourism demand models. *Annals of Tourism Research*, 24(4), 835-849.
- Lim, C., & McAleer, M. (2002). Time series forecasts of international travel demand for Australia. *Tourism Management*, 23(4), 389-396.
- Machado, T., Teixeira, J., & Fernandes, P. (2010). Modelação da procura turística em Portugal: regressão linear versus redes neuronais artificiais. *Revista Turismo & Desenvolvimento*, 1, 435-445.
- Marôco, J. (2014). *Análise Estatística com SPSS Statistics*. Edições Sílabo, Lda. (6.a Ed.).
- OANDA. (2015). *Taxas de câmbio Históricas*. Acedido em Março de 2015, disponível on-line em <http://www.oanda.com/lang/pt/currency/historical-rates/>.
- Palmer, A., Montano, J., & Sesé, A. (2006). Designing an artificial neural network for forecasting tourism time series. *Tourism Management*, 27(5), 781-790.
- Peng, B., Song, H., Crouch, G., & Witt, S. (2014). A meta-analysis of International tourism demand elasticities. *Journal of Travel Research*, 1-23.
- Pestana, M., & Gageiro, J. (2005). *Descobrimos a regressão: com a complementaridade do SPSS*. Edições Sílabo, Lda. (5.a Ed.).
- Preez, J., & Witt, S. (2003). Univariate versus multivariate time series forecasting: an application to international tourism demand. *International Journal of Forecasting*, 19, 435-451.
- Rodrigues, P. M. & Teixeira, João Paulo (2010). Classification of Electroencephalogram Signals Using Artificial Neural Networks. Proceedings of 3rd International Conference on BioMedical Engineering and Informatics (BMEI'10). Yantai – China.
- Santos, N., & Fernandes, P. (2010). *Análise e Previsão da Procura Turística na Região Norte de Portugal*. Proceedings of APDR; Madeira, Portugal.
- Santos, N., & Fernandes, P. (2011). Modelação e caracterização da procura turística: o caso da região Norte de Portugal. *TÉKHNE-Polytechnical Studies Review*, 9(16), 118-137.
- Sivanandam, S., & Paulraj, M. (2003). *Introduction to Artificial Neural Networks*. Vikas Publication India.
- Song, H., & Witt, S. (2000). *Tourism demand modelling and forecasting. Modern econometric approaches*. Routledge.
- Song, H., Li, G., Witt, S., & Fei, B. (2010). Tourism demand modelling and forecasting: how should demand be measured?. *Tourism Economics*, 16(1), 63-81.
- Song, H., & Witt, S. (2012). *Tourism demand modelling and forecasting*. Routledge.
- Song, H., Witt, S., & Li, G. (2003). Modelling and forecasting the demand for Thai tourism. *Tourism Economics*, 9(4), 363-387.
- SSA (2014). *Statistics South Africa*. Acedido em Março de 2015, disponível on-line em: <http://www.statssa.gov.za>.
- Tribe, J., & Xiao, H. (2011). Developments in tourism social science. *Annals of Tourism Research*, 38(1), 7-26.

- Uysal, M., & Roubi, M. (1999). Artificial neural networks versus multiple regression in tourism demand analysis. *Journal of Travel Research*, 38(2), 111-118.
- Witt, S., & Witt, C. (1995). Forecasting tourism demand: A review of empirical research. *International Journal of Forecasting*, 11, pp 447-475.
- Zhang, G. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159-175.
- Zhang, G., & Qi, M. (2005). Neural networks Forecasting and trend time series. *European Journal of Operational Research*, 160, 501-514.
- Zhang, G., Patuwo, B., & Hu, M. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14(1), 35-62.