Forecasting of a Non-Seasonal Tourism Time Series with ANN

Teixeira, João Paulo[1] & Fernandes, Paula Odete[1,2]

[1]Polytechnic Institute of Bragança; UNIAG
[2]NECE
joaopt@ipb.pt; pof@ipb.pt

Abstract. The paper present and discusses several alternative architectures of Artificial Neural Network models used to predict the time series of tourism demand for Cape Verde. This time series is particularly difficult to predict due to its non-seasonal characteristic usual in a similar time series for European Tourism destinations. The time index used as input and other input parameters variations improved the performance of the prediction over the test set to a relative error of 7.3% and a Pearson correlation coefficient of 0.92.

Keywords: ANN; Forecast; Non-seasonal time series; Tourism; Cape Verde.

1 Introduction

Modelling and forecasting tourism demand has received in the last decade's a substantial attention among researchers, policy makers, hospitality management, and other interest groups. Forecasting tourism demand being a significant activity for its beneficiaries and stakeholders', several forecasting models have been applied to estimate and forecast the tourism demand.

Tourism demand modelling and forecasting methods can be divided into two categories: quantitative and qualitative methods. The majority of the published studies used quantitative methods to forecast tourism demand (Son & Li, 2008). The quantitative forecasting is dominated by two sub-categories of methods: the causal econometric models and non-causal time series models. The difference among them is whether the forecasting model identifies any causal relationship between the tourism demand variable and its influencing factors. The casual econometric models found in the literature most frequently used are: cointegration and error correction (ECM) models (Algieri, 2006; Dritsakis, 2004), time varying parameter (TVP) models (Song & Wong, 2003), structural equation (SEQ) models (Turner & Witt, 2001), vector autoregressive (VAR) models (Song & Witt, 2006) and linear almost ideal system (LAIDS) models (Han et al., 2006). These methods have also been combined (Li et al., 2006). On the other hand, the most commonly used procedures in non-causal time series forecasting are the autoregressive integrated moving average (ARIMA) models (Goh & Law, 2002; Kulendran & Shan, 2002; Fernandes et al., 2008) and the exponential smoothing (ES) models (Cho, 2003). Recently, artificial intelligence (AI)

adfa, p. 1, 2011.

methods have also been implemented in tourism forecasting. The most commonly used AI methods are artificial neural network (ANN) models (Kon & Turner, 2005; Palmer et al., 2006; Fernandes et al., 2008; Fernandes et al., 2013). But, according to Song and Li (2008) there is no one model that stands out in terms of forecasting accuracy. Therefore academics and practitioners continue to improve and develop models and methods to bring about greater understanding of the economics and business principles as guidance for more effective management and planning in the tourism sector.

In this way, with this paper it is intended to present and discuss several alternative architectures of Artificial Neural Network models used to predict the time series of tourism demand for Cape Verde. For that it will be used the time series "Monthly Guest Nights in Hotels in Cape Verde" registered between the period January 2005 and December 2011, and collected by National Statistics Office of Cape Verde (INECV, 2013). It is one of the most used time series in the academic and scientific field, to describe the tourism demand.

The remaining sections of this paper are organized as follows: After this introduction; in section 2 the time series "Monthly Guest Nights in Hotels in Cape Verde" is presented and analysed; Artificial Neural Network models will be presented in section 3; the empirical results are given in section 4. A concluding discussion finishes the paper.

2 Tourism Demand Time Series of Cape Verde

The archipelago of Cape Verde is made up of ten volcanic islands (Santo Antão, São Vicente, Santa Luzia, São Nicolau, Sal, Boavista, Maio, Santiago, Fogo and Brava) and eight islets which make up a total area of 4033 km² (PEDTCV, 2010). There has been an exponential increase in the number of overnight stays which, in 2000, totalled 685 thousand and in 2012 already exceeded 3 million (INECV, 2013). Over this period the average annual increase in overnight stays in hotel establishments in the country was around to 14%, while, last year, the number increased by 17.9% in relation to 2011 (INECV, 2013). There continues to be strong reliance on foreign tourism – over 95% of overnight stays – almost exclusively represented by European markets. The most important markets in 2012 were: the United Kingdom, 31.7%; Germany, 14.9%; Portugal, 9.5%; and France, 9.0%. The islands of Boavista and Sal stand out clearly as being the destinations most capable of attracting tourists, representing 90% of overnight stays; 47.4% on the island of Boavista and 42.2% on the island of Sal (INECV, 2013).

The follow Fig. 1 presents the number of overnights of tourist that arriving and staying overnight at any of the ten islands in the Archipelago. The monthly overnights were registered between the period January 2005 and December 2011, a total of 84 observations, and collected by National Statistics Office of Cape Verde (INECV, 2013).

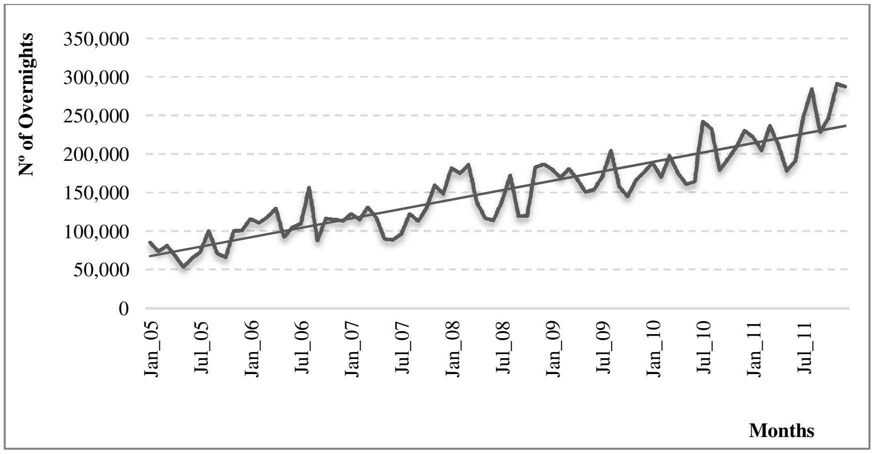

Fig. 1. Monthly Guest Nights in the Cape Verde, from 2005:01 to 2011:12.

Is possible to see clearly the non-seasonality of the time series, that is no common when compared with other tourism demand time series, along with a clear upward trend. However, it is noteworthy that the month of August and in some years the month of December presents the highest peaks.

An analysis using the Fourier transform of the time series is presented in Fig. 2. The Fourier transform was performed with a length N of 256. Since each sample corresponds to one month, the sampling frequency Fs is the inverse of one month. Therefore the 256 sample of Fourier transform corresponds to Fs frequency. In Fig. 1 only the first 128 samples (corresponding to Fs/2) of Fourier transform is presented and scaled to 6. This means that the visible peak that appears at 3, corresponds to frequency of half of Fs/2, meaning Fs/4 that is the inverse of 4 months. This corresponds that is usual to find 3 peaks per year or that there are a repetition of periods with 4 month long in this time series.

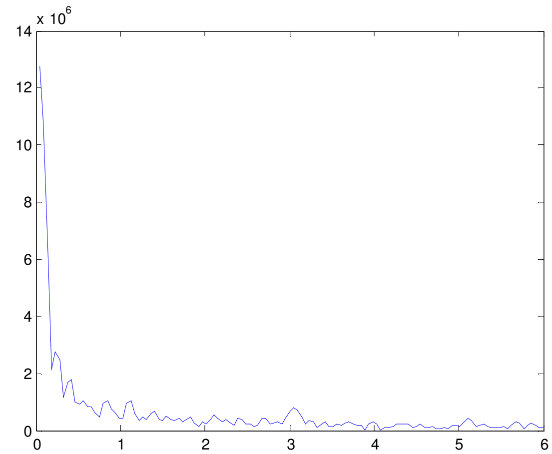

Fig. 2. Fourier analysis of the time series.

3 Artificial Neural Network Models

Due to the non classical time series that makes difficult the forecasting of future values because non seasonal characteristics seem to be present in the time series several experimental architectures and input data will be used in models in order to evaluate the possibility of make predictions with some degree of confidence. The experimented ANN models already have been used to predict a similar tourism time series but for other country and/or region of country with good results. Anyhow it should be mention that all time series used in those models has the characteristic of seasonality.

All the ANN used has one output node to predict the value of the time series for next month but the model should predict the values for next twelve months. With the ANNs with one output the prediction for twelve month is made in 12 iterations of prediction using the output of previous iteration as input for next prediction.

The following 5 models and inputs have been experimented:

- Model A – consists in a feedforward architecture ANN with 12 input nodes one output node and 4, 6 and 8 nodes in the hidden layer. The input consists in the values of the previous twelve months. The output is the forecast for next month. This model already has been used with success result for the time series of the North Region of Portugal (Fernandes et al., 2013).
- Model B – consists in a feedforward architecture ANN using the time index in its entrance. The ANN has 14 inputs, and one output and 4, 6 and 8 nodes in the hidden layer. The time index is modelled in two variables one for the month and the other for the year. The month varies from 1 to 12 corresponding to the month of the year (1 – January ... 12 – December). The year varies from 1 for the values of

the first year (2005) until 7 for the last year (2011). This model already has proved its ability to forecast a similar time series with a strong rising tendency (Fernandes & Teixeira, 2009).

- Model C – consists in a feedforward architecture ANN using only the variable with the year in its entrance. This model is very similar to model B but discard the variable month because the time series seem to do not be sensitive to the month. Anyhow the year is kept because of the rising tendency of the time series. The remaining inputs, output and hidden layer are the same as the ones of models B.
- Model D – consists also in a feedforward ANN using the year in its entrance but now using only the previous 5 months in the input. Therefore the model D ANN has 6 input nodes only and the remaining is similar as model C. This experiment intends to verify the importance or not of one year of data in the input.
- Model E – consists in a feedforward ANN similar as model B but using the time series in the logarithmic domain. This last model has used in order to verify if reducing the amplitude differences between months it could improve the forecast error. The architecture and input features were the same as model B, because after a first analysis model B had lower forecast error.

The five ANN models had been experimented with different number of nodes in the hidden layer. Four, six and eight (4, 6 and 8) nodes were experimented for each model. The available data of the time series is relatively short to consider more nodes in the hidden layer, because it would increase the number of weights to be adjusted during training process. Considering, for instance the model A with 12 inputs, and M the number of hidden nodes, the total number of weight to be adjusted are 12*M+M*1. Additionally there are also the bias weights that are also M+1. For M=8 it gives 104 weight and 9 bias. Considering only 64 input/output pair it is too much weight parameters to be adjusted.

The data set was divided in a training, validation and test set. The test set was used for a final evaluation of the model performance and consists in the last 12 month, corresponding to the last year (2011). The validation set was used for a cross validation process to early stop training. The set corresponds to the 12 months of the year 2010. The training set, used to train the ANN, corresponds to remaining data in a total of 60 months. Actually only 48 months were available to train for models A, B, C and E because the previous 12 months are need for the input. For model D 55 month were used for training because only 5 months are need for the input.

All the ANN models were trained with the Levenberg-Marquardt training function (Marquardt, 1963). The Resilient back propagation algorithm (Riedmiller & Braun, 1993) has also experimented but with worst results.

Each model was trained 50 times and selected the one with lower Mean Relative Error (MRE) in validation set, given by eq. 1. The results in the test set were not seen during this process.

4 Discussion of Results and Models

The quality of prediction is measured by the ability of the model to fit the curve of target data. This ability is analysed by the MRE and by the Pearson Correlation coefficient (***r***) between target and predicted data set. The MRE measures the distance between target and predicted values of a given set. The correlation coefficient measures the similarity between the two curves of the given set. It is possible to have a high similarity between the two curves (high ***r***) but the curves can be very different in their absolute values. The correlation coefficient can evaluate the ability of the model to catch the variation along the months of the year and MRE can evaluate the ability of the model to follow the accurate amplitude of the time series.

Target data is the original values of the time series.

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{T_i - P_i}{T_i}\right| \qquad [1]$$

with: P_i, Predicted value for month i and T_i, Target value for month i.

The MRE and correlation coefficient along test set, validation set and considering the all data for the 4 models using 4, 6 and 8 nodes in the hidden layer is presented in table 1.

The best models considering the all data set has model B with 8 nodes in hidden layer that reached the lower MRE (7.7%), and model C with 4 nodes in hidden layer that reached the higher r (0.92).

Considering the validation set clearly model A with 4 nodes in hidden layer achieved simultaneously lower MRE (4.6%) and higher r (0.82).

For the test set model C with 8 nodes in hidden layer has again the one with best result both considering MRE (7.3%) and r (0.85). Though model E with 8 nodes in hidden layer also reached the same r (0.85).

The result considering the All data set cannot be taken very seriously because most of the data were used in training process been used to adjust weights. Validation set was used during training process to stop train early and to select the best training session along 50 training session. And data of the test set were not seen during training process. Therefore the model with better performance along this set should be considered the selected one.

Model B with 8 nodes in hidden layer reached the better performance in test set and simultaneously the better MRE considering the all data set. Considering the All data set this model also reached a correlation coefficient r (0.91) at the level of the best value (0.92). Therefore it should be considered the model more adequate to predict this time series.

Table 1. Performance of models.

Model	Hidden nodes	Test		Validation		All	
		MRE (%)	***r***	MRE (%)	***r***	MRE (%)	***r***
Mod. A	4	11.3	0.44	**4.6**	**0.82**	11.8	0.88
	6	14.8	0.18	8.0	0.69	9.7	0.88
	8	12.2	0.73	8.5	0.67	10.4	0.88
Mod. B	4	12.1	0.57	9.5	0.60	11.1	0.89
	6	10.1	0.58	9.1	0.64	9.5	0.90
	8	**7.3**	**0.85**	8.7	0.67	**7.7**	0.91
Mod. C	4	10.4	0.66	6.4	0.81	8.4	**0.92**
	6	11.9	0.56	7.7	0.47	12.7	0.87
	8	15.1	0.47	10.1	0.57	11.7	0.91
Mod. D	4	19.0	0.51	10.8	0.16	19.4	0.81
	6	21.1	0.60	8.4	0.46	13.5	0.85
	8	26.0	0.05	8.7	0.60	16.3	0.76
Mod. E	4	15.9	0.79	9.4	0.67	10.8	0.88
	6	10.6	0.70	7.2	0.59	11.9	0.87
	8	8.1	**0.85**	7.2	0.53	10.2	0.91

Comparing the best results of the model B for this non-seasonal time series with the best performance of ANN for seasonal time series, this model reached a MRE of 7.3% against an MRE of level between 5.4% and 7.3 % using a model similar to model A in Fernandes et al. (2013) or 5.8% and 6.4% using models similar to present models B and A, respectively in Fernandes and Teixeira (2009). Considering the correlation coefficient ***r*** this model reach 0.85 in this non seasonal time-series against 0.98 in Fernandes et al. (2013) using a seasonal time series.

Figures 3, 4, 5 and 6 present the ANN output and target time series using Model A with 4 nodes in hidden layer, Model B with 8 node in hidden nodes, Model C with 4 nodes in hidden layer and Model E with 8 nodes in hidden layer, respectively. These models were the ones with best results in some of the data set. The first 12 values of the target and output are equal because no output exists for these months used as inputs for next month. The last 12 values correspond to the 12 months of the test set. The previous 12 values correspond to the 12 months of the validation set. The remaining values are the training data set. Generally the presented models show a relatively good but not perfect fitting in training data set. It is remarkable that the models can follow very well the rising of the time series along years.

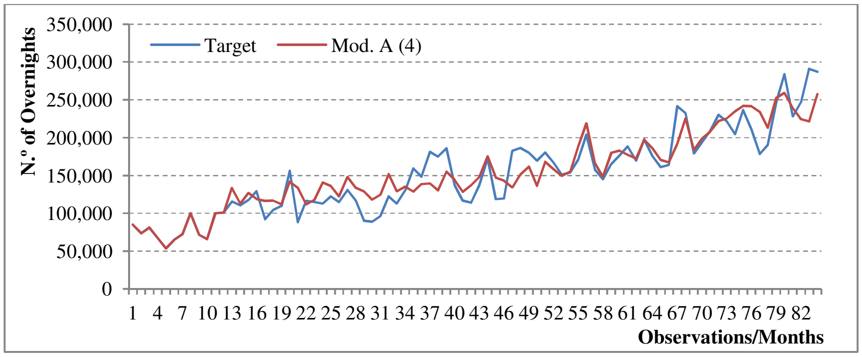

Fig. 3. Output of model A with 4 hidden nodes.

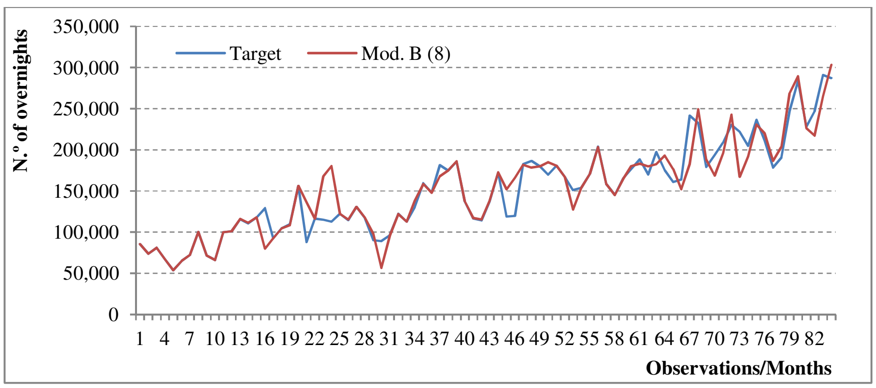

Fig. 4. Output of model B with 8 hidden nodes.

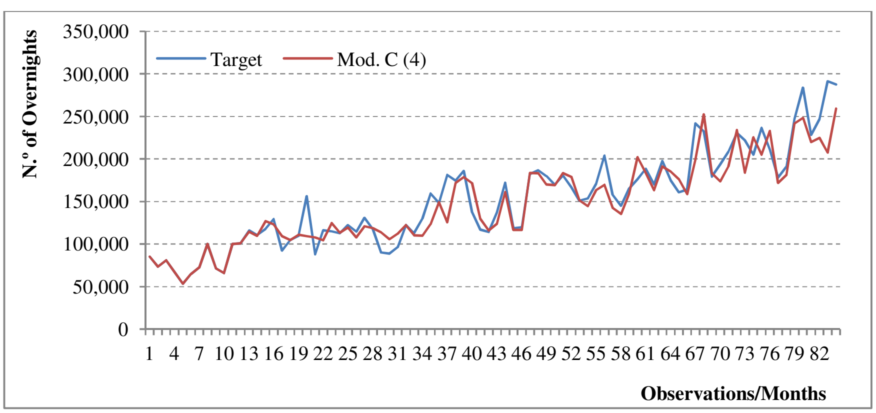

Fig. 5. Output of model C with 4 hidden nodes.

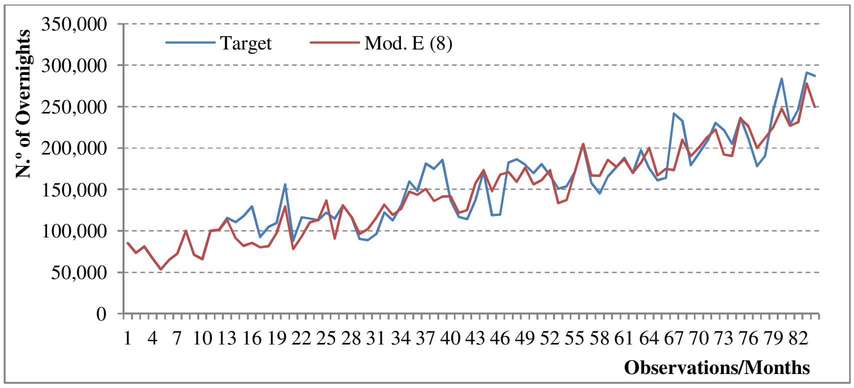

Fig. 6. Output of model E with 8 hidden nodes.

Fig. 7 presents the results of the same models but only in the 12 months of the test set. It can be seen that the models captured the 3 peaks along the year. Model B are the one with values more close to the target ones.

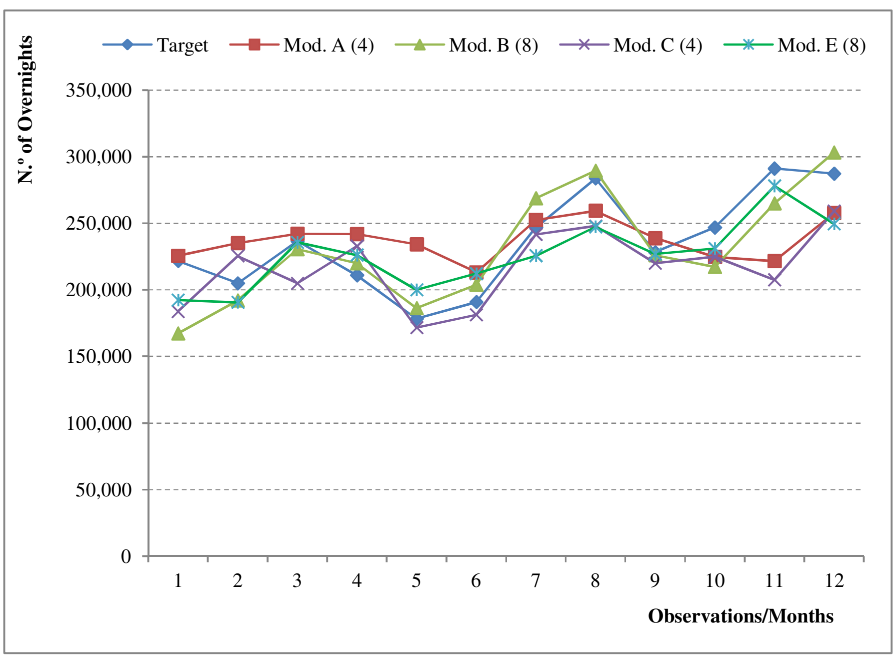

Fig. 7. Output of test set with models A, B, C and E.

5 Conclusion

The tourism time series used in this work contrary to usual tourism time series is non seasonal. This makes more challenging the task to predict future values. ANN models are adequate to solve no linear problems and were used to make predictions of this time series for twelve months of next year. Five different architectures/inputs were experimented. Model A with the 12 previous months in its input, Model B with two additional inputs for time index, Model C with only one input for the year of the time index, Model D using only 5 previous months and time index and model E with same inputs as model B but in logarithmic domain. Different number of nodes in hidden layer was experimented.

Model B with 8 nodes in hidden layer was selected as the best model achieving an MRE of 7.3% and a correlation coefficient of 0.85 in test set.

Comparing the results of this model for this non seasonal time series with similar models for a seasonal time series, Model B achieved a similar MRE but worst correlation coefficient denoting the difficulty of fit the curve of this time series.

Although the apparently random values of this time series the ANN model achieved a prediction with an error that can be considered at a good level according to (Lewis, 1982).

Model C is similar as Model B but was only one variable to model the year instead of two variables to model month and year. The results showed clearly worst results for this model C. Therefore to model the time index it is actually necessary this two variables.

The input of model D has only the previous 5 months instead of 12. The results for this model were very bad denoting the real need to use the previous 12 months in the input of the model.

Model E is the same as model B, but the time series were converted to the logarithm domain. The results showed no improvement with model E. Therefore there is no advantage to use the log domain.

References

Algieri, B. (2006). An econometric estimation of the demand for tourism: the case of Russia. *Tourism Economics*, 12, 5-20.

Cho, V. (2003). A comparison of three different approaches to tourist arrival forecasting. *Tourism Management*, 24, 323-330.

Dritsakis, N. (2004). Cointegration analysis of German and British tourism demand for Greece. *Tourism Management*, 25, 111-119.

Fernandes, P., & Teixeira, J. (2009). *New Approach of the ANN Methodology for Forecasting Time Series: Use of Time Index*. Proceeding of ICTDM, Kos-Greece.

Fernandes, P., Teixeira, J., Ferreira, J., & Azevedo, S. (2008). Modelling Tourism Demand: A Comparative Study between Artificial Neural Networks and the Box-Jenkins Methodology. *Romanian Journal of Economic Forecasting*, 5(3), 30-50.

Fernandes, P., Teixeira, J., Ferreira, J., & Azevedo, S. (2013). Training Neural Networks by Resilient Backpropagation Algorithm for Tourism Forecasting. In J. Casillas, F. Martínez-López, R. Vicari, F. De la Prieta (Editors), *Management Intelligent Systems, Advances in Intelligent Systems and Computing*, 220, 41-49.

Goh, C., & Law, R. (2002). Modelling and forecasting tourism demand for arrivals with stochastic nonstationarity seasonality and intervention. *Tourism Management*, 23, 499-510.

Han, Z., Dubarry, R., & Sinclair, M. (2006). Modelling US tourism demand for European destinations. *Tourism Management*, 27, 1-10.

INECV (2013). Anuários Estatísticos. Assessed on March 2013.

Kon, S., & Turner, W. (2005). Neural network forecasting of tourism demand. *Tourism Economics*, 11, 301-328.

Kulendran, N., & Shan, J. (2002). Forecasting China's monthly inbound travel demand. *Journal of Travel and Tourism marketing,* 13 (1/2), 5-19.

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

Li, G., Song, H., & Witt, S. (2006). Time varying parameter and fixed parameter linear AIDS: an application to tourism demand forecasting. *International Journal of Forecasting*, 22, 57-71.

Marquardt, D. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal on Applied Mathematics*, 11(2), 431-441.

Palmer, A., Montaño, J., & Sesé, A. (2006). Designing an artificial neural network for forecasting tourism time series. *Tourism Management*, 27, 781-790.

PEDTCV (2010). Plano Estratégico para o Desenvolvimento do Turismo em Cabo Verde. Assessed on March 2013.

Riedmiller, M., & Braun, H. (1993*). A direct adaptive method for faster back-propagation learning: The RPROP algorithm.* Proceedings of the IEEE International Conference on Neural Networks.

Song, H., & Li, G. (2008). Tourism demand modelling and forecasting - a review of recent research. *Tourism Management*, 29, 203-220.

Song, H., & Witt, S. (2006). Forecasting international tourist flows to Macau. *Tourism Management*, 27, 214-224.

Song, H., & Wong, K. (2003). Tourism Demand Modeling: A Time-Varying Parameter Approach. *Journal of Travel Research*, 42, 57-64.

Turner, L., & Witt, S. (2001). Factors influencing demand for international tourism: tourism demand analysis using structural equation modeling revisited. *Tourism Economics*, 7, 21-38.