# SARIMA and Exponential Smoothing model for forecasting ecotourism demand: A case study in National Park Kuala Tahan, Pahang

# SARIMA and Exponential Smoothing model for forecasting ecotourism demand: A case study in National Park Kuala Tahan, Pahang.

**Noratikah Abu[1], Wan Nur Syahidah[2], Megat Muhammad Afif[3], Syarifah Zyurina Nordin[4]**

[1,2,3]Centre for Mathematical Sciences, College of Computing and Applied Sciences, Universiti Malaysia Pahang, 26300 Gambang, Kuantan, Pahang, Malaysia.
[4]Malaysia-Japan International Institute of Technology, University Teknologi Malaysia, Jalan Sultan Yahya Petra, Kampung Datuk Keramat, 54100 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia.

[1]**Email:** atikahabu@ump.edu.my

**Abstract**. Tourism forecasting can lead to an important element in tourism industry to ensure that each investment by individuals, companies and government is profitable. From economy perspective, eco- tourism is a growing business, and it is an important indicator to the tourism industry. It also generates income revenue to the owner and surrounding communities. This research aims to forecast the eco-tourism demand based on number of tourist arrival for both local and foreign tourist at National Park Kuala Tahan, Pahang. The forecasting models used are seasonal autoregressive integrated moving average (SARIMA) and exponential smoothing. Both forecasting models are compared and assessed using mean absolute percentage error (MAPE), root mean square error (RMSE) and mean absolute error (MAE). The result demonstrated that SARIMA $(1,0,0)(1,0,1)_{12}$ the best model to forecast the number of tourist arrival in National Park Kuala Tahan, Pahang is which give the smallest forecast evaluation values. Hence, the exponential smoothing is not as good as the SARIMA model in forecasting tourist arrival for the data used. In future study, SARIMA model can be used to compare the local and foreign tourist arrival for eco-tourism destination.

## 1. Introduction

Tourism industry is one of the main economy indicators in the world and is growing very fast from day to day. Therefore, tourism forecasting can lead to an important element in the tourism industry to ensure that each investment by individuals, companies and government was profitable. The country can inhibit gains with fewer errors in determining the total number of tourists. According to The World Tourism Organization (UNWTO), tourism can be classified to 14 types such as cultural, ecotourism, rural, adventure, health, wellness, medical, business, gastronomy, urban and city, mountain, education, sports, and coastal, maritime and inland water tourism. In this study, we will focus on ecotourism industry as it received less attention among researchers. Ecotourism involved of natural, communities and sustainable travel. The International Ecotourism Society (TIES) defined ecotourism as responsible tourist to natural areas that maintain the environment, sustain the well-being of the local people and includes education. Pahang, which is located at East Coast Malaysia and the population around 5.1% from total Malaysian

population always, be the tourist attraction because of its natural beauty of flora and fauna, covered by thousands of hectares of rainforest. Estimated exist more than 130 million year's old, national park located at Jerantut, gazetted as National Park since 1938, it is the largest national park more than 4000 square kilometers. The beautiful and unique natural scenery undeniable natural lovers around world. Based on this information, it can be seen that nature is the main source of tourism product in Pahang which lead to ecotourism product. According to [1], the number of tourist arrivals is the most popular measure for tourism demand forecasting. Hence, this study proposes to forecast the ecotourism product in Pahang based on number of tourist arrival in National Park Kuala Tahan. The first section of this paper will present the introduction which the problem and background of the research. Then, the selected literature review is reviewed. Next, we will discuss the methodology used and analyses the findings and results.

## 2. Literature review

Tourism forecasting can lead to an important element in the tourism industry to ensure that each investment by individuals, companies and government was worth it. Past research has shown that modelling and forecasting tourism demand is numerous with various type of empirical analysis. [2] forecast international tourism demand in ASEAN countries, [3] forecast the demand for Hong Kong tourism, and [4] forecast number of tourist arrival in Malaysia. However, past research on ecotourism demand forecasting is limited. [5] studied on construction and application of forecast model for ecotourism market demand using annual data from 2008 until 2016 while [6] estimated economic value for ecotourism product in west Sumatera, Indonesia. Hence, this study will focus on forecasting ecotourism demand.

SARIMA method is one of the modelling approaches in forecasting tourism industry. Several studies have applied this method [7,8,9]. [7] forecast foreign tourist visits entering Bandung airport from 2010 to 2017 using SARIMA model and found that the model can be the best model to predict foreign tourist with good accuracy. [8] applied forecasting international tourist arrivals to Sri Lanka using SARIMA approach and found that SARIMA models fit tourism data well with lowest value of MAPE. [9] used SARIMA model in forecasting monthly tourist arrivals from ASEAN countries from January 2000 until December 2014. Then, [4], forecast international tourism demand in Malaysia by using Box-Jenkins SARIMA model and concluded smoothing techniques should be included to forecast tourism demand. Hence, we will apply the exponential smoothing technique to compare the forecast accuracy. The exponential smoothing has been used by many researchers in forecasting [10,11,12]. [10] examined that exponential smoothing models is a widely used method in time series analysis. Three forecast accuracy techniques had been used to select the most accurate forecast. [11] forecast annual international tourist arrivals in Zambia using Holt-Winter's exponential smoothing model and compared with ARIMA model. Results showed that Holt-Winter's model provide the smallest error. [12] compared Holt-Winter's and ARIMA for forecasting tourist arrival in India for the period 1981 to 2014. Hence, based on these literature review, we propose to forecast the ecotourism demand in Pahang using SARIMA and exponential smoothing model.

## 3. Methodology

### *3.1. Data*

This study used a 5-year monthly data of tourist arrival at National Park Kuala Tahan, Pahang. The data is obtained from Department of Wildlife and National Parks Peninsular Malaysia. This study focuses on total number of tourists entered the National Park Kuala Tahan from January 2013 until December 2017.

### *3.2. Methods*

*3.2.1. SARIMA.* Seasonal autoregressive integrated moving average (SARIMA) model is generally referred to as SARIMA$(p,d,q)(P,D,Q)_s$. This seasonal model is extended from autoregressive integrated moving average (ARIMA) and designed for seasonal series. $p$ and $P$ are the orders of autoregressive, $d$ and $D$ are the differences, and $q$ and $Q$ are the orders of moving average operator of non-seasonal and seasonal components, respectively. If the length of seasonal period is $n$, so that $n = 4$ for quarterly data and $n = 12$ for monthly.

The SARIMA model will follows the basic Box-Jenkins model procedure consist of four steps. The first step is model identification where the model is identify using training data. Next, ordinary least square method will be used to estimate the parameters in the model. The model with the significant parameters will be chosen to the next step. Then, after we get the selected model, diagnostic checking will be conducted to using Ljung-Box test to check the residuals. Finally, the selected model will be used to forecast the number of tourist arrival. Testing data will be used to evaluate the model.

*3.2.2. Exponential Smoothing.* Exponential smoothing models are procedure for continually revising a forecast in a light of more recent experience. Exponential smoothing models also most effective techniques when the parameters describing the time series may be changing slowly over time. There are three types of exponential smoothing; simple exponential smoothing for data without trend and seasonal, double exponential smoothing for data with trend and Holt-Winter's method for data with seasonal trend. In this study, we will used Holt-Winter's method. The prediction formula as follow:

$$\hat{y}_{t+1}(T) = [a_0(T) + b_1(T)\tau]sn_{T+1}(T+1-L) \tag{1}$$

where

$$a_o(T) = \bar{y}_1 - \frac{L}{2}b_1(T)$$

$$b_1(T) = \frac{\bar{y}_m - \bar{y}_1}{(m-1)L}$$

$$sn_t(T) = \gamma\frac{y_t}{a_0(T)} + (1-\gamma)sn_T(T-L)$$

In this method $\bar{y}_m$ is the average of the observations in year m. While $\bar{y}_1$, measure the average level of the time series in the middle of year 1. Then, $m$ refers to the year $m$ and $L$ is number of seasons.

### *3.3. Forecast Accuracy*

Then, for further examine of the forecasting accuracy for the ecotourism demand, the testing data and the actual data will be compared. Next, the data will be analysed using both methods. Forecast evaluation method that will be used to measure the forecasting accuracy and validate the forecasting model is mean absolute percentage error (MAPE), root mean square error (RMSE) and mean absolute error (MAE).

## 4. Results and Discussions

In this study, the number of tourist arrival at National Park Kuala Tahan from January 2013 to December 2016 will be the training data with 48 observations and number of tourist arrival from January 2017 to

December 2017 with 12 observations will be the testing data. In time series, training data is used to estimate model, whereas the testing data is used to validate model. Figure 1 shows the number of tourist arrivals for training data.

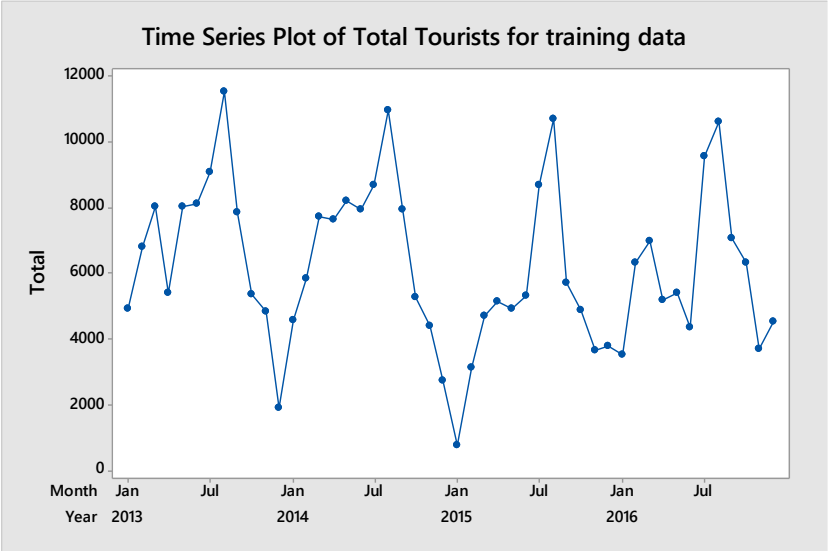


**Figure 1**. Training data for number of tourist arrival in National Park Kuala Tahan

Based on Figure 1, there is a large jump in number of tourists in July each year and shows that there were an unusually small number of tourists in December every year. Hence, this data will be analysed using seasonal model; SARIMA and Holt-Winters' exponential smoothing model.

Table 1 tabulates the results of forecast evaluation for SARIMA models. Based on empirical evidence, it indicates that SARIMA $(1,0,0)(1,0,1)_{12}$ was selected as the best model for SARIMA since the value of MAPE are the lowest as compared to the other SARIMA models. Then, this SARIMA model is applied to testing data consists of 12 observation. The testing data is used to obtain fitted forecast to calculate the forecast evaluation.

**Table 1**. Forecast evaluation result for model of SARIMA

| Forecast accuracy | SARIMA $(1,0,0)(1,0,0)_{12}$ | SARIMA $(1,0,0)(1,0,1)_{12}$ |
|---|---|---|
| MAPE (%) | 24 | 20 |
| RMSE | 7755 | 7853 |
| MAE | 764 | 820 |

Table 2 presents the comparison of forecast evaluation for SARIMA and Holt-Winter's models. Based on the result of forecast evaluation values in Table 2, it shows that SARIMA $(1,0,0)(1,0,1)_{12}$ is the best forecasting model to forecast the number of tourist's arrival in National Park Kuala Tahan, Pahang as the MAPE value is 5% which is smaller than Holt-Winter's model 39%. Besides that, RMSE value for SARIMA is 949 while Holt-Winter's 7378 and MAE shows 222 for SARIMA and 1810 for Holt-Winter's.

**Table 2.** Forecast evaluation values for testing data

| Forecast Evaluation | SARIMA $(1,0,0)(1,0,1)_{12}$ | Holt-Winter's |
|---|---|---|
| MAPE (%) | 5 | 39 |
| RMSE | 949 | 7378 |
| MAE | 222 | 1810 |

Figure 2 shows the comparison of testing and actual data for number of tourist arrival with SARIMA $(1,0,0)(1,0,1)_{12}$ and Holt-Winter's. Based on Figure 2, the gap for data fitted from SARIMA model and actual data is minimum compared to Holt-Winter's. While for Holt-Winter's model, there are several huge differences with the actual data. Therefore, graphically SARIMA $(1,0,0)(1,0,1)_{12}$ model was chosen as the best model for forecasting ecotourism demand in Pahang based on tourist arrival in National Park Kuala Tahan.

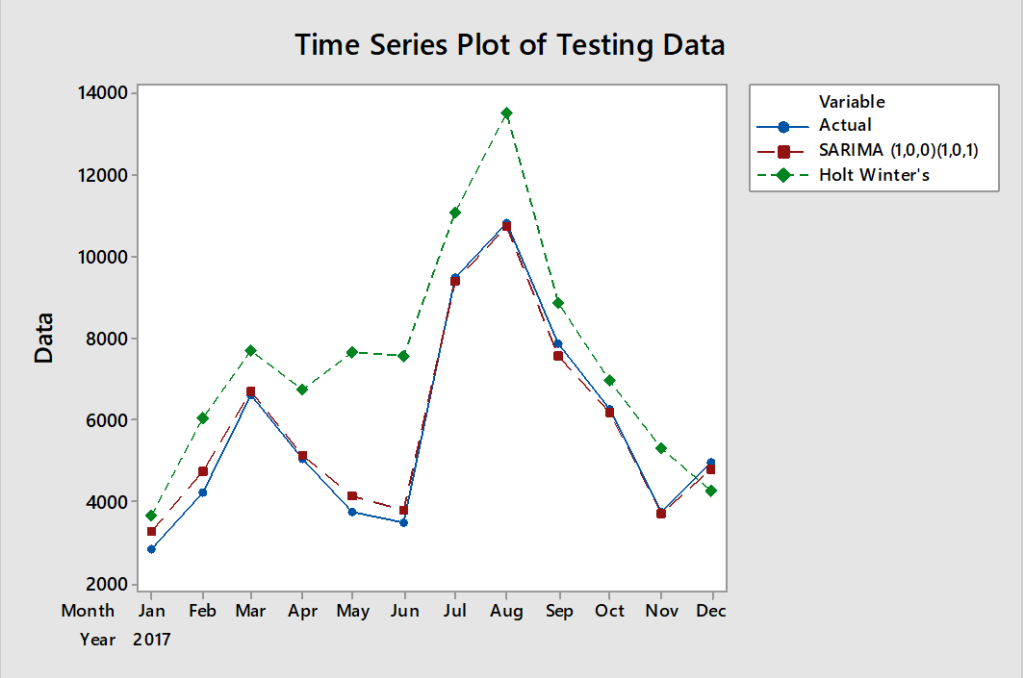


**Figure 2.** Actual and forecasted testing data number of tourist arrival in National Park Kuala Tahan

## 5. Conclusions

In this study, the best model for forecasting ecotourism demand in Pahang were presented. The data was analysed using SARIMA and exponential smoothing model. Based on the analysis, the best forecasting model to forecast number of tourist arrival at National Park Kuala Tahan, Pahang is SARIMA model. SARIMA model give the smallest error value compared to exponential smoothing. In future study, SARIMA model can be used to compare between the local and foreign tourist arrival for eco-tourism destination.

**Acknowledgement**

This research is supported by Centre for Mathematical Sciences, Universiti Malaysia Pahang and funded under UMP internal grant RDU1903124. The supports are gratefully acknowledged.

## References

[1] Song, H., and Li, G. (2008). Tourism Demand Modelling and Forecasting-A Review of Recent Research. Journal of Tourism Management, 29, 203-220.
[2] Karimi, A., Faroughi, P. and Rahim, K. A (2015). Modeling and Forecasting of International Tourism Demand in ASEAN Countries. American Journal of Applied Sciences, 12(7), 479-486.
[3] Song, H., Gao, B. Z. and Lin, V. S. (2013). Combining Statistical and Judgmental Forecasts Via a Web-Based Tourism Demand Forecasting System. International Journal of Forecasting, 29, 295-310.
[4] Loganathan, K. and Ibrahim, Y. (2010). Forecasting International Tourism Demand in Malaysia using Box-Jenkins Sarima Application. South Asian Journal of Tourism and Heritage, 3(2), 50-60.
[5] Chen, H. (2019). Construction and Application of Forecast Model for Eco-Tourism Market Demand Based on BP Neural Network. The Frontiers of Society, Science and Technology, 1(6), 87-95.
[6] Iasha, A., Yacob, M. R., Kabir, I. and Radam, A (2015). Estimating Economic Value for Potential Ecotourism Resources in Puncak Lawan Park, Agam District, West Sumatera, Indonesia. International Conference on Environmental Forensics 2015, 30, 326-331.
[7] Fahrudin, R. (2018). Forecasting Tourist Visits using Seasonal Autoregressive Integrated Moving Average Method. IOP Conf. Series: Materials Science and Engineering, 407.
[8] Thushara, S. C., Su, J. J. and Bandaralage, J. (2017). Forecasting International Tourist Arrivals to Sri Lanka using SARIMA Approach. GI Social Sciences Forum, 1, 13-23.
[9] Rufino, C. (2016). Forecasting Monthly Tourist Arrivals from ASEAN+3 Countries to The Philippines for 2015-2016 using Sarima Noise Modeling. DLSU Research Congress.
[10] Ostertagova, E. and Ostertag, O. (2012). Forecasting Using Simple Exponential Smoothing Method. Acta Electrotechnica et Informatica, 12(3),62–66.
[11] Jere, S., Banda, A., Kasense, B., Siluyele, I. and Moyo, E. (2019). Forecasting Annual International Tourist Arrivals in Zambia using Holt-Winters Exponential Smoothing. Open Journal Statistics, 9, 258-267.
[12] Sood, S. and Jain, K. (2017). Comparative Analysis of Techniques for Forecasting Tourists' Arrival. Journal of Tourism & Hospitality, 6(3), 1000285.