

A Seasonal Autoregressive Integrated Moving Average (SARIMA) Model to Forecasting Tourist Arrival in the Philippines: A Case Study in Moalboal, Cebu (Philippines)

Severina P. Velos, Marivel B. Go, Glynne P. Bate, and Elvira B. Joyohoy
Cebu Technological University - Moalboal Campus
seve767@yahoo.com

Abstract

Forecasting plays a critical part in implementing effective tourism management strategies. However, the role of tourism forecasting is not extensively studied in the Philippines, which is a key tourism destination in Southeast Asia. To address such gap, this paper explores the dynamics of tourist demand in the Philippines through a case study. It illustrates the tourist arrival using a SARIMA model. Results show that the adopted methodology was able to capture the dynamics of the tourist demand in the Philippines. By providing lenses to the Philippine tourism case, this paper would help shed light to the gaps in the literature's current understanding of tourism in the Philippines. Moreover, these findings would be beneficial for stakeholders in shaping policies, strategies, and other tourism initiatives.

Keywords: SARIMA, seasonal time series, stochastic processes, tourism demand forecasting, Philippine tourism

1.0 Introduction

Tourism plays a vital role in many economies across the globe, contributing to about 10% of the global gross domestic product (GDP) (Abellana, Rivero, D. M., Aparente & Rivero A., 2020; Ghalekhondabi, Aedjmand, Young & Weckman, 2019). With this, studies on the different levels of tourism (i.e., global, national, or local) have received extensive focus in the current literature ((Song, Qui & Park, 2019; Song & Li, 2008). While works in literature have grown in numbers, most studies focused on tourism from the perspective of

developed countries. In contrast, few attempts have been made to investigate tourism perspectives in developing countries. Several scholars, including Abellana et al. (2020), Song et al. (2019), and Rufino (2011), emphasize that perspectives in developing countries must be given significant attention to shape a more comprehensive understanding of tourism. One of the developing countries that require such attention is the Philippines, being a productive tourist destination in Southeast Asia (Abellana et al., 2020; Rufino, 2016; Rufino, 2011). The Philippines is an archipelagic country in

Southeast Asia consisting of about 7641 islands with three main geographical divisions from north to south: Luzon, Visayas, and Mindanao. The country has received a significant contribution from the tourism sector as reflected by several economic performance indicators, including a 12.2% share in total GDP since the year 2012 (Philippine Statistics Authority, 2018).

In this connection, Republic Act No. 9593, also known as the Tourism Act of 2009, recognizes tourism as an indispensable element of the Philippine economy that generates investment, employment, growth and national development (Abellana et al., 2020). In the current literature, forecasting tourist arrivals (also called as tourism demand forecasting) support stakeholders in achieving efficient business operations and tourism management, such as in staffing and resource allocation, tourism brochure preparation, and hotel investments, among others (Bi, Liu & Li, 2020; Li, Ge, Liu & Zheng, 2020). Despite the tourism sector's role in the Philippine economy, very little research has been done on developing forecasting models for tourist arrivals in the country. As such, very scant evidence has been provided on the role of tourist arrival forecasting in improving operations of the tourist destinations in the country. To date, only Abellana et al. (2020), Maliberan (2019), Rufino (2016), and Rufino (2011) have developed forecasting models for tourist arrivals in the Philippines. Among such works, only Maliberan (2019) has focused on tourist arrivals at a local level while the rest of the works focused on either a national or international level.

While national and international aggregation levels are crucial for analyzing tourism at a large scale, forecasts at local levels deserve significant attention in the literature as they would be more directly useful to local stakeholders in strategy

formulations. To address such gaps in the literature, this study models the tourist arrivals using a seasonal autoregressive integrated moving average (SARIMA) on historical data of a locality (i.e., municipality) in Cebu (Philippines). The SARIMA model is suitable for tourism data due to the inherent seasonality of tourist arrivals (Abellana et al., 2020; Fernandez, Po III, Montero, & Addawe, 2017). As such, SARIMA is an extension of the autoregressive integrated moving averages (ARIMA) to handle seasonal components in the time series. In the Philippine setting, Abellana et al. (2020) used a combination of the support vector regression (SVR) and SARIMA to model the tourist arrivals in the country from 1988 to 2019. The hybridization was crucial due to the presence of a nonlinear trend in the historical data. If a nonlinear trend is not present in the historical data, SARIMA can be used without combining it with other models (Abellana et al., 2020). Hence, SARIMA is used as a single model in this study due to the absence of a nonlinear trend in the case subject.

This study analyzes the monthly tourist arrivals of Moalboal, Cebu (Philippines) from 2010 to 2018. Tourist arrivals are considered homogenous. In other words, the inbound to the municipality is focused instead of counting arrivals from each destination in the municipality. The choice of study design is motivated by two objectives: (i) to provide lenses on the dynamics of tourist arrivals at a municipal perspective, and (ii) to draw insights that would be directly useful for local governments and stakeholders. Moalboal, a municipality in Cebu (Philippines), is chosen for the case study due to being one of the most visited tourist destinations in the province. In fact, the latest report of the Department of Tourism (Department of Tourism, 2018) ranks the municipality as the 9th most visited tourist destination out of 43 municipalities/

cities in Cebu (Philippines). Despite being highly rated, the municipality has not extensively focused on developing strategies based on its tourist arrival forecasts. As such, the study would be significant to the literature as it is one of the very few to model the tourist arrivals in the Philippines, especially at a local level. The findings in the study can be used to develop a more comprehensive framework for understanding tourism holistically. Finally, the results and findings of the study can be used by stakeholders who are looking into how local tourist arrivals can be modeled using time series approaches.

2.0 Methodology

2.1 The Seasonal Autoregressive Integrated Moving Averages (SARIMA) model

The SARIMA (p, d, q) (P, D, Q, s) model is an extension of the autoregressive integrated moving averages (ARIMA), which can model the seasonality of a time series. Seasonality is the characteristic of a time series that represents the predictable and absolute changes that repeatedly take place in the data (Stewart, 2005). Seasonality occurs in several areas of application. For instance, in fluid mechanics, seasonality has been observed to occur in submesoscale turbulence (Callies, Ferrari, Klymak, & Gula, 2015). Similarly, in the biological and physical sciences, the hourly aggregation of the seasonality of physical and biological controls on the surface ocean was observed (Shadwick, Trull, & Tilbrook, 2015). Likewise, in tourism, the frequency of arrival of tourists has been observed to follow a seasonal pattern (Baldigara & Mamula, 2015). With the tourist demand following a seasonal pattern, the SARIMA model becomes an appropriate model to use. To introduce the SARIMA model, it is mathematically defined below:

If d and D are nonnegative integers, then $\{X_t\}$ is a seasonal ARIMA(p,d,q) (P,D,Q)_s process with period s if the differenced series $Y_t=(1-B)^d(1-B^s)^D$ is a causal ARMA process defined by

$$\Phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t, \{Z_t\} \sim WN(0, \sigma^2)$$

Where:

$$\begin{aligned} \phi(z) &= 1 - \sum_{i=1}^p \phi_i z^i, \\ \Phi(z) &= 1 - \sum_{I=1}^P \Phi_I z^I, \\ \theta(z) &= 1 - \sum_{j=1}^q \theta_j z^j, \\ \Theta(z) &= 1 - \sum_{J=1}^Q \Theta_J z^J, \end{aligned}$$

2.2 Box-Jenkins Methodology

WISCONSIN UNIV MADISON DEPT OF STATISTICS, Box & Jenkins (1970) forged a procedure for discrete systems and time series modeling called the Box-Jenkins Methodology. Several scholars adopt it in literature for developing forecasting models such as AR, MA, ARMA, ARIMA, SARIMA, and SARIMAX (Gairaa, Khellaf, Messlem, & Chellali, 2016; Anvari, Tuna, Canci, & Turkay, 2016; Jackson, Sillah, & Tamuke, 2018). In the tourism forecasting literature, it has also taken a dominant position as a modeling methodology. For instance, Baldigara & Mamula (2015) made use of the Box-Jenkins methodology in building a SARIMA model that describes the behavior of the German tourists' flows to Croatia. Similarly, Liang (2014) used the Box-Jenkins methodology in constructing several forecasting models such as ARIMA and SARIMA, among others.

The Box-Jenkins methodology will be adopted using three iterative steps to construct

the SARIMA model used in this paper (WISCON UNIV MADISON DEPT OF STATISTICS, Box & Jenkins, 1970; Geurts, Buchman, & Ibrahim, 1976; Baldigara & Mamula, 2015):

Step 1: Model Identification. Determining the appropriate model to be used (e.g., AR, MA, ARMA, ARIMA, and SARIMA) that best exemplifies the underlying process is done through a close examination of the sample autocorrelation function (ACF) and partial autocorrelation function (PACF).

Step 2: Parameter Estimation. After step 1, the sample ACF and PACF can be further used to determine the significant parameters of the model. Significant lags of the sample ACF can be used to estimate the MA(q) parameter, while the considerable lags of the sample PACF can be used to determine the AR(p) parameter. Moreover, the estimated parameters will be used to fit the model.

Step 3: Diagnostic Checking. After estimating the suitable parameters, the fitted model is checked to identify the adequacy of the model in representing the observed series, and as such, checking the model requires an examination of the sample ACF and PACF of the differences between the forecasted and actual values. If the sample autocorrelations do not satisfy specific statistical tests, then the pattern of autocorrelations should indicate methods for model improvement. The process can then reiterate back to Step 1 until the most appropriate model is obtained.

Case Study

The Philippines is a Southeast Asian country that is recorded to have one of the highest tourist demands in Asia. Although there is a prominence of typhoons due to its geographical location, it does not lead tourists to postpone their travel schedules to the country (Etulle-Tapanan, 2015). Its prominent tourist destinations are the Boracay Beach, Chocolate Hills, Underground River, Panglao Island, and Vigan City, among others (Department of Tourism, 2018). Tourism has been a significant activity in the Philippines and has made a massive impact on its economy. It has contributed 12.2 percent share to the country's total gross domestic products (GDP) (PSA, 2018). Tourism is also highly active in the southern parts in the Philippines as evident in the prevalence of tourist activities such as diving, heading to the beach, trips on heritage and cultural sites, tours on wildlife sites, caving, mountain biking, sea kayaking, surfing, trekking, white water action, and rock climbing, among others (Department of Tourism, 2018). Hence, a case study is conducted in Moalboal, Cebu, Philippines, one of the highly active tourist destinations in the country, especially in terms of island hopping (Department of Tourism, 2018), to explore the forecasting models' performance in developing countries. Figures on the monthly arrival of tourists for the duration of January 01, 2010 up to December 31, 2018 were gathered from the local government of Moalboal, Cebu, Philippines. As such, a total of 108 observations are obtained from the dataset. With this, the 108 observations are distributed into two set divisions: (i) training set, and (ii) test set. The first 84 observations will be used as a training set to construct the time series-forecasting model. The remaining observations will be used as test data to compare with the out-sample forecast data of the time series forecasting model. Finally, to provide a comparison with the adopted time

series-forecasting model's performance, the model is compared to benchmark forecasting models. This is done to determine if an adopted forecasting model provides significant improvement to "simply using today's value in predicting tomorrow's value" (Hyndman & Koehler, 2006; Ülke, Sahin, & Subasi, 2018).

3.0 Results and Discussion

3.1 Time Series Data from January 2010 to December 2018

Shown below is the time-series data of the tourist arrival in Moalboal, Cebu. The data is plotted using Python's matplotlib library, as presented in Figure 1. The SARIMA analysis is carried out using the time series analysis package offered by Scientific Python (SciPy). Furthermore, the platform used for coding the analysis is Google's Colaboratory, a free-to-use cloud platform that provides a GPU-accelerated virtual machine. By simple inspection of the data in the temporal domain, a seasonal pattern seems to occur in the data. Such a claim will be verified in later procedures. To determine the data distribution without reference to the temporal domain, the data in the frequency domain present a non-normal behavior of the data, as shown in Figure 2.

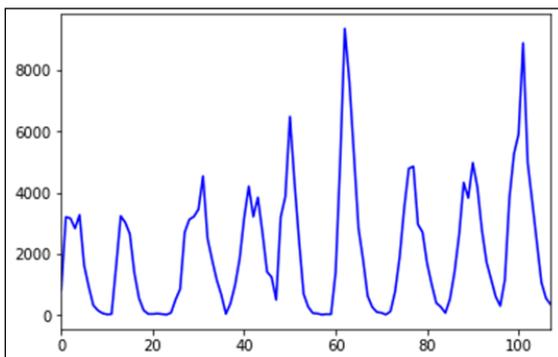


Figure 1. Tourist arrivals in Moalboal, Cebu

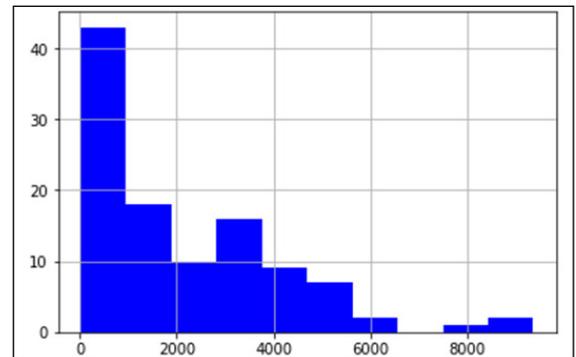


Figure 2. Data distribution in the frequency domain

3.2 Augmented Dickey-Fuller (ADF) Test for Detecting Significant Unit Roots

The ADF test allows the determination of unit root in the time series data. A statistically significant unit root is indicated when the null hypothesis is failed to be rejected, which implies that the series needs to be differenced. The Python code used to perform the ADF test is presented below:

```
from pandas import Series #Importing Series
from statsmodels.tsa.stattools import adfuller #Importing ADF test
result_arrivals = adfuller(arrivals_train) #Assigning the training data to an internal variable
print('arrivals_ADF Statistic: %f' % result_arrivals[0]) #Displays the test statistic
print('p-value: %f' % result_arrivals[1]) #Displays the p-value
print('Critical Values:') #Displays critical values for the hypothesis test
if result_arrivals[1] < 0.05:
    print('The series is already stationary')
else:
    print('The series is still non-stationary. Apply differencing')
for key, value in result_arrivals[4].items():
    print('\t%s %3f' % (key, value))
```

The ADF test arrived at the following results with the use of the code snippet displayed in this section:

arrivals_ADF Statistic: -6.644895
 p-value: 0.000000
 Critical Values:
 The series is already stationary
 1%: -3.494
 5%: -2.889
 10%: -2.582

The successful rejection of the null hypothesis provides sufficient evidence to support the claim that the series is already stationary; hence, no differencing is required.

3.3 Determining the autocorrelation function (ACF) and partial autocorrelation function (PACF)

The ACF and PACF are plotted in Figure 3 and Figure 4, respectively. The shaded region in Figures 3 and 4 represent the confidence interval upon which the autocorrelations and partial autocorrelations are significant. They provide insights into the AR(p) and MA(q) parameters. Moreover, significant lags that are preceded by non-significant lags may give a signal on seasonal parameters.

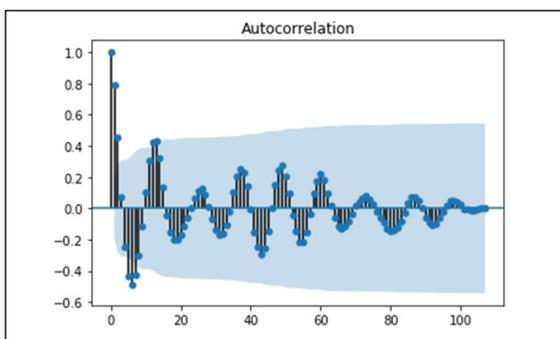


Figure 3. Autocorrelation function (ACF) correlogram of series

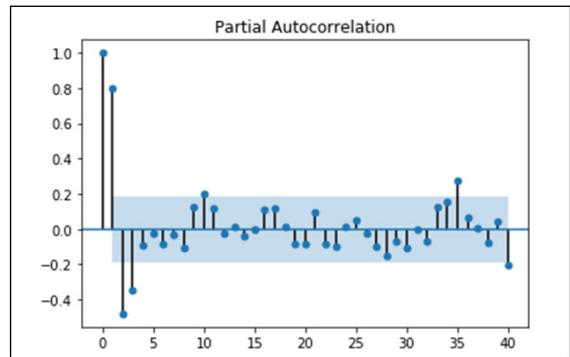


Figure 4. Partial autocorrelation function (PACF) correlogram of series

From the plots, the relevant lags at which the autocorrelation is significant and follow a decaying structure are 1 and 2. This result provides insights on the MA(q) parameter. All other significant lags in the ACF plot may provide insights on the seasonal parameter of the series. In the PACF plot, the significant lags that follow a decaying form are 1, 2, and 3, which provide insights on the AR(p). All other significant lags in the PACF plot provide insights on the seasonal parameter.

3.4 Determining the SARIMA (p,d,q)(P,D,Q,S) parameters

The determination of the model parameters is aided by the rules of thumb provided by the Box-Jenkins methodology. Moreover, the Aikake Information Criterion (AIC) is also used to determine the model fitness. The potential seasonal parameters lie between lags 10 to 12. Due to the behavior of the tourism arrival, it is reasonable to select lag 12 as the seasonal parameter. Moreover, this is the only parameter that does not make the estimation approach invertible, hence, signals model validity. The model to be used, then, is the SARIMA(2,0,2)(0,0,0,12).

3.5 SARIMA forecasting

The SARIMA(2,0,2)(0,0,0,12) is used to generate a two-year (or 24 months) forecast of the tourist arrivals in Moalboal, Cebu. The forecasts are presented in Figure 5. The code snippets below are used to implement the SARIMA model in Python code:

```
from statsmodels.tsa.statespace.sarimax import SARIMAX
arrivals_SARIMAX=SARIMAX(arrivals_train,order=(2,0,2),seasonal_order=(0,0,0,12))
arrivals_SARIMAX=arrivals_SARIMAX.fit()
SARIMAX_forecast=arrivals_SARIMAX.predict(start=84,end=107)
SARIMAX_forecast
```

Using the SARIMAX class of the time series library of SciPy, the forecast values below are generated and are plotted against the original time series using the following Python code:

```
#plt.plot(ARIMA_forecast)
import matplotlib.pyplot as plt
plt.plot(arrivals,color='c')
plt.plot(SARIMAX_forecast,color='r')
plt.xlabel('Period')
plt.ylabel('Number of Tourist Arrivals')
plt.show()
```

The above code snippet uses the matplotlib library of Python and results in the plot below:

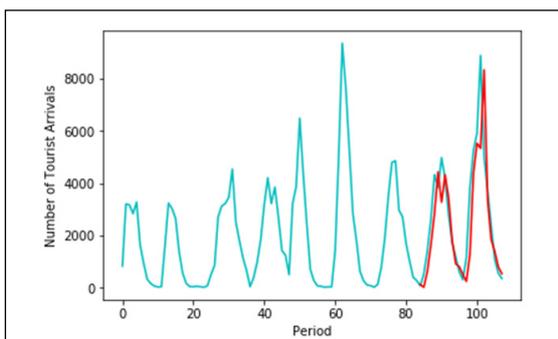


Figure 5. Forecasted Values vs. Original Time Series

3.6 Forecast evaluation

The mean absolute percentage error (MAPE) is the performance criterion used in this study because of its ability to quantify the fitness of the forecasted values to the actual values without dependence on the scale used, hence, making it scale-invariant. As such, the interpretation is more intuitive in contrast to other performance criteria. As such, the SARIMA model utilized in this paper is put to comparison with a NAIVE model, seasonal NAÏVE model, and Holt-Winters method (i.e., considered a benchmark model in literature in testing new models due to its simplicity in which a future value is equal to its directly previous value) following Abellana et al. (2020). The table of comparison is presented in Table 1. As stipulated in Table 1, the SARIMA model has lower forecast error than the NAÏVE model; hence, it has higher forecast accuracy. In order to determine how much improvement the proposed SARIMA model offers, it is helpful to compare the model's percentage improvement relative to the Naïve or Random Walk model (Ashley, 2003). As such, the MAPE and MSE of the SARIMA model are compared to the Naïve or Random Walk model by taking their ratio using Equation 2 and 3:

$$MAPE \text{ ratio} = \frac{SARIMA-NAIVE}{NAIVE} \times 100 \quad (2)$$

$$MSE \text{ ratio} = \frac{SARIMA-NAIVE}{NAIVE} \times 100 \quad (3)$$

With Equation 2 and 3, the MAPE ratio and MSE ratio of the SARIMA model in comparison to the Naïve or Random Walk model are -43.33 percent and -16.44 percent, respectively (i.e., the negative sign implies that a reduction has occurred in the performance measure). Although there is a lack of general agreement in the literature as to how much one can say that improvement is

significant in the context of using MAPE and MSE, among other criteria, such ratios are considered statistically significant. In fact, Ashley (2003) argues that a 16 percent reduction in MSE is statistically significant at a ten percent confidence level, assuming normally distributed residuals. Based on such arguments, it can be inferred that the Naïve or Random Walk model was significantly outperformed by the SARIMA model; hence, it may be a useful representation of the tourist arrivals used in the case study.

Discussions

Tourist demand has long been observed in the literature to follow a seasonal pattern, as found by several scholars such as Baldigara & Mamula (2015), Song et al. (2019), and Abellana et al. (2020), among others. Hence, these patterns exhibited by tourist demand conform to that of a stochastic process and can be mathematically represented with a time series model. A time series can be broken down into four components: (1) cycle, (2) seasonal patterns, (3) trend, and (4) noise. With cyclical components reflecting the repeated but non-periodic fluctuations and the trend reflecting the series progression in the long-run, the additional information brought about by the seasonal component allows the development of a more accurate forecasting model that captures essential statistical information about the stochastic process. The most undesirable case in developing a time series model with the available data is when only noise or random fluctuations can be captured from such a dataset. This is the case when a stochastic process behaves only like white noise (Stewart, 2005; Brockwell & Davis, 2016). Subsequently, white noise is a random

fluctuation having same-level intensities at varying frequencies. It is composed of serially unrelated random variables with zero mean and finite variance put into a sequence, and depending on the context, may need an independent and identical distribution (or iid). In other words, it would be impossible to trace patterns in the sequence of a white noise process. On the contrary, the best case is when all the components can be effectively taken into account by the model (Stewart, 2005; Brockwell & Davis, 2016). Remarkably, most of the valuable information in the dataset were effectively captured by the SARIMA(2,0,2)(0,0,0,12) model adopted in this paper. The MAPE obtained by the SARIMA model is only 36%, while the considered benchmark models obtained significantly higher MAPE and MSE, as shown in Table 1.

These results conform to the general observation in the literature that in the presence of seasonal patterns, the SARIMA model outperforms the Random Walk or Naïve model as well as other forecasting models as well. Since Moalboal ranked as 9th most visited municipality by tourists (both domestic and foreign), it is crucial to adopt an efficient allocation of its resources. However, with the municipality's lack of strategies and techniques to facilitate such allocation scheme, the city may not be able to maximize its benefit from tourism. In this study, SARIMA was used to model tourist flow in the municipality. Modeling such data enables the elicitation of strategies that would be directly useful for the LGU and business stakeholders. The findings of this study are beneficial to the local government unit (LGU) of Moalboal, Cebu. For one, it would be possible to study the effect of new activity in inducing tourist arrivals by being able to analyze the behavior of the time series.

Table 1. Performance Evaluation of the Forecast Models using MAPE

No.	Actual Values	SARIMA	Absolute Percentage Error	Squared Error	NAïVE	Absolute Percentage Error	Squared Error	Seasonal NAïVE	Absolute Percentage Error	Squared Error	Holt-Winters Method	% Absolute Error	Squared Error
84	80	134	0.68	2916	277	2.46	38809	448	4.60	135652	317	2.96	56169
85	530	44	0.92	236196	80	0.85	202500	209	0.61	103071	303	0.43	51529
86	1413	592	0.58	674041	530	0.62	779689	650	0.54	582921	112	0.92	1692601
87	2630	1614	0.39	1032256	1413	0.46	1481089	823	0.69	3264185	336	0.87	5262436
88	4336	2822	0.35	2292196	2630	0.39	2910436	1742	0.60	6728971	2211	0.49	4515625
89	3836	4451	0.16	378225	4336	0.13	250000	3633	0.05	41177	2396	0.38	2073600
90	4983	3309	0.34	2802276	3836	0.23	1315609	3999	0.20	968126	4887	0.02	9216
91	4184	4345	0.04	25921	4983	0.19	638401	6893	0.65	7338804	4173	0.00	121
92	2769	3493	0.26	524176	4184	0.51	2002225	6338	1.29	12734589	2688	0.03	6561
93	1736	1654	0.05	6724	2769	0.60	1067089	6447	2.71	22198159	1754	0.01	324
94	1179	808	0.31	137641	1736	0.47	310249	5168	3.38	15912121	1166	0.01	169
95	613	738	0.20	15625	1179	0.92	320356	4654	6.59	16326657	639	0.04	676
96	310	358	0.15	2304	613	0.98	91809	3418	10.03	9658172	80	0.74	52900
97	1154	98	0.92	1115136	310	0.73	712336	2699	1.34	2386277	530	0.54	389376
98	3837	1249	0.67	6697744	1154	0.70	7198489	2663	0.31	1379144	1413	0.63	5875776
99	5276	4427	0.16	720801	3837	0.27	2070721	3985	0.24	1665728	2630	0.50	7001316
100	5914	5509	0.07	164025	5276	0.11	407044	5285	0.11	395666	4336	0.27	2490084
101	8892	5287	0.41	12996025	5914	0.33	8868484	7214	0.19	2815753	3836	0.57	25563136
102	4989	8487	0.70	12236004	8892	0.78	15233409	9016	0.81	16213906	4983	0.00	36
103	3700	3428	0.07	73984	4989	0.35	1661521	8881	1.40	26845962	4184	0.13	234256
104	2384	1625	0.32	576081	3700	0.55	1731856	8830	2.70	41548475	2769	0.16	148225
105	1076	1439	0.34	131769	2384	1.22	1710864	7388	5.87	39838905	1736	0.61	435600
106	562	528	0.06	1156	1076	0.91	264196	6200	10.03	31792657	1179	0.10	380689
107	357	155	0.57	40804	562	0.57	42025	4885	12.68	20499123	613	0.72	65536
Mean			36%	1786834.42		64%	2137883.58		282%	11723925.04		51%	2346081.54

Moreover, having knowledge of the tourist flows in the municipality would enable LGUs to plan out their budget and schedule for tourism in proportion with the forecasted volume of tourists. The LGU can also extend the forecasts to determine the proportion of the arrivals that visit the different tourist destinations in the locality. Aside from LGUs, the individual tourist destinations in Moalboal such as beach resorts can also benefit from the model. Since the arrival of customers in each destination is stochastic, business operations can be made more efficient by appropriately allocating resources based on the forecasts. With this, more efficient allocation algorithms can be adopted by firms.

4.0 Conclusions and Future Works

With the crucial role of forecasting in the management of tourism activities, interests in both academic and practical contexts on forecasting spurred across the globe. However, the case of tourism in developing countries is relatively less studied in contrast to developed countries. Such a gap in the literature presents several concerns. For one, it may potentially induce a lack of understanding of the tourism characteristics in developing countries. It is a concern that needs to be addressed since developing countries have become growing stakeholders around the globe. To address the gap in the literature as well as the resulting concerns, this study aims to model the tourist arrivals in the Philippines with monthly tourism data from January 2010 to December 2018 using SARIMA. Results show that the MAPE and MSE of the SARIMA model outperformed the benchmark models. As such, the SARIMA model can be a useful representation of the tourist arrivals in the case study and would provide valuable insights

that help in understanding the tourist demand characteristics of developing countries. The results of this paper contribute significantly to the literature. First, it models the tourist demand in a developing country using a case study in a province located in the Philippines. Second, it showed how a SARIMA model could use the seasonal patterns inherent in the time series data of tourist arrival to improve the forecasting accuracy.

Collectively, these contributions provide scholars insights that shed light on the tourism dynamics in a developing country, which is relatively less studied in the literature. Moreover, the results of this paper are beneficial to tourism planners, government agencies, private organizations, and other stakeholders because the study may enable them to design better tourism strategies such as planning tourist activities ahead of time, improving existing tourist destinations, or launching new tourist destinations. Furthermore, it allows them to answer managerial questions relevant to capacity, location, and marketing decisions, among others. This paper, however, is limited in its scope. Despite being able to show the improvement in the forecast accuracy by adopting a SARIMA model, this paper was not able to directly test its statistical significance; however, it relied only on relevant results from the literature to decide if such results have significance. Moreover, sensitivity analyses regarding the number of lags at which the model remains accurate have not been explored due to the scope of the work. Such undertakings are left in this paper for future works. Likewise, hybrid models such as artificial neural networks, deep neural networks, classical neural networks, and causal models, among others, may be adopted to increase the forecasting accuracy of the model.

Reference

- Abellana, D. P. M., Rivero, D. M. C., Aparente, M. E., & Rivero, A. (2020). Hybrid SVR-SARIMA model for tourism forecasting using PROMETHEE II as a selection methodology: a Philippine scenario. *Journal of Tourism Futures*.
- Anvari, S., Tuna, S., Canci, M., & Turkay, M. (2016). Automated Box–Jenkins forecasting tool with an application for passenger demand in urban rail systems. *Journal of Advanced Transportation*, 50(1), 25-49
- Ashley, R. (2003). Statistically significant forecasting improvements: how much out-of-sample data is likely necessary? *International Journal of Forecasting*, 19, 229-239.
- Baldigara, T., & Mamula, M. (2015). Modeling international tourism demand using seasonal ARIMA models. *Tourism and hospitality management*, 21(1), 19-31
- Bi, J. W., Liu, Y., & Li, H. (2020). Daily tourism volume forecasting for tourist attractions. *Annals of Tourism Research*, 83, 102923.
- Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting*. Springer.
- Callies, J., Ferrari, R., Klymak, J., & Gula, J. (2015). Seasonality in submesoscale turbulence. *Nature communications*, 6, 6862. <https://doi.org/10.1038/ncomms7862>
- Department of Tourism (2018). *Visitor arrivals to the Philippines by country of residence: January – December 2017*. Retrieved from http://www.tourism.gov.ph/industry_performance/ December2018/January%20December%202018%20Table%202.pdf
- Etulle-Tapanan, H. (2015). Path Analysis of Climate and Tourism to the Economic Growth in the Philippines. *Recoletos Multidisciplinary Research Journal*, 3(1). <https://doi.org/10.32871/rmrj1503.01.03>
- Fernandez, F. R., Po III, R., Montero, N., & Addawe, R. (2017, November). Prediction of South China sea level using seasonal ARIMA models. In *AIP Conference Proceedings 1905*(1), 050018. AIP Publishing LLC.
- Gairaa, K., Khellaf, A., Messlem, Y., & Chellali, F. (2016). Estimation of the daily global solar radiation based on Box–Jenkins and ANN models: A combined approach. *Renewable and Sustainable Energy Reviews*, 57, 238-249
- Geurts, M., Buchman, T., & Ibrahim, I. (1976). Use of the Box-Jenkins approach to forecast tourist arrivals. *Journal of Travel Research*, 14(4), 5-8.
- Ghalehkhondabi, I., Ardjmand, E., Young, W.A. & Weckman, G.R., (2019). A review of demand forecasting models and methodological developments within tourism and passenger transportation industry. *Journal of Tourism Futures*, 5(1), 75-93. <https://doi.org/10.1108/JTF-10-2018-0061>
- Hyndman, R., & Koehler, A. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4), 679-688.
- Jackson, E., Sillah, A., & Tamuke, E. (2018). Modeling monthly headline consumer price index

- (HCPI) through seasonal Box-Jenkins methodology. *International Journal of Sciences*, 7(1), 51-56.
- Li, C., Ge, P., Liu, Z., & Zheng, W. (2020). Forecasting tourist arrivals using denoising and potential factors. *Annals of Tourism Research*, 83, 102943.
- Liang, Y. (2014). Forecasting models for Taiwanese tourism demand after allowance for Mainland China tourists visiting Taiwan. *Computers & Industrial Engineering*, 74, 111-119.
- Maliberan, R. M. E. (2019). Forecasting tourist arrival in the Province of Surigao del Sur, Philippines using time series analysis. *JOIV: International Journal on Informatics Visualization*, 3(3), 255-261.
- Philippine Statistics Authority (2018). Contribution of Tourism to the Economy is 12.2 Percent in 2017. Republic of the Philippines: Philippine Statistics Authority. Retrieved from <https://psa.gov.ph/content/contribution-tourism-economy-122-percent-2017>
- Rufino, C. C. (2011). Forecasting international demand for Philippine Tourism. *DLSU Business & Economics Review*, 21(1), 61-76.
- Rufino, C. C. (2016, March). Forecasting monthly tourist arrivals from ASEAN+ 3 countries to the Philippines for 2015-2016 using SARIMA Noise Modeling. In *Presented at 2016 DLSU Research Congress*.
- Shadwick, E., Trull, T., & Tilbrook, B. (2015). Seasonality of biological and physical controls on surface ocean CO₂ from hourly observations at the Southern Ocean Time Series site south of Australia. *Global Biogeochemical Cycles*, 29(2), 223-238.
- Shumway, R. H., & Stoffer, D. S. (2017). ARIMA models. In *Time series analysis and its applications* (pp. 75-163). Springer, Cham.
- Song, H., & Li, G. (2008). Tourism demand modelling and forecasting—A review of recent research. *Tourism management*, 29(2), 203-220.
- Song, H., Qiu, R. T., & Park, J. (2019). A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting. *Annals of Tourism Research*, 75, 338-362.
- Stewart, K. G. (2005). *Introduction to applied econometrics*. Belmont, CA: Thomson Brooks/Cole.
- Ülke, V., Sahin, A., & Subasi, A. (2018). A comparison of time series and machine learning models for inflation forecasting: empirical evidence from the USA. *Neural Computing and Applications*, 30(5), 1519-1527.
- WISCONSIN UNIV MADISON DEPT OF STATISTICS, Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis forecasting and control*.