International Conference on Business Research
University of Moratuwa, Sri Lanka
December 8, 2025

https://doi.org/10.31705/ICBR.2025.22

A COMPARATIVE STUDY OF SARIMA AND PROPHET MODELS FOR FORECASTING TOURIST ARRIVALS TO SRI LANKA

P.G.S. Upeksha* and H.G.G. Kaushalya

Department of Interdisciplinary Studies, Faculty of Information Technology, University of Moratuwa, Sri Lanka
*shaliniu@uom.lk**

ABSTRACT

The Tourism sector in Sri Lanka predominantly contributes to the nation's economy, making it vital for numerous stakeholders to understand the tourist arrival patterns to make informed decisions. Thus, primary objective of this research is to model tourist arrivals to Sri Lanka from January 2009 to June 2025 and provide accurate forecasts for the upcoming year, i.e., from July 2025 to June 2026. This univariate time series modelling was done using two approaches: SARIMA and FB Prophet modelling. Forecasts were generated using the SARIMA model and Prophet model, whereas performance metrics indicated that ARIMA (1,1,0) $(1,0,1)_{12}$ model predicts with better accuracy. Moreover, the visual inspections of the time series identified major disruptions beyond the seasonal fluctuations which had a long-term impact on the tourism industry.

Keywords: Tourism, Forecast, SARIMA, FB Prophet

1. Introduction

Tourism sector has seen swift expansion and has become crucial for the economies of both developed and developing nations. In Sri Lanka, the tourist arrivals significantly influence the economy. Though the industry was confronted with challenges posed by the Easter Attack in 2019 and COVID-19 Pandemic, tourist arrivals started to increase over the recent past. As per the Sri Lankan Tourism Authority, Sri Lanka recorded tourist arrivals of 1.5 million which is the highest since the year 2019. Touristdemand not only boosts Gross Domestic Product but also enhances employment opportunities and generates new sources of income for local communities, as well as public and private sectors (SLTDA, 2024).

Given the significant impact of Sri Lankan tourism sector on country's economy, there is a growing interest in researching and accurately predicting the international tourist arrivals. However, comprehensive studies focusing on developing models and projections of tourism inflows to Sri Lanka remain relatively scarce in existing literature. Furthermore, Sri Lanka currently lacks a precise forecasting methodology, leading to significant discrepancies between projected figures and actual outcomes (Fernando, 2016). According to Silva et al. (2019), in planning and implementing the strategies and policies in the tourism sector, it is vital to conduct accurate forecasting of international tourist arrivals. Furthermore, it provides important insights for infrastructure development, destination management, and effective decision making on tourism investments (Yang & Zhang, 2019). Thus, countries try to develop different strategies and set targets in order to improve their tourism sectors, especially by increasing international tourist arrivals. However, one of the major issues is that they are often overly ambitious due to erroneous forecasting (Fernando, 2016).

Based on the available body of literature, there are a few studies done to forecast tourist arrivals in Sri Lanka; however, most were based on data prior to COVID-19. Therefore, this research is aimed at developing appropriate models to forecast tourist arrivals to Sri Lanka based on data up to the most recent period, employing robust statistical models and accurately forecast future arrivals which would be able to provide valuable up-to-date insights to policymakers and other stakeholders.

2. Literature Review

A substantial amount of literature is available on time series modelling and forecasting pertaining to the tourism industry and some of them can be listed as follows.

ARIMA modelling and exponential smoothing are two primary techniques used in time series forecasting. The exponential smoothing method focuses on identifying the trends and seasonal patterns in data. ARIMA modelling focuses on identifying the autocorrelations between current and past observations in data. The ARIMA approach contains seasonal components in its model, therefore, it can handle the seasonal patterns in data (Hyndman & Athanasopoulos, 2021).

Kodippili and Senaratne (2017) conducted a study applying SARIMA modelling in forecasting monthly international tourist arrivals to Sri Lanka. Data from June 2009 to March 2017 was used in this study. ARIMA (4,0,0) $(0,1,0)_{12}$ was found to be the best model to produce forecasts of monthly international tourist arrivals to Sri Lanka for the upcoming year.

Peiris (2016) forecasted international tourist arrivals to Sri Lanka with a SARIMA model using data spanning from January 1995 till July 2016. The study found ARIMA (1,0,16) $(36,0,24)_{12}$ as the optimal model. The HEGY method was used to test seasonality and metrics such as Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) were utilized to examine the performance of the models.

Nagendrakumar et al. (2021) employed ARIMA modelling to model monthly tourist arrivals to Sri Lanka with monthly observation between January 2000 to July 2021. It generated forecasts for five years beginning in August 2021using ARIMA (12,1,3) which was the optimal model.

FB Prophet model is an open-source tool which was designed to overcome two main challenges: the limitations of automated forecasting methods and the shortage of skilled analysts. (Taylor & Letham, 2018). Patandung and Jatnika (2021) conducted a study to forecast the international tourist arrival growth in Indonesia amid the COVID-19 pandemic. FB Prophet model was employed, and it was found that there would be a decline in the international tourist arrival growth until February 2022 alongside a downward trend in the number of daily COVID-19 cases in Indonesia until December 2021.

Anggreini & Karmilasari (2021) also employed the Prophet model to forecast tourist arrivals to Bali. For the study data from June 2019 to May 2021 were considered and generated forecasts for the six-month period from June 2021. The accuracy was assessed using MAPE and the low value suggested that Prophet model was a reliable tool to generate forecasts of international tourist arrivals.

Based on the literature survey done, it was found that the ARIMA model and Prophet Model are among the most effective approaches for time series forecasting. However, due to the post-COVID 19 data abnormalities related to Sri Lanka, it was decided to apply the above models to the most recently available data and evaluate their respective performance to derive insights, thereby addressing the current gap in empirical forecasting related studies for tourist arrivals to Sri Lanka.

3. Methodology

The available literature indicates that for short span of time, ARIMA modelling is performing well to generate forecasts whereas FB Prophet model is particularly suited for data having seasonal effects (Hyndman

& Athanasopoulos, 2021). Therefore, both the models are employed to obtain projections of tourist arrivals to Sri Lanka, and their model performance is compared.

3.1. Data Sample

This research is entirely done utilizing secondary data published by Sri Lanka Tourism Development Authority. Monthly total tourist arrivals to Sri Lanka extending from January 2009 to June 2025 are considered for this study.

3.2. Data Preprocessing

Since both SARIMA and FB Prophet models are sensitive to variance, prior to model fitting, the time series was first scrutinized for variance. With the visual inspection of the time series, stabilizing the variance was done and for this data, logarithmic transformation was utilized. When modeling with SARIMA, the data series was further examined for the presence of missing values and outliers as SARIMA model does not inherently handle them. Given that the disruptions such as Easter attack and COVID-19 outbreak are not expected to repeat in future, such anomalies were detected and replaced with linear interpolation, done using 'tsoutliers' and 'tsclean' functions in the 'forecast' package in R. The alternative approaches of handling outliers such as incorporating external regressors and intervention variables were not adopted for this study as these shocks were one-off and not expected to recur. Including them as regressors could cause the model to learn the short-term crisis effects as part of the long-term pattern, leading them to project similar declines in future. Instead, using interpolation approach allows estimate a baseline view of how tourist arrivals to the country would have trended without those extreme events. When modelling data with FB Prophet, the model itself would be more robust in handling missing values, trend shifts, and outliers.

3.3. SARIMA Model

General representation of SARIMA process can be expressed as equation 1.

$$\phi_p(B)\Phi_p(B^s)W_t = \theta_q(B)\Theta_q(B^s)\varepsilon_t \tag{1}$$

where B is the backward shift operator while ϕ_p, Φ_p, θ_q, $\Theta_q$represent polynomials of order p, P, q, Q respectively. ε_t is the white noise at time t. $W_t = \nabla_d\ \nabla_s^d Y_t$ is the differenced series (Priyangika et al., 2016). ∇_d and ∇_s are ordinary and seasonal difference components (Chang et al., 2012).

3.4. FB Prophet Model

Facebook developed this model as an additive regression model that can be written as follows (Equation 2).

$$y(t) = \alpha(t) + \beta(t) + \eta(t) + \varepsilon(t) \qquad (2)$$

where $y(t)$ is the forecast value. $\alpha(t), \beta(t)$, and $\eta(t)$ are trend, seasonal and holiday effects respectively. $\varepsilon(t)$ denotes the forecast error term. This model utilizes a Fourier series to capture and fit seasonal effects (Agyemang et al., 2023).

3.5. Statistical Techniques Used

For this study, R was employed as the programming language and several packages such as readxl, fpp3, tseries, forecast, sarima, and prophet are used to run required statistical tests and data visualizations. When employing SARIMA modelling, model adequacy is checked with employing residual diagnostic plots, characteristic roots of the model, Ljung-Box test and Augmented Dickey-Fuller (ADF) test. Further, once the model fitting is done with SARIMA and FB Prophet models, the respective accuracies will be evaluated with performance metrics namely MAPE, RMSE and MAE. Lewis (1982) has proposed a classification of forecasting accuracy based on MAPE: less than 50% as reasonable, less than 20% as good and less than 10% as very good forecasts.

4. Results and Discussion

4.1. Visualizing the Time Series

Initial visual inspections of the time series were done using different plots (Figures 1, 2 and 3). Visual inspection of the tourist arrivals from January 2009 to June 2025 revealed some seasonal fluctuations alongwith several irregular shocks.

The seasonal plot (Figure 2) displays the monthly tourist arrival patterns across

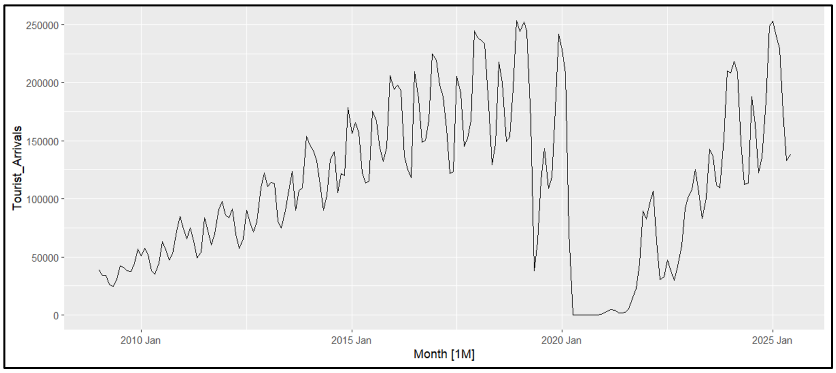

Figure 1: Time series plot: January 2009 – June 2025

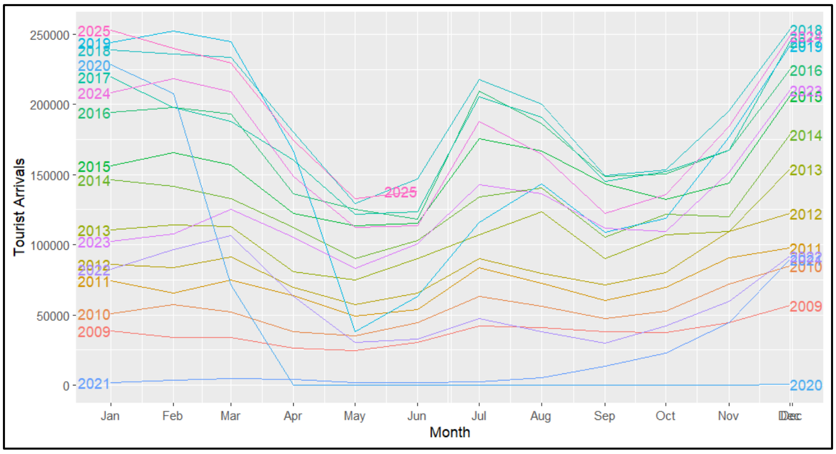

Figure 2: Seasonal plot: January 2009 – June 2025

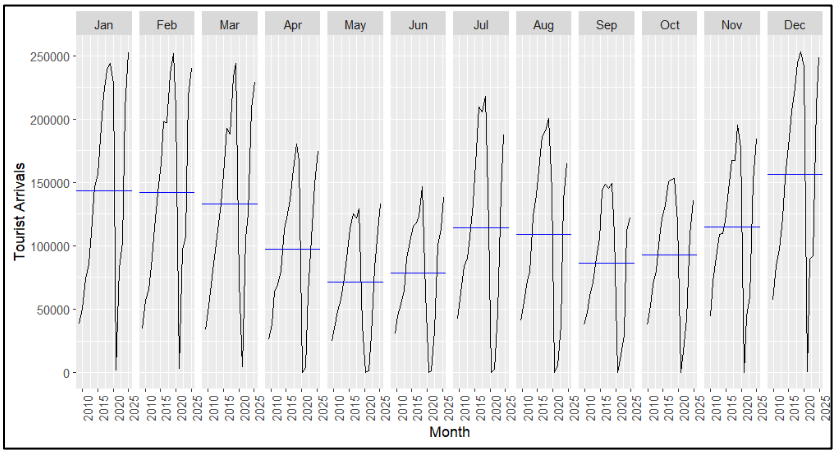

Figure 3: Sub-series plot: January 2009 – June 2025

the years, whereas subseries plot (Figure 3) illustrates the patterns within each month over the period of study. They both show that tourist arrivals are typically high in the months December to March, and this can be attributed to the winter holidays in Western countries and favorable weather in Sri Lanka. In addition to that, July to August also attract more tourists to Sri Lanka due to European summer holidays and cultural events like Kandy Esala perahera.

The tourism industry to Sri Lanka is highly vulnerable to external events as demonstrated by the sudden fluctuations in arrivals during certain periods. When scrutinizing such fluctuations from the whole series plot, one significant drop can be observed in May 2019, following the Easter attack in April, reflecting the sector's sensitivity to country's security situation. In addition to that, in 2020, there is a noteworthy decline in tourist arrivals recording zero tourist arrivals from the month April to November which is due to the COVID-19 pandemic. The disruptions occurred due to lockdowns, border closures, and stringent health protocols affected global mobility, causing major adverse impact on the tourism sector in Sri Lanka. It began to recover in 2021, but the recovery was slowed by the peak of the economic crisis in Sri Lanka in 2022. However, in December 2023, there has been a notable increase in the arrival of international tourists, reaching the year's peak, which may be linked to the recognition of Sri Lanka as one of the top destinations by travel-related entities in addition to the factors like peak tourist season and favorable conditions in Sri Lanka.

4.2. Data Transformation

As per Figure 1, the data demonstrates an increasing variance over time, thus, the stabilization of variance was done through logarithmic transformation (Figure 4). In doing so, as the dataset had zero tourist arrivals for May 2020 to November 2020 due to COVID-19 pandemic, those records were first imputed by a small value to accommodate logarithmic transformation.

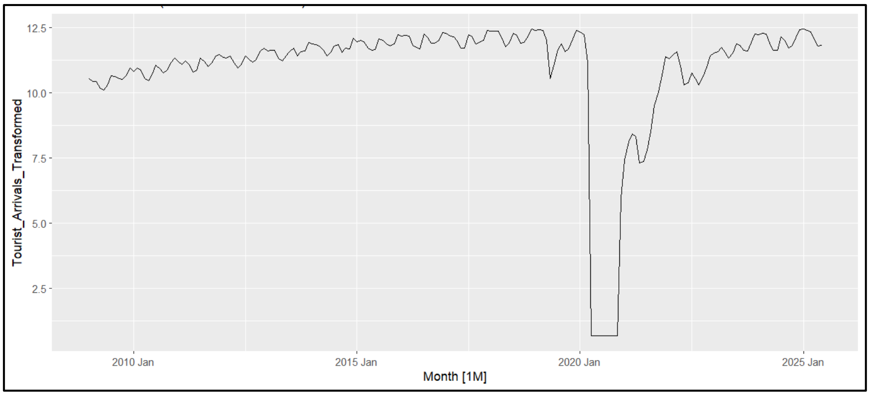

Figure 4: Time Series after Logarithmic Transformation

4.3. SARIMA Model

For modelling the series with SARIMA, the data was split into a training set (January 2009–June 2024) and a test set (July 2024–June 2025).
As the SARIMA model does not inherently deal with outliers, they were closely examined. As the log transformed series had significant outliers, they were detected and replaced with 'tsoutliers' and 'tsclean' functions in the 'forecast' package in R.

Fitting SARIMA Model

After transforming the data and replacing the outliers, the data was suitable for modelling with SARIMA as it can handle the non-stationarity. As the initial estimates of SARIMA model parameters, ACF/PACF plots were inspected (Figure 5). Then, a manual grid was done for different orders and the respective model with the lowest AICc value out of the candidate models was determined as the best fit. Accordingly, the selected model was ARIMA (1,1,0) $(1,0,1)_{12}$.

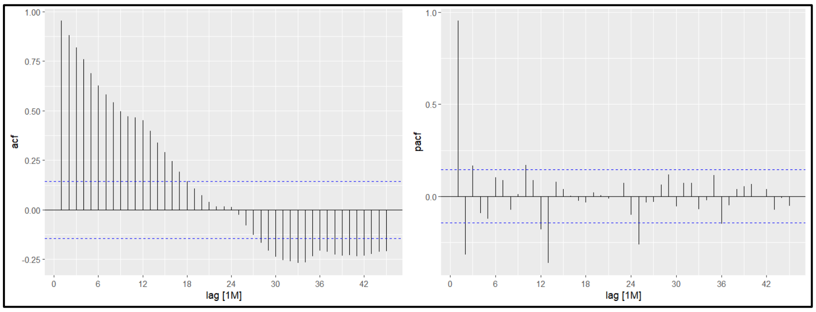

Figure 5: ACF and PACF Plots

Table 1 below provides a summary of the generated output of ARIMA (1,1,0) (1,0,1) model.

Table 1: Model Summary

Model	Coefficients			σ^2	LogLik	AIC	AICc	BIC
	ar1	sar1	sar2					
ARIMA (1,1,0) (1,0,1)$_{12}$	0.304	0.986	-0.854	0.045	18.68	-29.36	-29.14	-16.48

The fitted model can be represented by the following equations 3 and 4.

$$(1-\phi_1 B)\,(1-\Phi_2 B^{12})\,(1-B)Y_t = (1+\theta_1 B^{12})\varepsilon_t \qquad (3)$$

$$(1-0.304B)(1-0.986B^{12})(1-B)Y_t = (1-0.854B^{12})\varepsilon_t \qquad (4)$$

where, Y_t is the total tourist arrivals to Sri Lanka at period t while ε_t denotes the residual at period t.

Evaluation of Model Adequacy

Model adequacy was evaluated with the residual diagnostics graphs, Ljung-Box test, ADF test and analysis of characteristic roots.

- **Residual Diagnostic Graphs**

To assess the adequacy of the best-fitted model, i.e., ARIMA (1,1,0) (1,0,1)$_{12}$, its residual plots (Figure 6) were first inspected.

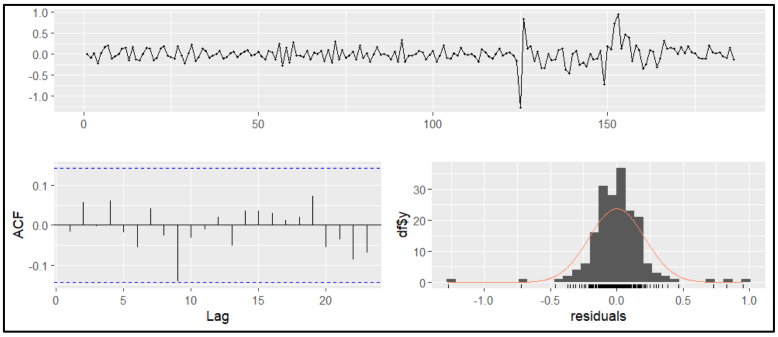

Figure 6: Residual Diagnostic Graphs

The residuals of the fitted model are close to zero, with variance as a whole stays within a constant band except for a few exceptions. Further, the histogram demonstrates a normal distribution. Moreover, in the ACF plot, all spikes lie within the threshold limits, and no significant spikes can be observed, thus, it is consistent with white noise. This can be further confirmed by employing the Ljung-Box test.

- **Ljung-Box Test**

Lag value for the test was set to 24 (twice the period of seasonality) to test residual autocorrelation (Hyndman and Athanasopoulos, 2021). The output is shown in Table 2.

Table 2: Ljung-Box Test Output

Test Statistic	p value
13.093	0.9053

As per the output, Ljung Box statistic (X-squared) is 13.093, whereas p-value is 0.9053 (> 0.05). Therefore, there is not enough evidence to reject the null hypothesis, implying that the residuals of the best-fitted model are not autocorrelated.

- **ADF Test**

In order to find the stationarity of the residual series, ADF test was performed and the output is shown in Table 3.

Table 3: ADF Output

Test Statistic	p value
-5.467	0.01

As per the output, ADF statistic is -5.467, whereas p-value is 0.01 (< 0.05). Therefore, there is enough evidence to reject the null hypothesis. This implies that the residual series of the best-fitted model is stationary.

- **Characteristic Roots Analysis**

ARIMA (1,1,0) $(1,0,1)_{12}$ had all its roots lying within the unit circle having moduli less than one; therefore, it is adequate for forecasting (Figure 7).

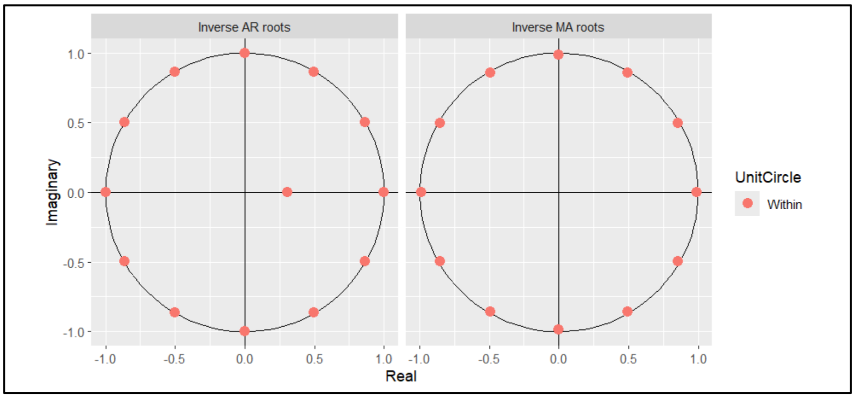

Figure 7: Plot of Inverse Characteristic Roots

Forecasting Tourist Arrivals using SARIMA Model

First, the forecasts were generated for the test period (July 2024 to June 2025). Then, the future values were predicted for 12 months beyond the test period which is from July 2025 to June 2026.

The forecasts generated beyond test period which were converted back to the original scale using back transformation with bias correction are shown below (Table 4).

Table 4: Forecasts from July 2025 to June 2026 - SARIMA Model

Year/ Month	Point Forecast (Log Transformed)	Point Forecast (Original Scale)
2025 Jul	11.92347	154,115
2025 Aug	11.93266	155,537
2025 Sep	11.85139	143,397
2025 Oct	11.95130	158,464
2025 Nov	12.13016	189,499
2025 Dec	12.37125	241,162
2026 Jan	12.31746	228,533
2026 Feb	12.31132	227,136
2026 Mar	12.29715	223,940
2026 Apr	12.04503	174,034
2026 May	11.65610	117,957
2026 Jun	11.74221	128,565

The following plot of future forecasts shows that they have well-captured the trend as well as the seasonal patterns of tourist arrivals (Figure 8).

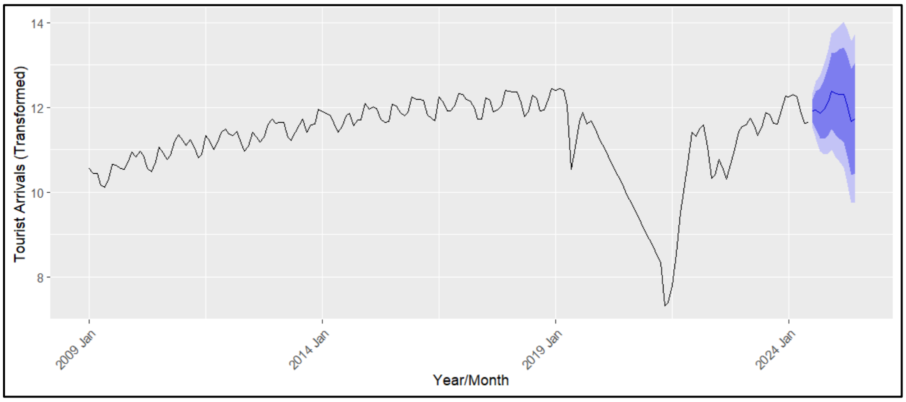

Figure 8: Forecasts of Tourist Arrivals to Sri Lanka - SARIMA Model

4.4. FB Prophet Model

Though the Prophet model can handle the outliers by fitting them with trend changes, some extreme outliers which were identified using tsoutliers function were removed in order to avoid their lasting impact.

Forecasting Tourist Arrivals using Prophet Model

Prophet model was fitted to generate forecasts, and plot below illustrates actual vs forecasted tourist arrivals for the period of study 9).

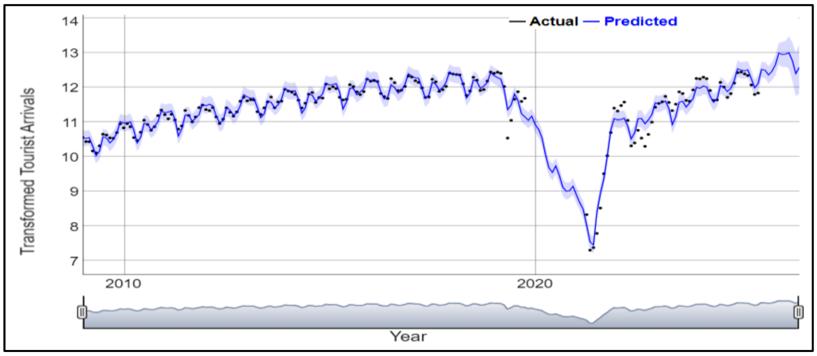

Figure 9: Forecasts of Tourist Arrivals to Sri Lanka - Prophet Model

4.5. Checking Accuracy of Forecasts of Candidate Models

Forecasts were generated from both SARIMA Model and Prophet model as discussed above. The forecast accuracies were compared to determine better performing models using performance metrics such as MAPE, RMSE and MAE.

When checking accuracy related to ARIMA (1,1,0) $(1,0,1)_{12}$, forecasts for the test period, i.e., from July 2024 to June 2025 and actual values during that period were used to derive the performance metrics. Forecast accuracies for the fitted Prophet model was derived performing cross validation which was done with initial training set from January 2009 to December 2011 with a rolling window of 180 days. For the time series, this made 23 total forecasts. Then, overall performance metrics were generated for the fitted Prophet model. Based on the performance metrics, the SARIMA model has generated forecasts with better accuracy. These values are on log scale. Therefore, to derive values in the original scale, back transformation was done for the data with bias correction. SARIMA Model generated the following values for the performance metrics (Table 5).

Table 5: Comparison of Performance Metrics of SARIMA and Prophet Models

Model	Performance Metric	For Log-Transformed Dataset	For Data in Original Scale and Back-Transformed (Bias-Corrected)
ARIMA (1,1,0) $(1,0,1)_{12}$	MAPE	0.79%	8.27%
	RMSE	0.11	16,749
	MAE	0.09	13,935

Prophet Model	MAPE	7.01%
	RMSE	1.20
	MAE	0.74

The MAPE value indicates that tourist arrival forecasts can differ from the actual values by approximately 8.27%. As for Lewis (1982), this falls into the category of very good forecasts.

5. Conclusion

The visual inspections of the tourist data series for the study period identified that the fluctuations are primarily due to holiday seasons in tourist source markets and favorable weather conditions. Further, major shocks were identified in 2019 due to the Easter attack and April 2020 to November 2020 due to COVID-19. Further, in 2022, due to the peak of the then-prevailing economic crisis in the country, the recovery lost momentum, however, in 2023 onwards, it showed a noteworthy improvement.

The modelling of the tourist arrival series was done using SARIMA model and Prophet model and forecasts were generated for the period: July 2025 to June 2026. Considering the performance metrics, the fitted ARIMA $(1,1,0)$ $(1,0,1)_{12}$ model outperformed Prophet model. This is likely to be due to the existence of short-term temporal dependencies in the tourist arrival data rather than complex seasonal patterns which can be better captured by Prophet models. ARIMA models are well-suited for capturing short-term dependencies in time series data (Elseidi, 2024). When back transformed the values with bias correction for the best-fitted SARIMA model, MAPE value was 8.27% indicating high accuracy. Therefore, it can be utilized to generate further forecasts under necessary modifications. Further, the forecasts for the period of July 2025 to June 2026 utilizing ARIMA $(1,1,0)$ $(1,0,1)_{12}$ model demonstrate a growth in tourist arrivals as compared to the previous year. Therefore, these findings will offer valuable insights to relevant stakeholders, including the government, hotels and airlines, by supporting policy formulation, tourism capacity planning and strategic decision making in order to better manage resources and improve service delivery in tourism industry. Further, they provide a basis for effective infrastructure development, destination management, and informed decision-making regarding tourism investments (Yang & Zhang, 2019).
However, the main limitation of this study is that it conducted a univariate analysis. Future studies can consider and incorporate the factors that involve additional economic indicators that affect tourist arrivals when modelling the data and expand the analysis to include multivariate time series modelling.

References

Agyemang, E. F., Mensah, J. A., Ocran, E., Opoku, E., & Nortey, E. N. N. (2023). Time series-based road traffic accidents forecasting via SARIMA and Facebook Prophet model with potential changepoints. Heliyon, 9(12). https://doi.org/10.1016/j.heliyon.2023.e22544

Chang, X., Gao, M., Wang, Y., & Hou, X. (2012). SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE MODEL FOR PRECIPITATION TIME SERIES. *Journal of Mathematics and Statistics*, *8*(4), 500–505. https://doi.org/10.3844/jmssp.2012.500.505

Elseidi, M. (2024). A hybrid Facebook Prophet-ARIMA framework for forecasting high-frequency temperature data. *Modeling Earth Systems and Environment*, *10*(2), 1855–1867. https://doi.org/10.1007/s40808-023-01874-4

Fernando, S. (2016). Managing the post-war tourism development in Sri Lanka. *International Journal of Business and Social Science, 7*(11), 90–100.

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice (3rd ed)*. Retrieved September 9, 2024, from https://otexts.com/fpp3/

Kodippili, A., & Senaratne, D. (2017). Forecasting Tourist Arrivals to Sri Lanka Using Seasonal ARIMA. *Journal of Tourism, Hospitality and Sports, 29*.

Lelwala, E. I., & Gunaratne, L. H. P. (2008). Modelling Tourism Demand using Cointegration Analysis: A Case Study for Tourists Arriving from United Kingdom to Sri Lanka. *Tropical Agricultural Research, 20*, 50–59.

Lewis, C. D. (1982). *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*. Butterworth Scientific. https://iucat.iu.edu/iue/3760302

Nagendrakumar, N., Lokeshwara, A. A., Gunawardana, S. A. D. C. K., Kodikara, U. P., Rajapaksha, R. W. N. H., & Rathnayake, K. R. M. C. S. (2021). Modelling and Forecasting Tourist Arrivals in Sri Lanka. *SLIIT Business Review*, *02*(01), 95–120.

Patandung, S., & Jatnika, I. (2021). The FB Prophet Model Application to the Growth Prediction of International Tourists in Indonesia during the COVID-19 Pandemic. *International Research Journal* *of Advanced Engineering and Science, 6*(2),110-115.

Peiris, H. R. I. (2016). A Seasonal ARIMA Model of Tourism Forecasting: The Case of Sri Lanka. *Journal of Tourism, Hospitality and Sports, 22*.

Priyangika, J. H., Pallawala, P. K. B. N. M., & Sooriyaarachchi, D. J. C. (2016). *Modelling and Forecasting Tourist Arrivals in Sri Lanka*. http://repository.kln.ac.lk/handle/123456789/15547

Silva, E. S., Hassani, H., Heravi, S., & Huang, X. (2019). Forecasting tourism demand with denoised neural networks. *Annals of Tourism Research, 74*, 134–154.

SLTDA | Sri Lanka Tourism Development Authority. (n.d.). SLTDA. Retrieved September 6, 2024, from https://www.sltda.gov.lk/en/monthly-tourist-arrivals-reports-2024

Taylor, S. J., & Letham, B. (2018). Forecasting at Scale. *The American Statistician*,*72*(1),37–45. https://doi.org/10.1080/00031305.2017.1380080

Yang, Y., & Zhang, H. (2019). Spatial-temporal forecasting of tourism demand. *Annals of Tourism Research, 75*, 106–119.