MethodsX 12 (2024) 102723

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/methodsx

Forecasting of Indian tourism industry using modeling approach

Renuka Devi[a], Alok Agrawal[a], Joydip Dhar[b], A.K. Misra[c,*]

[a] Amity School of Applied Sciences, Amity University Uttar Pradesh, Lucknow Campus, Lucknow 226 028, India
[b] Department of Applied Sciences, ABV-Indian Institute of Information Technology and Management, Gwalior 474 015, India
[c] Department of Mathematics, Institute of Science, Banaras Hindu University, Varanasi 221 005, India

REVIEW HIGHLIGHTS

- We have analyzed the tourism industry time-series data.
- We have considered the international tourist arrival variable to understand the relationship between the tourism industry and economic development.
- The forecasting method is applied to predict the future growth of tourism industry.

ARTICLE INFO

Method name:
Autoregressive integrated moving average method.

Keywords:
Mathematical model
SARIMA model
Tourism
Regression analysis
Forecasting

ABSTRACT

Currently, India has become one of the largest economies of the world in which tourism and hospitality have significantly contributed; however, the growth rate of tourism industry has been greatly affected during the COVID-19 pandemic. In this study, we have used the modeling approach to analyze and understand the growth pattern of Indian tourism industry. To achieve this, we consider the data of international tourist arrivals before and after the lockdown. The Dickey-Fuller test, AIC and BIC methods are used to obtain the best fitted model and further, the accuracy of obtained model is also analyzed. Data and forecasting indicate that the weather and public holidays significantly affect the tourism industry.

Specifications table

Subject area:	Mathematics and Statistics
More specific subject area:	Mathematical Modeling
Name of the reviewed methodology:	Modeling and forecasting the tourism industry
Keywords:	Mathematical model; SARIMA model; Tourism; Regression analysis; Forecasting.
Resource availability:	The data that support the findings of this study are available within the article.
Review question:	1. How the growth of tourism industry of India will grow? 2. What are the methods, we can use to forecast the tourist arrival? 3. How the international tourists arrival affected the growth of tourism industry after COVID-19?

* Corresponding author.
E-mail address: akmisra@bhu.ac.in (A.K. Misra).

https://doi.org/10.1016/j.mex.2024.102723
Received 5 March 2024; Accepted 16 April 2024
Available online 19 April 2024

Method details

Overview

The tourism industry stands as a significant pillar of the world's economy. The nature of tourism industry is highly complex due to its profound relationship with social, political, and cultural factors. To collect the data of tourism industry and analyze it, the United Nations World Tourism Organization (UNWTO) was formed, which later on renamed as UN Tourism and presently 160 countries are members of it. India became a member of the UN Tourism in 1975, and since then India is continuously improving its position in terms of ranking [1,2]. Tourism serves as a lifeline to preserve heritage sites and also generates employment opportunities for the local population, thereby alleviates poverty levels by offering income sources within communities. However, during the COVID -19 pandemic, lockdowns were abruptly imposed and borders were sealed, which completely shattered the world tourism industry due to restrictions on the movement of people. After the pandemic, the pace of movement has gradually returned to its original rhythm across the globe. Recent data of the UN Tourism reveals a significant recovery, with approximately 90% of the global tourism operations to pre-pandemic levels. In the initial eight months of 2023, an impressive 975 million individuals embarked on international journeys [3]. As the momentum of tourist arrivals continues to surge, the economy is poised for a robust upswing in 2024. In a yearly report of the UN Tourism, it is indicated that international tourism receipts have soared to an astounding USD 1.4 trillion in 2023 and contributed to the Indian economy only 3.5% less in comparison to 2019 [3]. India has several attractive places for tourists and the Indian government is paying much attention to the development of tourist places to attract both domestic and international tourists [4]. To foster the development of tourism industry, the Indian government has unveiled numerous schemes, such as Swadesh Darshan, Atmanirbhar Bharat, Azadi Ka Amrit Mahotsav, and Dekho Apna Desh, etc. [5–8]. These schemes motivate the rural as well as urban population to develop their regions for tourism purposes. Recently, the Indian government has administered the G-20 summit 2023 and organized tours at different places for the delegates to promote the tourism industry. Also, the Indian government has made efforts to establish strong connectivity between important tourist destinations under the scheme RCS: Udan-3 [9]. Further, various facilities such as medical tourism services, a 24/7 toll-free multilingual tourist helpline, transportation options, diverse accommodation choices, heritage tours, trekking packages, etc., are also being strengthened to promote the tourism industry. In order to preserve the authenticity and cater the tourists effectively, it is vital to effectively oversee current amenities while also expanding available facilities to accommodate increasing tourist numbers. However, increasing these facilities is not a simple endeavor; it necessitates adequate funding and meticulous planning. Therefore, anticipating tourist arrivals well in advance is crucial to mitigate uncertainties and bolster decision-making capabilities. In this regard, forecasting methods are crucial and play a vital role in the economic planning and policymaking of government to boost performance [10,11]. Accuracy in forecasting is pivotal for crafting effective tourism strategies. The SARIMA method, known for its precision in dealing with seasonality and provides valuable forecasts [12]. Additionally, other proxies, like tourist expenditures, sightseeing expenses, shopping records, and meal expenditures can also be utilized to forecast tourist arrivals [13].

To gain a comprehensive understanding of the long-term relationship between economic growth and international tourism, utilizing unit root tests and Johansen techniques proves to be beneficial [14]. To predict future growth accurately, the utilization of forecasting methodologies becomes essential. To address the complexities inherent in non-stationary data and intricate seasonal patterns, it is common to employ fundamental techniques, such as ARIMA (Auto Regressive Integrated Moving Average), seasonal ARIMA, Super Vector Regression, FB Prophet model, and Neural Networking methods [15–17]. To assess the accuracy of the forecast model, the Box-Jenkins method proves to be more suitable. This method relies on four performance criteria: Schwarz Bayesian Criterion (SBC), Akaike Information Criterion (AIC), standard error (SE), and maximum likelihood [18–20]. The accuracy of predictions relies on well-structured data and becomes crucial when dealing with large and complex time series data, encompassing multiple variables, such as weather temperature, seasonal variations influenced by weekends and public holidays, and web search queries related to the destination [21]. To analyze these relationships effectively, the Vector Autoregression (VAR) method may be used, which includes tests like the Granger causality test [22,23].

Literature review

Tourism occupies a central position globally and its impact is two-fold, encompassing both negative and positive dimensions [24]. The qualitative and quantitative research, leveraging ANOVA, provide insights into both dimensions. There are numerous challenges associated with rural tourism, e.g., underdeveloped river-side rural tourist spots, dealing with different cultures keeping people safe, handling the effects of climate change, improvement in healthcare facilities, fixing the conditions of roads, etc. [25]. To promote medical tourism, the Indian government organized special flights to Myanmar during the pandemic and when it recedes, the Indian government introduces luxury travel options and expands rural and spiritual tourism to attract more tourists [26]. For sustainable growth of tourism, the Indian government and tourism investors are working together to boost rural tourism by building infrastructure, providing amenities and arranging cultural experiences like home-stays, but improving road transportation services remains a challenging task. These challenges are not specific to India; they affect numerous other countries as well. Research conducted by Kumar and Stauvermann [27] on five small Pacific Islands Fiji, Samoa, Solomon Islands, Tonga, and Vanuatu has elucidated that tourism serves as a primary pillar for their economies; however, the insufficient infrastructure and amenities pose significant challenges to attract foreign tourists and thus hinders the potential growth. In this study, it is also pointed out that tourism and foreign direct investment share a symbiotic relationship, wherein sustainable and robust growth in tourism is essential to foster economic development. In fact, the considered model in [28] revolves around five key factors; namely, per capita GDP, visitor arrivals, FDI inflows to GDP, domestic credit and remittances inflow as a percent of GDP. Further, authors have conducted the unit root test analysis by using the Dickey-Fuller (ADF), Phillips-Perron (PP), and KPSS (Kwiatkowski) tests. Research conducted on BRICS countries using the ARDL model affirms the stable long-term relationship between tourism, financial development, and economic growth.

Forecasting methods are simple in nature but complex in application. There are several methods, like neural networking, artificial intelligence, random forest, support vector machine and multilayer perception, etc., which are applied to analyse the complicated time series data [29]. But, there is no single method of forecasting that can provide accurate results in all circumstances. The precision of a forecast relies on various parameters including duration, data size, number of variables, etc. However, in order to determine the best fitted model, it is necessary to assess the error [30]. To enhance the accuracy of predictions, the tourism and hospitality industry widely adopt advanced forecasting methods, such as ARIMA and ANN (Artificial Neural Networks). These sophisticated methods analyze historical data and patterns to generate reliable forecasts for future tourists' arrivals. By leveraging these forecasting methods, stakeholders in the tourism industry can make informed decisions, allocate resources efficiently, and adapt to changing market conditions with greater confidence. Thus, the integration of predictive analytics plays a crucial role in shaping the strategies and policies that drive the success of the tourism industry [31]. India's neighbouring countries, like Nepal, Bangladesh and Bhutan have done similar research to analyse the seasonal variation by considering the time series data [32,33]. The Zimbabwe national statistics agency has conducted the Box-Jenkins analysis by applying the ACF, PACF, RMSE methods and revealed that seasonal ARIMA fits well to the data. The analysis concludes that there is a seasonal and gradual increase in international tourist arrivals [34]. It may be noted that ARIMA modeling is not only applied to tourism and hospitality industry but also in financial sectors, marketing, stock market, etc. [35].

Data information

The UN Tourism evaluates country's performance on various criteria, one of which is International Tourist Arrivals (ITA). ITA consists of two primary components: foreign tourists' arrivals and arrivals of non-resident nationals. In this study, the data has been segregated into two segments due to the COVID - 19 pandemic lockdown period (year 2020 and 2021). The first segment encompasses yearly data of pre-pandemic international tourists' arrivals in India from 1995 to 2019, and the ARIMA method is applied to forecast the yearly future growth of tourist arrivals. The second segment comprises post-pandemic monthly data spanning from January 2022 to November 2023, aimed to calculate the monthly growth rate of tourist arrival by applying the SARIMA method. Forecasted ITA data will be proved invaluable in optimizing and implementing tourism policies and forecasting India's growth on the UN Tourism barometer. The forecasted ITA data will serve as a guide for the government in formulating strategies for international tourism planning. The data source for this research is the UN Tourism database [36]. In this study, R-programming and Excel software are used to calculate the growth rate of ITA and to avoid the data biasness and inconsistency, we have used a valid data source and there is no missing value to incorporate.

Model derivation

This research work has a primary focus on the use of autoregressive integrated moving average (ARIMA) model to study the growth of tourism industry. This model is also known as the Box-Jenkins (1976) forecasting model and is listed among the best forecasting methodologies [37]. Box-Jenkins model is a combination of three methods, i.e., autoregressive, integration, and moving average. This model follows a step by step process; namely, (i) identification, (ii) estimation, (iii) diagnostic checking, and (iv) forecasting.

Identification

In time series analysis, a sequence of observations is arranged in chronological order. In ARIMA modeling, the identification stage plays a crucial role in determining whether the time series data is stationary or non-stationary. If the time series data is non-stationary, then it is necessary to assess how many differences are needed to render it stationary. One common method to visually inspect the trend in the series is graphical representation i.e., plotting the variable against time. Moreover, to check the stationarity of the time series, there are also some mathematical conditions, which are listed below:

(i) The mean value of the time series should be constant ($\mu = constant$).

(ii) The variance of the time series should be constant ($\sigma^2 = constant$).

(iii) No seasonality should be in the time series.

In this research work, we have applied three methods to check the stationarity of the time series, which are mentioned as follows.

Autocorrelation Function

ACF is the coefficient correlation between the time and lagged time value. If the time series is y_t, then

$$\mathtt{ACF} = \mathtt{corr}\ (y_t, y_{t-k}). \tag{1}$$

Here, k is any natural number representing the time lag.

Partial Autocorrelation Function

The PACF plot represents the partial correlation coefficients between the series and lags of itself. The partial correlation between two variables is the amount of correlation between them which is not explained by their mutual correlations with a specified set of other variables.

Augmented Dickey-Fuller Test

ADF test is one of the approach of unit root test to check the stationarity in the time series data. If the observed time series data is $y_1, y_2, ..., y_n$, then the ADF considers the three differential forms of the autoregressive equations:

$$\Delta y_t = \gamma y_{t-1} + \sum_{j=1}^{p}(\delta_j \Delta y_{t-j}) + \epsilon_t, \tag{2}$$

$$\Delta y_t = \alpha + \gamma y_{t-1} + \sum_{j=1}^{p}(\delta_j \Delta y_{t-j}) + \epsilon_t, \tag{3}$$

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{j=1}^{p}(\delta_j \Delta y_{t-j}) + \epsilon_t, \tag{4}$$

where t, α, β, γ and ϵ_t sequentially represents the time index, an intercept constant called a drift, the coefficient on a time trend, the coefficient presenting process root, i.e., the focus of testing and an independent identically distribution residual term.

Estimation

This research is focused on the autoregressive integrated moving average (ARIMA) and seasonal autoregressive integrated moving average (SARIMA) modeling. ARIMA is defined in terms of (p, d, q) model and SARIMA in terms of $(p, d, q)(P, D, Q)s$, where

p = non seasonal autoregressive (AR) order,
d = non seasonal differencing,
q = non seasonal moving average (MA) order,
P = seasonal autoregressive (AR) order,
D = seasonal differencing,
Q = seasonal moving average (MA) order,
s = seasonal order.

The ARIMA and SARIMA model can be expressed respectively as follows:

$$y_t = c + \phi_1 y_{t-1} + ... + \phi_p y_{(t-p)} + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q} + \epsilon_t, \tag{5}$$

$$\phi_p(O)\Phi_P(O^s)(1-O)^d(1-O^s)^D + y_t = \theta_q(O)\Theta_Q(O^s)\epsilon_t. \tag{6}$$

In the following, we define the terms occurring in the above equation. The nonseasonal AR(p) order model:

$$\phi_p(O) = 1 - \phi_1 O - \phi_2 O^2 - .... - \phi_p O^p. \tag{7}$$

The nonseasonal MA(q) order model:

$$\theta_q(O) = 1 - \theta_1 O - \theta_2 O^2 - .... - \theta_q O^q. \tag{8}$$

The seasonal AR(P) order model:

$$\Phi_P(O^s) = 1 - \Phi_1 O^s - \Phi_2 O^{2s} - .... - \Phi_P O^{Ps}. \tag{9}$$

The seasonal MA(Q) order model:

$$\Theta_Q(O^s) = 1 - \Theta_1 O^s - \Theta_2 O^{2s} - .... - \Theta_Q O^{Qs}. \tag{10}$$

$(1-O)^d$ = Nonseasonal differncing of order d.

$(1-O)^D$ = Seasonal differencing of order D.

Here, ϵ_t and O are error term and backshift operator, respectively.

Diagnostic checking

Diagnostic checking means to find out the best fitted model out of the suggested models. This step is primarily based on the two information criteria, AIC and BIC. AIC and BIC criterion are applied to compare the models of the ARIMA.

Akaike Information Criterion (AIC)

AIC is a unbiased predictive accuracy model. AIC is applied when the data size is big. AIC is defined as

$$AIC = -2 \ln L + 2k. \tag{11}$$

Fig. 1. International tourist arrivals (ITA) in India during 1995–2019.

However, when the sample size is small, AIC is biased and the corrected version of AIC is AICc. AICc is defined as

$$AICc = AIC + \frac{2k(k+1)}{N-k-1}. \tag{12}$$

Schwarz Bayesian Information Criterion (BIC)
BIC evaluates that whether the generated estimate is true or probably close to the truth. BIC is defined as

$$BIC = -2\ln L + k\ln N. \tag{13}$$

Here, $\ln L$, k and N represent the maximized log likelihood, number of estimated parameters and number of observations, respectively.

Model validation and forecasting
Forecasting is the method to calculate the future values based on the past time series. ARIMA model provides the forecasted values on the basis of the best fitted model with minimum standard error. The accuracy of the model is based on the root mean square error (RMSE), mean percentage error (MPE), mean absolute error (MAE), and these can be calculated using the following formulas.
Root mean square error (RMSE)

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}. \tag{14}$$

Mean absolute error (MAE)

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|. \tag{15}$$

Mean absolute percentage error (MAPE)

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{y_i} \times 100. \tag{16}$$

To analyse the growth of Indian tourism industry and the future aspects, this research is conducted for two different time intervals, i.e., pre and post COVID-19 pandemic. As this study follows the process of the ARIMA modeling that we have discussed before and so in accordance with the process, here we have plotted the ITA (ARIMA) and ITAM (SARIMA) time series.

Non-stationary time series: The graphical representation of the data set in Figs. 1 and 2 explains that international tourists arrival growth rates are non-stationary. Also, Fig. 1 shows an upward trend; however, Fig. 2 explains that there is an upward trend & seasonality in the month of July and December.

To make any conclusion regarding the stationarity of the time series data, we have shown the ACF and PACF lag values in Figs. 3 and 4, respectively. ACF spikes are moving towards zero and there is also a pattern in spikes.

We have applied Augmented Dickey-Fuller test, which provides the t-statistic values as (−0.74027, −1.2463) and the p-values as (0.955, 0.8595) for ARIMA and SARIMA, respectively as depicted in Table 1. Here, it may be noted that the p value is greater than 0.05 and thus the time series data is non-stationary.

Fig. 2. International tourist arrivals (ITAM) in India during 2022–2023.

Table 1
Augmented Dickey-Fuller test.

Model	Test statistic	lag order	p-value
ARIMA	−0.74027	2	0.955
SARIMA	−1.2463	2	0.8595

Table 2
ARIMA and SARIMA models summary.

	ARIMA models			SARIMA models	
(p,d,q)	AIC value	BIC value	(p,d,q)(P,D,Q)s	AIC value	BIC value
(2,2,2)	722.4141	728.0916	(0,1,0)(0,1,0)[12]	185.1339	185.0798
(0,2,0)	729.7172	730.8527	(0,1,1)(0,1,0)[12]	184.2807	184.1725
(1,2,0)	722.8065	725.0775	(0,1,2)(0,1,0)[12]	185.9365	185.7742
(0,2,1)	717.1264	719.3974	(1,1,0)(0,1,0)[12]	183.9314	183.8232
(1,2,1)	718.4687	721.8751	(1,1,1)(0,1,0)[12]	185.8102	185.6479
(0,2,2)	718.4168	721.8233	(1,1,2)(0,1,0)[12]	187.8024	187.5861
(1,2,2)	720.4142	724.9562	(2,1,0)(0,1,0)[12]	185.8178	185.6555

The time series should be stationary to apply the ARIMA and SARIMA modeling. So, we have applied differencing method to make the non-stationary time series data stationary. For the second difference, the ITA time series (I = 2) is transformed into stationary. Same method has been applied in ITAM series, where the first difference (I = 1) converted the time series data into stationary time series data. Figs. 5 and 6 represent the ACF and PACF stationary time series graphical image. The spikes of the ACF and PACF are towards zero.

Model Summary: It is interesting to note that the ARIMA (p,d,q) and SARIMA (p,d,q)(P,D,Q)s consist different values. There are multiple models that define the different combination of these values but the main challenge is to figure out the best fitted model that forecasts the future international tourist arrivals based on the available data. As discussed in the previous section that to select the best fitted model, the AIC and BIC information criterion methods are used and so in Table 2, we have shown the AIC and BIC values for the ARIMA and SARIMA models.

Model summary yields different models on the basis of AIC and BIC values and after careful evaluation of the outcomes, we select ARIMA (0,2,1) model (where p = 0, d = 2, q = 1) and SARIMA (1,1,0)(0,1,0)[12] model (where p=1, d=1, q=0, P = 0, D = 1, Q = 0) as the best forecasted models. This selection is based on the smallest values of the AIC and BIC, a criterion outlined earlier. Here, we note that for the ARIMA model: AIC = 717.1264 and BIC = 719.3974, and for the SARIMA model: AIC = 183.9314 and BIC = 183.8232 (Table 3).

Accuracy testing: Box-Ljung and error anaylising are two methods to check the accuracy level of the ARIMA and SARIMA models, which are given in Tables 4 and 5, respectively.

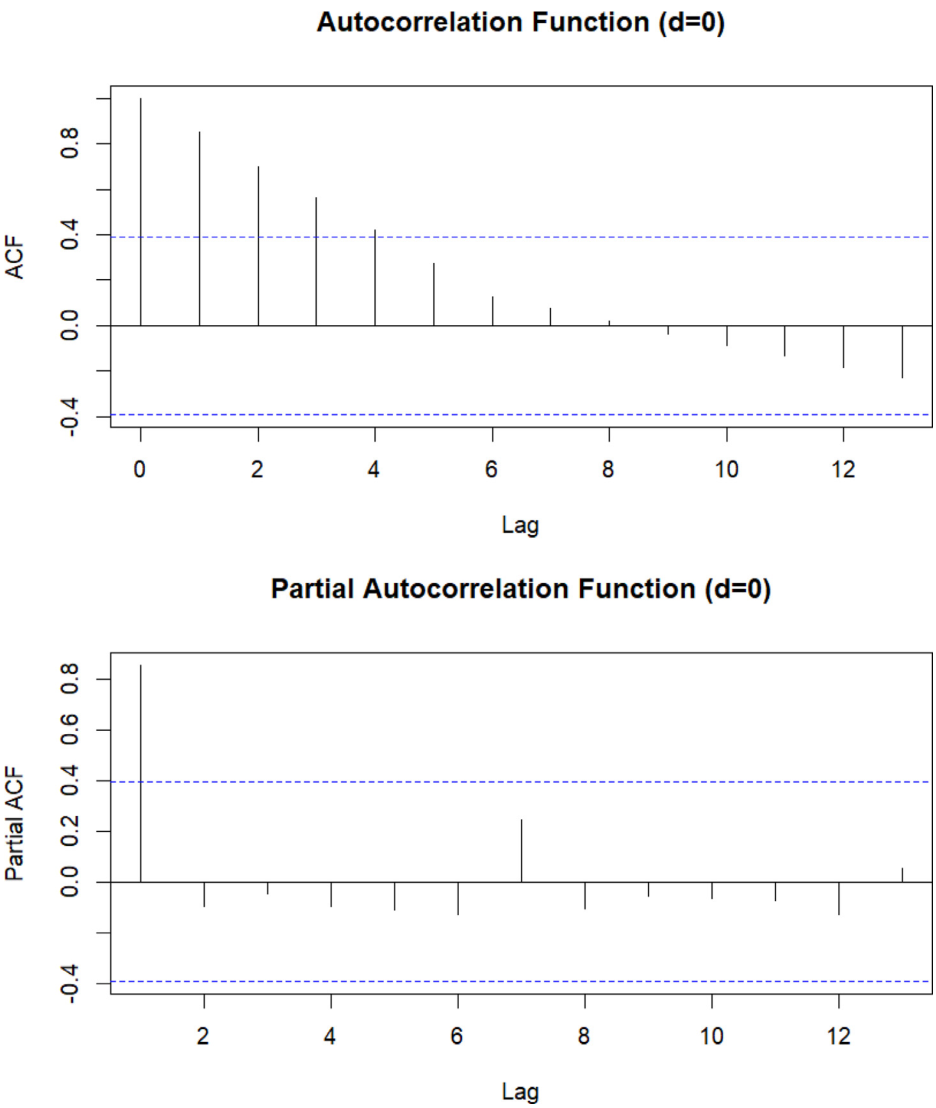

Fig. 3. ITA Autocorrelation and partial- autocorrelation function (d=0).

Table 3
Model fitted test result.

Coefficients regarding ARIMA model		Coefficients regarding SARIMA model	
coefficients:		coefficients:	
ma1	−0.8372	ar1	0.5720
s.e. 0.1156	σ^2 1.697e+12	s.e. 0.2658	σ^2 9401915806
AIC 1717.13	AICc 717.73	AIC 183.9314	AICc 186.93
BIC 179.4	loglikelihood −356.56	BIC 183.8232	loglikelihood −89.97

Table 4
Box-Ljung test.

ARIMA			SARIMA		
x-squard value	df	p-value	x-squard value	df	p-value
4.7501	5	0.4471	2.2661	5	0.8112
5.2014	10	0.8773	2.3448	10	0.9929
5.6321	15	0.9853	2.3929	12	0.9985
5.9389	20	0.999	2.4602	15	0.9999

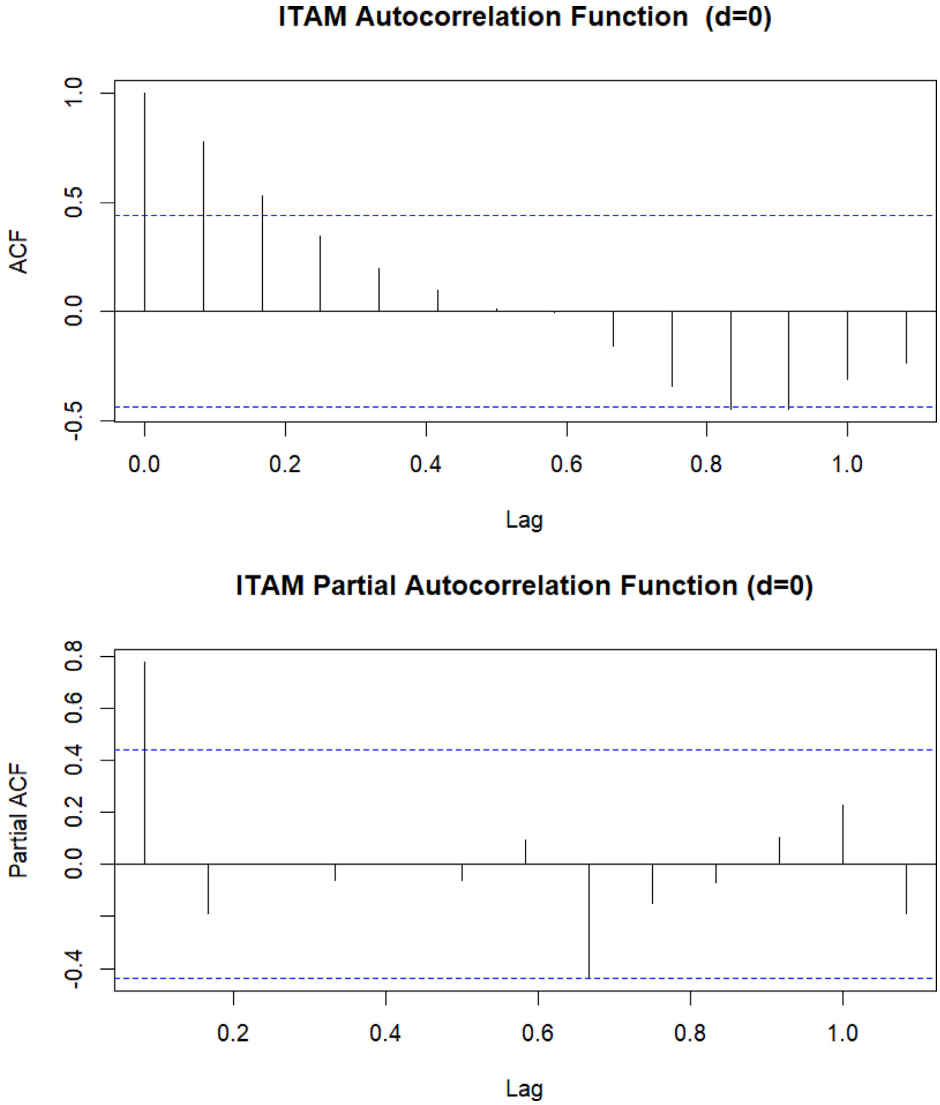

Fig. 4. ITAM- Autocorrelation and partial autocorrelation function (d=0).

Table 5
Accuracy measures.

	ARIMA	SARIMA
ME	242756.3	−9921.779
RMSE	1221862	53109.08
MAE	505907.5	25562.88
MPE	2.573478	−1.289702
MAPE	6.654961	3.79307
MASE	0.7323149	0.081127
ACF_1	−0.2321292	0.1149802

In Table 4, we can see that p - values for ARIMA and SARIMA models at the difference (5) are 0.4471 and 0.8112, respectively; however these are 0.9853 and 0.9999, respectively at the difference (15). It may be noted that the p - values at the difference (15) are greater than at the difference (5) and there is also a consistence increase in the p - values, which suggests that this model is well-equipped to forecast the future growth of international tourists' arrivals (y_t) up to the year 2025.

We have conducted error analysis for different models and compared them to obtain the best fitted model with the smallest value of ME, RMSE, MAE. The values of ME, RSME, MAE, MPE, MAPE, MASE, ACF_1 are provided in Table 5 for the selected ARIMA and SARIMA models.

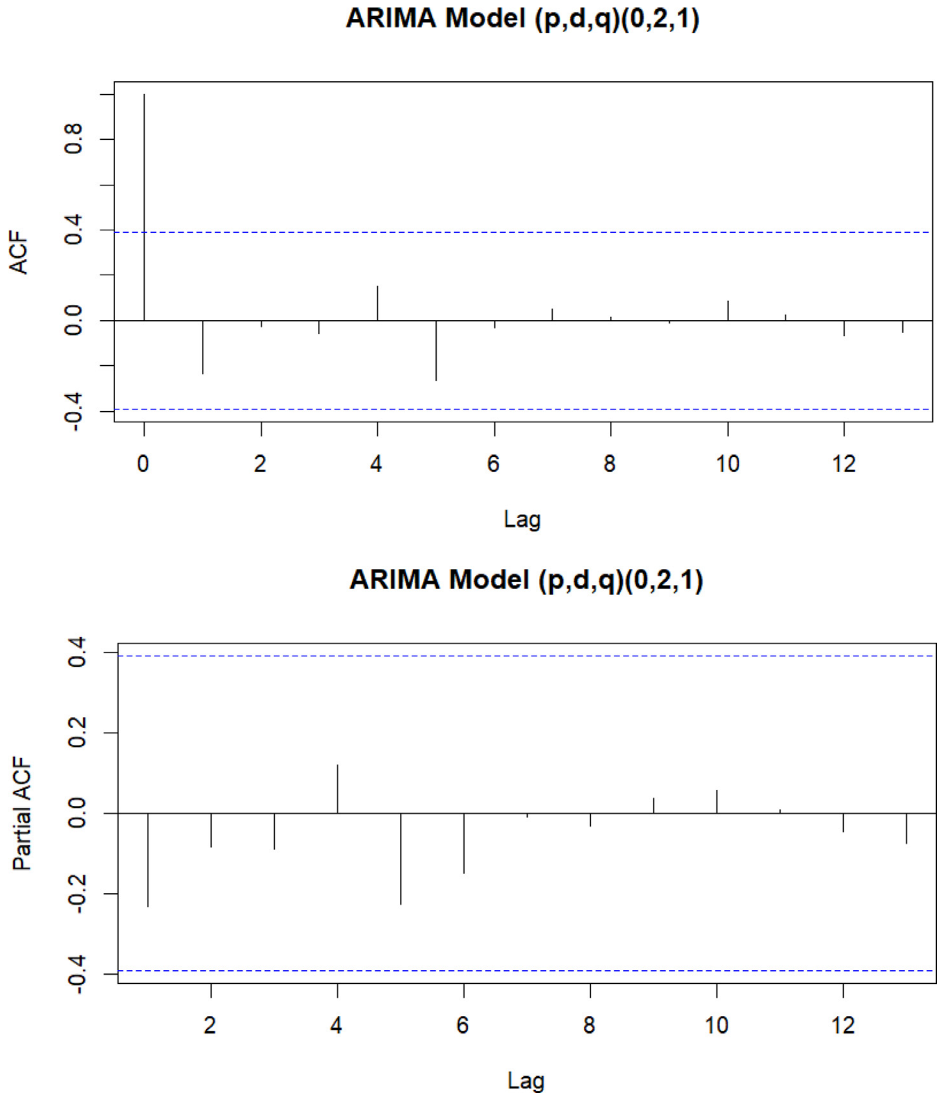

Fig. 5. ITA Autocorrelation and partial autocorrelation function (d=2).

Forecast results: In Figs. 7 and 8, we have provided a graphical representation of the future international (yearly and monthly, respectively) tourists' arrivals growth in India for the coming year 2025. While generating these graphs, we have excluded the available ITAM data for three months September, October and November, 2023. The reason behind this exclusion is that we want to compare our forecasted results with the actual values. It is interesting to note that forecasted values are very close to the actual values, as the actual values for September, October, November in the year 2023 are 648213, 811411, 922265, respectively and forecasted values for September, October, November in the year 2023 are 695696, 824490, 940983, respectively. This demonstrates that the forecasted values are close to the actual values. In Fig. 9 and Table 6, we have provided the forecasted values of monthly international tourists' arrivals in India from September 2023 to August 2025.

This research specifically examines the growth of international tourists' arrivals (ITA). Analysis of yearly ITA data reveals a consistent positive trend from 2022 to 2025. The sustained positive growth of international tourists' arrivals not only facilitates the establishment of new international relationships, but also serve as a magnet for investors, encouraging investments across various sectors, such as healthcare facilities, real estate, infrastructure, and financial markets. Consequently, this consistent growth trajectory will alleviate government expenditures. Moreover, the influx of international tourists injects significant capital into the country, contributing to its financial and social stability. The primary focus of this research study is to analyze seasonal trends of tourists' arrival between the period January, 2022 and November, 2023. The findings of this study indicate that growth rates of tourists arrival are notably high in July and December months in comparison to other months. During the period from November to January

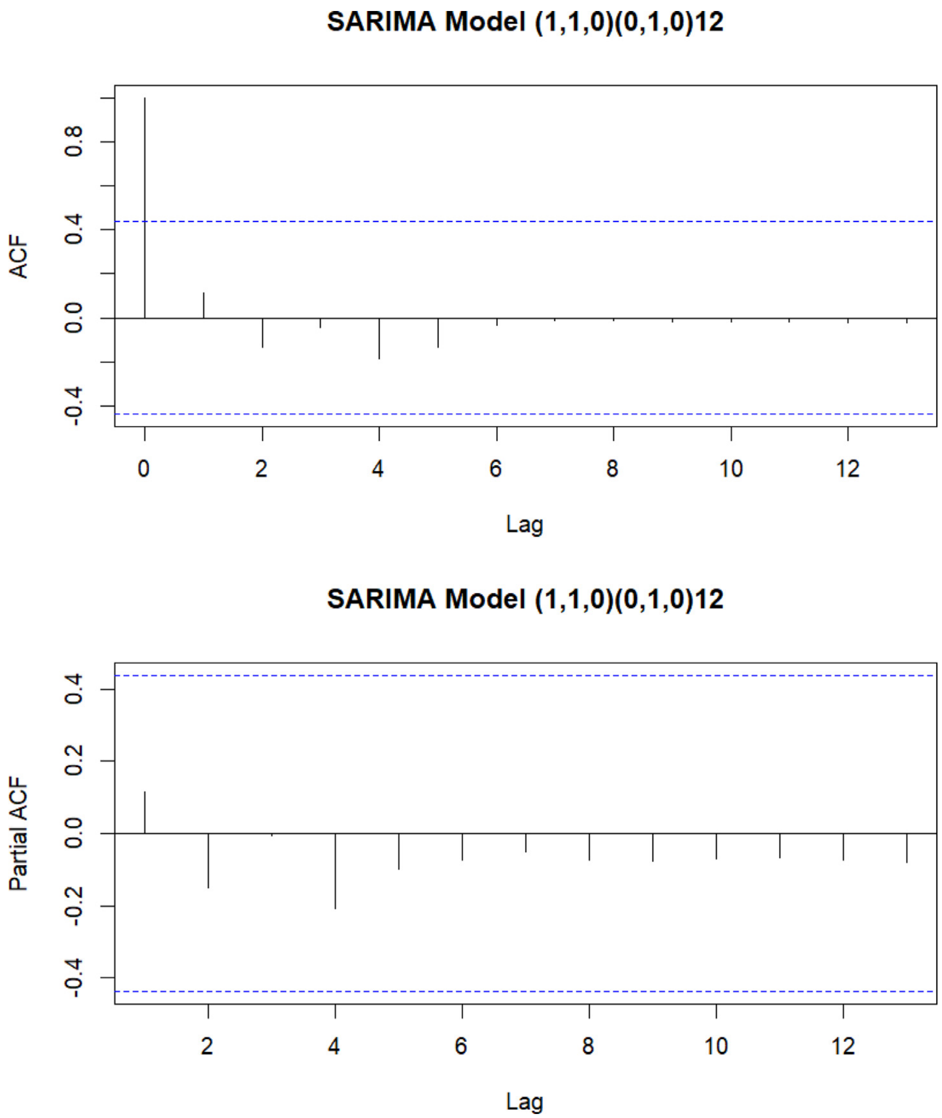

Fig. 6. ITAM Autocorrelation and partial autocorrelation function (d=1).

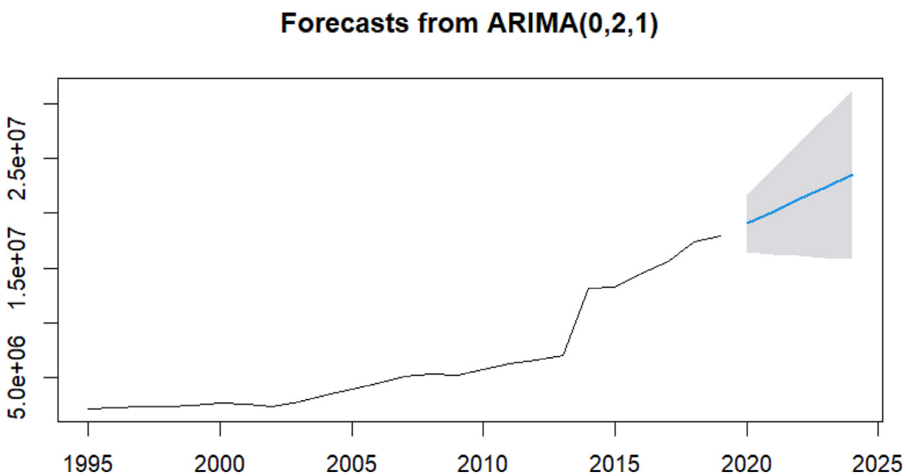

Fig. 7. Forecast international tourist arrival (ITA) in India year from 2020–2025.

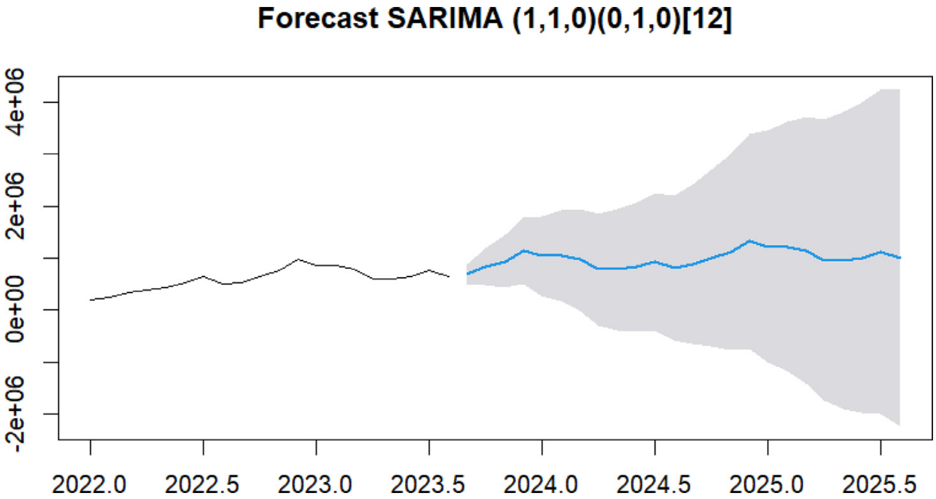

Fig. 8. Forecast international tourist arrival (ITAM) in India year from 2024–2025.

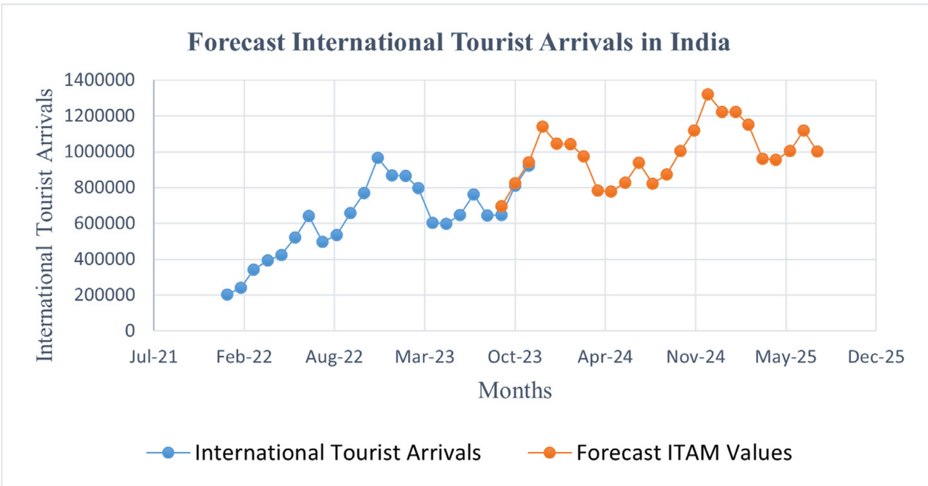

Fig. 9. Actual and forecasted international tourist arrival in India year from 2024–2025.

and June to August, the growth rate peaks before gradually tapering off. The forecasted values for international tourists' arrivals also reflect a similar trend. Here, it may be noted that the factors, like public holidays, weather temperatures, and festivals may contribute to this growth but the data related to these factors is not explicitly incorporated in this study. The forecasted results can serve as a valuable guide for the government to enhance various facilities, such as road transportation, specialized helpline services, security measures, hospitality standards, healthcare facilities, etc. timely to fulfil the demands of tourists and ensure a satisfactory experience by the tourists. Additionally, the findings from this model can guide tourism marketing agencies to develop effective tourist policies, stimulating currency inflow and investment in the tourism industry. Forecasting models offer valuable insights, yet they also come with certain limitations as no single model can accommodate all factors, such as the number of variables, time series patterns, data size, hidden variables, etc. In the machine learing, there also exists some methods, such as GARCH, ARCH, and ANN, LSTM approach, neural networking, etc., each suited to specific situations [38,39]. The challenge lies in identifying the most suitable method for both endogenous and exogenous variables. Further, in case of more variables and their dependency on each other, the other modeling approaches, like deterministic, stochastic, etc., can be used to predict the growth of tourism industry. However, the utilization of mathematical tools can be a complex endeavor due to the unavailability of detailed datasets. To handle the missing values and inconsistency of the data is the biggest challenge. To endure these complexities, in the future, we will analyse the big dataset and apply the machine learning tools for deep analysis.

Table 6
Forecasted values.

Time	Forecast Value	Lo 95	Hi 95
Dec 2023	1141273.8	498766.010	1783781.7
Jan 2024	1044705.5	278840.203	1810570.9
Feb 2024	1043206.3	165643.777	1920768.8
Mar 2024	973758.7	−5745.321	1953262.6
Apr 2024	782205.1	−291138.360	1855548.6
May 2023	776805.1	−383634.404	1937244.6
Jun 2023	826487.5	−415395.500	2068370.5
Jul 2023	939156.5	−379387.157	2257700.1
Aug 2023	821758.3	−569357.136	2212873.8
Sep 2023	874278.4	−654384.085	2402941.0
Oct 2023	1003082.9	−696387.305	2702553.1
Nov 2023	1119581.3	−760819.840	2999982.4
Dec 2023	1319875.0	−739783.380	3379533.4
Jan 2025	1223308.6	−1008981.575	3455598.8
Feb 2025	1221810.5	−1174902.503	3618523.4
Mar 2025	1152363.4	−1400481.027	3705207.9
Apr 2025	960810.2	−1740395.079	3662015.6
May 2025	955410.4	−1887098.771	3797919.7
Jun 2025	1005092.9	−1972391.763	3982577.6
Jul 2025	1117762.0	−1989042.795	4224566.7
Aug 2025	1000363.9	−2230700.993	4231428.8

Ethics statements

Not applicable

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Renuka Devi: Formal analysis, Methodology, Software, Writing – review & editing. **Alok Agrawal:** Formal analysis, Methodology, Writing – review & editing, Supervision. **Joydip Dhar:** Software, Validation, Writing – original draft, Writing – review & editing, Supervision. **A.K. Misra:** Writing – review & editing, Supervision.

Data availability

The data was used for the research described in the article.

Acknowledgment

Authors are thankful to the section editor and both the anonymous reviewers for their constructive suggestions which have improved the quality of this paper. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] UNWTO, UN Tourism World Organisation, About UN Tourism, Member States, 2024, Available at: https://www.unwto.org/member-states#heading64.
[2] UNWTO, UN Tourism World Organisation, The UN Tourism Data Dashboard, 2023a, Available at: https://www.unwto.org/un-tourism-world-tourism-barometer-data.
[3] UNWTO, UN Tourism world organisation, International Tourism to End 2023 Close to 90% of Pre-Pandemic Levels, 2023b, Available at: https://www.unwto.org/news/international-tourism-to-end-2023-close-to-90-of-pre-pandemic-levels.
[4] Vice President's Secretariat, Press Information Bureau Government of India, explore domestic tourist destinations first before looking at international travel- Vice President to Indian Tourists, 2022, Available at: https://pib.gov.in/PressReleasePage.aspx?PRID=1862624.
[5] Government of India, Ministry of Labour & Employment, Aatmanirbhar Bharat Rojgar Yojana (ABRY), 2024, Available at: https://labour.gov.in/aatmanirbhar-bharat-rojgar-yojana-abry.
[6] Ministry of Tourism, Government of India, Dekho Apna Desh Webinars under Azadi Ka Amrit Mahotsav (AKAM), 2024, Available at: https://tourism.gov.in/dekho-apna-desh-webinars-under-azadi-ka-amrit-mahotsav-akam.
[7] Ministry of Tourism, Government of India, Swadesh Darshan Guideline, Booklet, 2023, Available at: https://tourism.gov.in/sites/default/files/2023-04/Swadesh.
[8] Ministry of Culture, Government of India, Azadi Ka Amrit Mahotsav, 2024, Available at: https://amritmahotsav.nic.in/.
[9] Press Information Bureau, Government of India, Ministry of Civil Aviation, Posted On: 26 OCT 2023 by PIB Delhi, Available at: https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1971506.

[10] A.G. Georgantopoulos, Tourism expansion and economic development: VAR VECM analysis and forecasts for the case of India, Asian Econ. Financ. Rev. 3 (4) (2013) 464–482. Available at SSRN: https://ssrn.com/abstract=2250382
[11] R. Dong, L. Sun, L. Ma, A.A. Heidari, X. Zhou, H. Chen, Boosting kernel search optimizer with slime mould foraging behavior for combined economic emission dispatch problems, J. Bionic Eng. 20 (6) (2023) 2863–2895, doi:10.1007/s42235-023-00408-z.
[12] S.C. Thushara, J.J. Su, J.S. Bandara, Forecasting international tourist arrivals in formulating tourism strategies and planning: the case of Sri Lanka, Cogent Econ. Finance 7 (1) (2019) 1699884, doi:10.1080/23322039.2019.1699884.
[13] P.C. Padhan, Forecasting international tourists footfalls in India: an assortment of competing models, Int. J. Bus. Manag. 6 (5) (2011) 190, doi:10.5539/ijbm.v6n5p190.
[14] J. Balaguer, M. Cantavella-Jordá, Tourism as a long-run economic growth factor: the spanish case, Appl. Econ. 34 (7) (2002) 877–884, doi:10.1080/00036840110058923.
[15] M. Akin, A novel approach to model selection in tourism demand modeling, Tour. Manag. 48 (2015) 64–72, doi:10.1016/j.tourman.2014.11.004.
[16] Y. Liu, G. Feng, K.S. Chin, S. Sun, S. Wang, Daily tourism demand forecasting: the impact of complex seasonal patterns and holiday effects, Curr. Issues Tourism 26 (10) (2023) 1573–1592, doi:10.1080/13683500.2022.2060067.
[17] Y. Jin, A. Zhao, Bert-based graph unlinked embedding for sentiment analysis, Complex Intell. Syst. (2023) 1–12, doi:10.1007/s40747-023-01289-9.
[18] F.L. Chu, Forecasting tourism demand with ARMA-based methods, Tour. Manag. 30 (5) (2009) 740–751, doi:10.1016/j.tourman.2008.10.016.
[19] J. Fattah, L. Ezzine, Z. Aman, H.E. Moussami, A. Lachhab, Forecasting of demand using ARIMA model, Int. J. Eng. Bus. Manag. 10 (2018), doi:10.1177/1847979018808673. 1847979018808673
[20] B. Petrevska, Predicting tourism demand by ARIMA models, Econ. Res.-Ekon. Istraz. 30 (1) (2017) 939–950, doi:10.1080/1331677X.2017.1314822.
[21] D. Hou, W. Zhou, Q. Zhang, K. Zhang, J. Fang, A comparative study of different variable selection methods based on numerical simulation and empirical analysis, PeerJ Comput. Sci. 9 (2023) e1522, doi:10.7717/peerj-cs.1522.
[22] Y.Y. Liu, F.M. Tseng, Y.H. Tseng, Big data analytics for forecasting tourism destination arrivals with the applied vector autoregression model, Technol. Forecast. Soc. Change 130 (2018) 123–134, doi:10.1016/j.techfore.2018.01.018.
[23] J. Svane, T. Wiktorski, S. Ørn, T.C. Eftestøl, Optimizing support vector machines and autoregressive integrated moving average methods for heart rate variability data correction, MethodsX 11 (2023) 102381, doi:10.1016/j.mex.2023.102381.
[24] V. Saluja, S. Anand, H. Kumar, J. Peng, The perceived impact of tourism development and sustainable strategies for residents of Varkala, South India, Int. J. Geoheritage Parks 10 (2) (2022) 184–195, doi:10.1016/j.ijgeop.2022.03.003.
[25] R. Jangra, S. Kaushik, S.S. Saini, An analysis of tourist's perceptions toward tourism development: study of cold desert destination, India, Geogr. Sustain. 2 (2021) 48–58, doi:10.1016/j.geosus.2021.02.004.
[26] S.B. Dash, P. Sharma, Reviving indian tourism amid the COVID-19 pandemic: Challenges and workable solutions, J. Destin. Mark. Manag. 22 (2021) 100648, doi:10.1016/j.jdmm.2021.100648.
[27] R.R. Kumar, P.J. Stauvermann, Tourism and economic growth in the pacific region: evidence from five small island economies, J. Asia Pac. Econ. 28 (3) (2023) 894–921, doi:10.1080/13547860.2021.1944796.
[28] H. Rasool, S. Maqbool, M. Tarique, The relationship between tourism and economic growth among BRICS countries: a panel cointegration analysis, Future Bus. J. 7 (1) (2021) 1–11, doi:10.1186/s43093-020-00048-3.
[29] H. Song, G. Li, Y. Cai, Tourism forecasting competition in the time of COVID-19: an assessment of ex ante forecasts, Ann. Tour. Res. 96 (2022) 103445, doi:10.1016/j.annals.2022.103445.
[30] H. Hassani, E.S. Silva, N. Antonakakis, G. Filis, R. Gupta, Forecasting accuracy evaluation of tourist arrivals, Ann. Tour. Res. 63 (2017) 112–127, doi:10.1016/j.annals.2017.01.008.
[31] W. Jun, L. Yuyan, T. Lingyu, G. Peng, Modeling a combined forecast algorithm based on sequence patterns and near characteristics: an application for tourism demand forecasting, Chaos Solit. Fractals 108 (2018) 136–147, doi:10.1016/j.chaos.2018.01.028.
[32] E.H. Singh, Forecasting tourist inflow in bhutan using seasonal ARIMA, IJSR 2 (9) (2013) 242–245. Available at: https://www.ijsr.net/getabstract.php?paperid=10091306
[33] N.K. Tharu, Forecasting international tourists arrival in nepal: an application of ARIMA, department of statistics, 2019, Available at: https://www.unescap.org/sites/default/files/APS2020/81_Forecasting_of_International_Tourists_Arrival_in_Nepal_An_application_of_ARIMA.pdf.
[34] T. Makoni, G. Mazuruse, B. Nyagadza, International tourist arrivals modelling and forecasting: a case of zimbabwe, Sustain. Technol. Entrepreneurship 2 (1) (2023) 100027, doi:10.1016/j.stae.2022.100027.
[35] S. Guefano, J.G. Tamba, T.E.W. Azong, L. Monkam, Methodology for forecasting electricity consumption by grey and vector autoregressive models, MethodsX 8 (2021) 101296, doi:10.1016/j.mex.2021.101296.
[36] UNWTO, UN tourism world organisation, global and regional tourism performance, 2024, Available at: https://www.unwto.org/tourism-data/global-and-regional-tourism-performance.
[37] G.E. Box, G.M. Jenkins, Time Series Analysis: Forecasting and Control, Revised Edition, California: Holden Day, San Francisco, 1970.
[38] H. Yang, H. Liu, G. Li, A novel prediction model based on decomposition-integration and error correction for COVID-19 daily confirmed and death cases, Comput. Biol. Med. 156 (2023) 106674, doi:10.1016/j.compbiomed.2023.106674.
[39] D. Gursoy, Y. Li, H. Song, ChatGPT and the hospitality and tourism industry: an overview of current trends and future research directions, J. Hosp. Mark. Manag. 32 (5) (2023) 579–592, doi:10.1080/19368623.2023.2211993.