

A Framework of Predictive Analysis of Tourist Inflow in the Beaches of West Bengal: A Study of Digha-Mandarmoni Beach

Sanjana Mondal¹ and Jaydip Sen²(✉)

¹ Department of Marketing Management, Calcutta Business School,
Diamond Harbour Road, Bishnupur 743503, West Bengal, India
sanjanam@calcuttabusinessschool.org

² Department of Information Technology and Analytics, Praxis Business School,
Bakrahat Road, Thakurpukur, Off Diamond Harbor Road, Kolkata 700104, India
jaydip.sen@acm.org

Abstract. Tourism is increasingly becoming an extremely important sector with its rapidly increasing contribution to GDP of any state or country as a whole. Analyzing and predicting tourist inflow not only enables us to make an accurate estimate of the number of tourists that is likely to visit a destination, but it also provides us with an opportunity to gear up the capacities of that place in terms of logistics, hospitality etc. in order to cater to the tourists leading to an overall socio economic development of the place. This paper presents a study on the tourism demand for two very popular beaches of the state of West Bengal in India. In this work, time series values of the domestic tourist inflow to Digha and Mandarmoni beaches in West Bengal are used for the period of January 2008–December 2014. The time series is decomposed into its components – trend, seasonal, and random – in order to make further analysis. Based on the structural analysis, five different approaches of forecasting are formulated and the forecast accuracy is computed for each of the methods. Using R statistical tool, extensive results have been presented that provide very meaningful insights to the tourists' inflow time series. The results also demonstrate the effectiveness of our proposed forecasting framework.

Keywords: Forecasting · Tourist inflow · Holt-Winters forecasting model · ARIMA · Time series decomposition · Trend · Seasonality · R programming language

1 Introduction

The contribution of tourism industry in a country like India has grown significantly over the last few decades. Accordingly, tourism research has also become a topic of research that is of critical importance. The inherent potential of the tourism industry has driven the interest among the researchers to make comprehensive scan on various aspects of this industry with an aim to understand how to harness several benefits while avoiding the pitfalls associated with it. There are various approaches to tourism research based on the interest and objective of the researchers [1–6]. In the literature,

tourism research has found itself in several different perspectives. For example, the anthropologists and sociologists have investigated the cultural implications, the economists have examined the monetary aspects, the ecologists have explored the effects of tourism on local biospheres, while the analysts have attempted to investigate and understand the tourist pattern in order to forecast tourist inflow in future for formulating various strategies. In recent years, there has been an even more growing emphasis on combining multiple disciplines for a more holistic understanding of the tourism sector. Considering the tourists as the major stake holders of the tourism sector, researchers of marketing have intended to understand the behavioral pattern and preferences of the tourists in order to provide best possible value to the services provided to them. Since each destination is somewhat unique in terms of its attraction and offer to the tourists, it is mandatory to focus on each destination in order to understand its specific tourism pattern. Analyzing tourists' arrival pattern at a destination over a period of time helps in understanding the popularity the place and the seasonal nature of the data so that the demand can be modeled and forecasted effectively.

This research work aims in creating a predictive model by using the time series of the tourist inflow for a destination. The time series is decomposed into three components – trend, seasonal, and random – in order to have a deeper understating of the behavior of the time series. For the purpose of tourists' records, we have considered two very popular beaches in the state of West Bengal in India – Digha and Mandarmoni beaches. In our study, we have used data of domestic tourists' inflow in these two beaches during the period January 2008 to December 2014. Decomposition of the time series has revealed several interesting features about the trend, seasonality and randomness in the data. These features are utilized to develop some robust forecasting models for predicting future tourist inflow into the two beaches.

The rest of the paper is organized as follows. Section 2 discusses the methodology applied to justify the framework of the research. In Sect. 3, we have presented the graphical representation of the time-series data and the results after decomposing the data into three distinctive components of trend, random and seasonal. The numeric data table after decomposition with the observation and analysis of the data in general is also presented in this section. Section 4 demonstrates the various forecasting techniques applied on the data set to achieve the forecasting accuracies. Section 5 specifically presents each method applied with their outcome achieved. Section 6 concludes the paper while highlighting some future scope of work.

2 Methodology

For the purpose of our study, we have used the monthly data of aggregated domestic tourist inflow to Digha and Mandarmoni beach of West Bengal for the period of January 2008 to December 2014. The daily data of tourist visit to both the destinations are aggregated to monthly summary figures that resulted into 84 values of the time series data, one record for each month of the seven years. The aggregated monthly data is then converted into an R time series using the *ts()* function defined in the *TTR* library of the R programming tool [7]. In order to analyze the time series data more closely, the time series is decomposed into (i) trend, (ii) seasonal and (iii) random components,

using the *decompose()* function defined in the *TTR* library in the R statistical tool. After the decomposition, the three components of the tourist inflow time series data are studied in detail for deeper understanding of the visiting pattern of tourist. Five robust forecasting techniques like *ARIMA* and *HoltWinter*'s with forecast horizon of 1 month and 12 months, and aggregated trend and seasonality forecasting are applied on this time series data. We have also critically analyzed the accuracy of each of those forecasting methods.

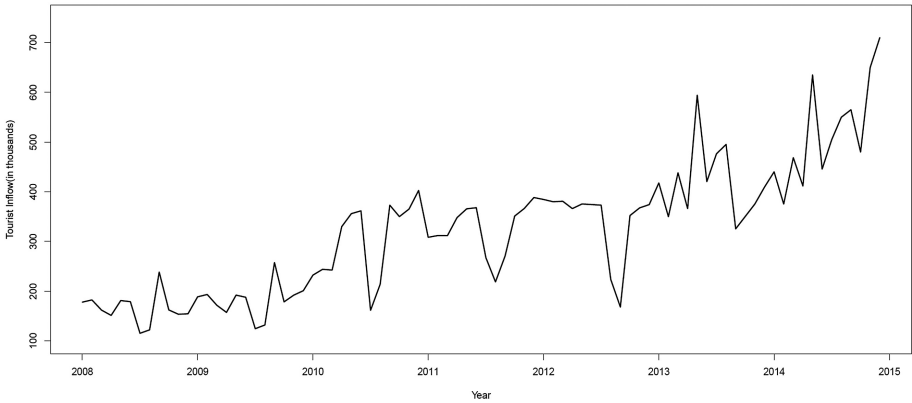


Fig. 1. Domestic tourist inflow to Digha and Mandarmoni beach (Period: Jan 2008–Dec 2014)

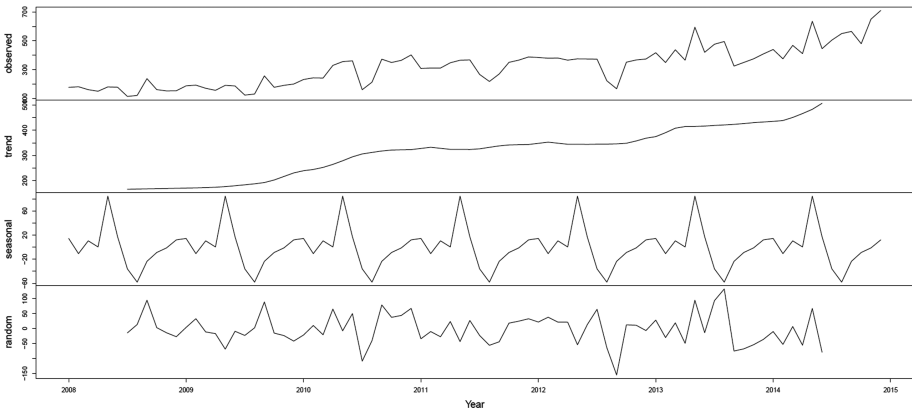


Fig. 2. Decomposition of tourist inflow time series into its three components

Extensive results are presented to demonstrate the significance and effectiveness of the decomposition approaches of time series. We have also explained the efficiency of different forecasting techniques in comparison to other approaches while the time series has a significant presence of seasonal and random components.

3 The Results of Decomposition

The outcome of the decomposition technique is presented this section. The decomposition is done on the time series of aggregated domestic tourist inflow to Digha and Mandarmoni beach of West Bengal. Figure 1 shows the observed time series of aggregated tourist inflow of Digha and Mandarmoni beaches of West Bengal for the period Jan 2008 to Dec 2014. It may be seen that the time series has an increasing slope which is accompanied by a strong seasonal behavior exhibited by regular stiff up and downturns. Figure 2 shows the decomposition outcome of the time series of Fig. 1. Here the three components of the time series – trend, seasonal, and random - are shown separately, so that their relative pattern can be easily visualized. Table 1 presents the numerical values of the aggregate time series and its three components. It is interesting to note that the trend and random components for the period January 2008–June 2008 and July 2014–December 2014 are not available [10].

Table 1. Time series components of tourist inflow data (Period: Jan 2008–Dec 2014)

Year	Month	Aggregate	Trend	Seasonal	Random
2008	January	178082		14369	
	February	182524		-10794	
	March	162058		10432	
	April	151400		264	
	May	181290		84448	
	June	178983		17102	
	July	115390	165550	-36075	-14085
	August	122270	166452	-57998	13816
	September	238470	167314	-23571	94727
	October	162433	167971	-8902	3364
	November	153680	168678	-1278	-13719
	December	154685	169505	12003	-26824
2009	January	188767	170263	14369	4136
	February	193475	171055	-10794	33214
	March	171780	172257	10432	-10909
	April	157456	173729	264	-16537
	May	192200	176007	84448	-68255
	June	187930	179541	17102	-8713
	July	124621	183294	-36075	-22598
	August	132052	187219	-57998	2831
	September	257548	192276	-23571	88843
	October	178676	202415	-8902	-14837
	November	192100	216429	-1278	-23050
	December	201090	230496	12003	-41409

(continued)

Table 1. (continued)

Year	Month	Aggregate	Trend	Seasonal	Random
2010	January	232430	239282	14369	-21221
	February	244010	244245	-10794	10559
	March	242616	252474	10432	-20290
	April	329950	264430	264	65256
	May	356044	278788	84448	-7192
	June	361689	294399	17102	50188
	July	161742	305964	-36075	-108147
	August	214034	311966	-57998	-39935
	October	350112	321320	-8902	37693
	November	365254	322480	-1278	44052
	December	402598	323149	12003	67446
	2011	January	308490	327818	14369
February		311998	332423	-10794	-9631
March		311791	328358	10432	-26999
April		348105	324136	264	23705
May		365732	324230	84448	-42946
June		368040	323689	17102	27249
July		267446	326268	-36075	-22747
August		218857	332270	-57998	-55415
September		270670	337985	-23571	-43744
October		351181	341624	-8902	18459
November		366452	342784	-1278	24947
December		388424	343452	12003	32968
2012	January	384550	348121	14369	22060
	February	379985	352726	-10794	38053
	March	380965	348661	10432	21872
	April	366260	344439	264	21557
	May	375420	344533	84448	-53561
	June	374390	343993	17102	13295
	July	373150	344781	-36075	64444
	August	223680	344911	-57998	-63233
	September	168276	346042	-23571	-154195
	October	352250	348423	-8902	12729
	November	367650	357539	-1278	11390
	December	374250	368577	12003	-6331
2013	January	417650	374801	14369	28480
	February	350000	390419	-10794	-29625
	March	438100	408286	10432	19382
	April	366260	414764	264	-48768
	May	594200	415018	84448	94734

(continued)

Table 1. (continued)

Year	Month	Aggregate	Trend	Seasonal	Random
2014	June	420530		17102	-13378
	July	476390	419198	-36075	93268
	August	495265	421191	-57998	132072
	September	325500	423520	-23571	-74449
	October	350500	426674	-8902	-67273
	November	375500	430261	-1278	-53483
	December	409300	433012	12003	-35716
	January	440000	435256	14369	-9624
	February	375500	438729	-10794	-52435
	March	468500	450988	10432	7080
	April	411550	466363	264	-55077
	May	635000	483197	84448	67355
2014	June	445760	507163	17102	-78505
	July	505000		-36075	
	August	550000		-57998	
	September	565000		-23571	
	October	480000		-8902	
	November	650000		-1278	
	December	710000		12003	

Observations: From Table 1, we can make the following observations: (i) the time series is highly seasonal in nature. A strong positive seasonality is observed during the months of May and June, while weaker positive seasonality is exhibited during the months of December, January and March. However, the months of July to November experience negative seasonality with the month of August depicting the maximum negative seasonality. Even the month of February is observed to carry a mild negative seasonal component. (ii) The seasonality of the tourist arrival pattern in the two beaches can be attributed to the seasons in a calendar year. The summer and the winter seasons seem to attract more tourists, while the monsoon experiences a lull.

4 Forecasting Techniques Applied on Tourist Data

In this section, we present a robust forecasting framework for estimating the future tourist inflow the two beaches of the state of West Bengal. Motivated from the work in [8–11], we present five different approaches for forecasting. We will critically analyze each approach for its merits and demerits.

4.1 Forecasting Method I

In this method, the time series data of the aggregated domestic tourist inflow from January 2008–December 2013 is used for forecasting the monthly inflow of domestic

tourist to Digha and Mandarmoni beaches for the year 2014. The forecasting is done at the end of the year 2013 (i.e., December 2013) and the forecast is made for each of the months of 2014, resulting in a forecast horizon of 12 months. Since we also have the actual data for the year 2014, we compute the percentage of deviation (i.e., error) of the forecasted value from the actual value for each month of 2014. Forecasting is done using the *HoltWinters* function in R with prediction horizon of 12 months [12].

4.2 Forecasting Method II

In this approach, for building the predictive model, we have considered time series data from January 2008 till the month previous to the month of forecast in 2014. For example, for forecasting tourist inflow for the month of March 2014, we build the model using data from January 2008 till February 2014, and apply the model to predict the tourist inflow for the month of March. Since the forecast horizon is 1 month, we need to rebuild the model by including the actual value of the last month in the model, every time before we make the prediction for a month. Forecasting is done using the *HoltWinters* function in R with a prediction horizon of 1 month [12].

4.3 Forecasting Method III

The fundamental objective of this method is to construct a forecasting framework which is based on the trend and seasonal components of a time series, without any consideration to the random component values. The premise is based on the fact that the random components are stochastic in nature, and they can never be predicted. In this method, the time series data of domestic tourist inflow from January 2008–December 2014 is decomposed in order to find its trend and seasonal components. As mentioned in Sect. 2, the trend component for the period July 2008 to June 2014 can only be derived using this method. In the same line, we use the time series data from January 2008 to December 2013 and compute the trend values for the period July 2008 till June 2013. The seasonal values, however, will be available for all the months during the period. Now, based on the trend values till June 2013, the trend values for the period July 2013 to June 2014 are computed with forecast horizon of 12 months using *HoltWinters* function in R [12]. The forecasted trend values for the period January 2014 to June 2014 are added to the seasonal component values of the corresponding months (based on the time series data from January 2008 till December 2013) to get the forecasted aggregate of the trend and seasonal components. Finally to have an idea about forecasting accuracy, the percentage of deviation of the actual aggregate of trend and seasonal component values with their forecasted values are computed for each month for January 2013 to June 2014.

4.4 Forecasting Method IV

In this method, *Auto Regressive Integrated Moving Average* (ARIMA) is used as a model of forecasting in this method [12]. The time series data of aggregated domestic tourist inflow of Digha and Mandarmoni beach for the period January 2008–June 2013 is used to build the ARIMA model. Based on the training data set, the three parameters

of the *Auto Regressive Moving Average* (ARMA) model is computed with *auto regression* parameter (p), the *difference* parameter (d), and the *moving average* parameter (q). Using the values of the three parameters of the ARIMA model, forecasting of the time series values is done for each month of the year 2014. The prediction horizon for the ARIMA model for this method is 12 months. For the purpose of using ARIMA, we use the in-built functions in R language for estimating the values of the parameters. The function *auto.arima* () defined in the forecast package in the R environment is used for estimating the ARIMA parameters [12]. However, we cross-verify the parameter values by computing the *auto correlation functions* (ACFs) and the *partial auto correlation functions* (PACFs) as well [12].

4.5 Forecasting Method V

In this method, we use ARIMA model with a forecast horizon of 1 month [12]. Hence for the purpose of prediction, the training data set of the time series data for the period January 2008 till the last month is considered and fitted in the ARIMA model. For example, to forecast the monthly domestic tourist inflow for the month of June 2014, time series data from January 2008 till May 2014 is used for building the ARIMA model. The percentage error in forecast is then computed using the deviation between the forecasted result and the actual value of the time series for each month. Since, in this method, the training dataset constantly changes for the ARIMA model, the parameters are evaluated every time before a forecasting is made for a month. As in Method IV, the parameters in the model are estimated using the *auto.arima* () function and they are also verified by computing the ACFs and PACFs of the series.

5 Forecasted Results

We applied all the five forecasting methods discussed in Sect. 4, on the tourist data of Digha and Mandarmoni beach so as to analyze the forecasting accuracy of the techniques. In this Section, we present the results.

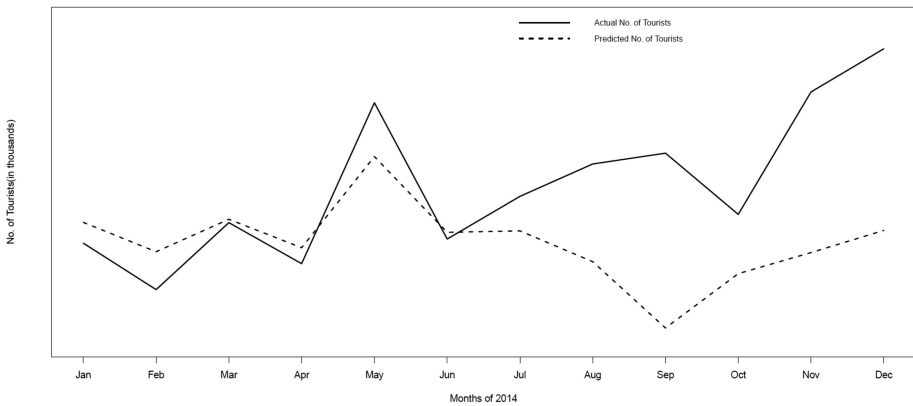
5.1 Results of Forecasting Using Method I

As discussed in Sect. 4, we apply Method I of forecasting for predicting the monthly tourist inflow into Digha and Mandarmoni beach for each month of the year 2014. *HoltWinters* function defined in R is used with varying trend and an additive seasonal component for designing a robust predictive model. The forecast horizon is chosen to be 12 to obtain the forecasted values for all months of the year 2014. The result obtained using this technique is presented in Table 2. Figure 3 depicts the actual and the forecasted number of tourists for each month of 2014 using this method.

It is observed from Table 2, *HoltWinters* forecast with time horizon 12 produces relatively high magnitude of error in forecasting. It is also noted that the percentage of error attained its highest value of 42% in the month of September before falling in October 2014 and increasing again in the months of November and December.

Table 2. Results of forecasting using Method I

Year	Month	Actual tourist flow	Forecasted tourist flow	Forecast error (%)
	A	B	C	$ C - B /B * 100$
2014	January	440000	468855	6.56
	February	375500	428114	14.01
	March	468500	473183	1.00
	April	411550	433670	5.37
	May	635000	560306	11.76
	June	445760	455079	2.09
	July	505000	457063	9.49
	August	550000	414300	24.67
	September	565000	322184	42.98
	October	480000	397669	17.15
	November	650000	426950	34.32
	December	710000	457706	35.53

**Fig. 3.** Actual vs. Forecasted Tourist Inflow using Method I (Period: Jan 2008–Dec 2014)

The relatively large error is mainly attributed to the longer forecast horizon and also due to the presence of a strong random component in the time series.

5.2 Results of Forecasting Using Method II

In this approach, as discussed in Sect. 4, prediction is made for each month of 2014 using the *HoltWinters* function in R with a forecast horizon of 1 month. Forecast model is built every time by including data from January 2008 till the month previous to the month for which forecasting is being made. An additive model with a changing trend and additive seasonal component is used for the *HoltWinters* function [12]. Since the prediction horizon is smaller, this model can capture a change in trend and seasonal components more efficiently than Method I. However, a significant and abrupt change

in the random component may adversely affect the performance of the model. The result of forecasting for this method is presented in Table 3. Figure 4 depicts the actual and forecasted values of tourists' inflow using this method.

Table 3. Results of forecasting using Method II

Year	Month	Actual tourist flow	Forecasted tourist flow	Forecast error (%)
	A	B	C	$ C - B /B * 100$
2014	January	440000	468855	6.56
	February	375500	421699	12.30
	March	468500	449948	3.96
	April	411550	417766	1.51
	May	635000	538590	15.18
	June	445760	451019	1.18
	July	505000	471556	6.62
	August	550000	445875	18.93
	September	565000	322734	42.88
	October	480000	512551	6.78
	November	650000	515118	20.75
	December	710000	587577	17.24

From Table 3 and Fig. 4, we observe that the forecasted values are, in general, quite close to the actual values, except for the month of September 2014. In the month of September 2014, the time series experienced the presence of a very strong random component, which the model could not properly catch, resulting in a large percentage of error. It is also noted that the error has not exceeded the threshold of 20% except for the month of September 2014, indicating a very robust and effective forecasting framework being used.

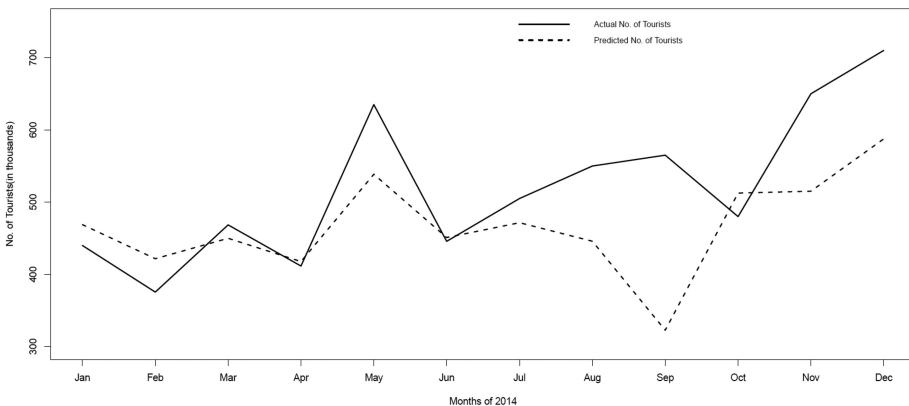


Fig. 4. Actual vs. Forecasted Tourist Inflow using Method II (Period: Jan 2008–Dec 2014).

Table 4. Results of forecasting using Method III (Period: Jan 2014–Jun 2014)

Month	Actual Trend (AT)	Actual Seasonal (AS)	Actual (AT + AS)	Forecast Trend (FT)	Past Seasonal (PS)	Forecast (FT + PS)	Error (%)
	B	C	D	E	F	G	
Jan	435256	14369	449625	429318	14180	443498	1.4
Feb	438729	-10794	427935	431106	-2420	428686	0.2
Mar	450988	10432	461420	432893	6903	439796	4.7
Apr	466363	264	466627	434681	9166	443847	4.9
May	483197	84448	567645	436468	68864	505332	11.0
Jun	507163	17102	524265	438256	30690	468946	10.6

5.3 Results of Forecasting Using Method III

As discussed in Sect. 4, this method of prediction is based on forecasting of the aggregate of trend and seasonality components. The time series data of domestic tourist inflow from January 2008–December 2014 is decomposed into its trend, seasonal and random components. Since it is not possible to determine the actual values of the trend component for the period July 2014–December 2014, we concentrate only on the period January 2014 to June 2014 for the purpose of forecasting. In Columns *B*, *C*, *D* of Table 4, the actual trend component, the actual seasonal component, and their aggregated monthly values are recorded respectively. Now, considering the trend values derived by using the time series data for the period January 2008–December 2013, the trend values for the period January 2014–June 2014 is forecasted using *HoltWinters* function in R with a changing trend and an additive seasonal component. The forecasted trend values, the past seasonal values, and their corresponding aggregate values are recorded in column *E*, *F* and *G* respectively at Table 4. The percentage error values are also computed. Figure 5 depicts the actual and forecasted values for the period Jan 2014–Jun 2014 using Method III.

The results of forecasting using Method III is presented in Table 4. Figure 5 exhibits the actual and the forecasted values for the aggregate of trend and seasonal components for this method. It can be easily observed from Table 4 that the method has yielded very low percentages of error. The results indicate that the time series has behavioral characteristics that are primarily guided by its trend and seasonal components. Hence the forecasted trend values and past seasonal components represent a good indicator for understanding and predicting the future trend and seasonal behavior of the time series. It is also evident that the seasonality pattern of the time series did not change substantially over period Jan 2008–Dec 2014 while the trend had been the most dominant component in the time series. This is surely a very positive indication for the time series to be amenable to various computationally sophisticated forecasting techniques for achieving higher level of accuracy in forecasts.

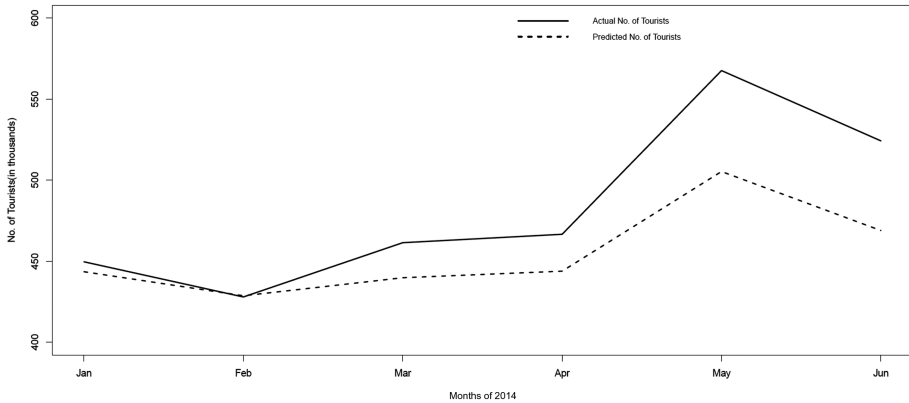


Fig. 5. Actual vs. Forecasted Tourist Inflow using Method III (Period: Jan 2008–Dec 2014)

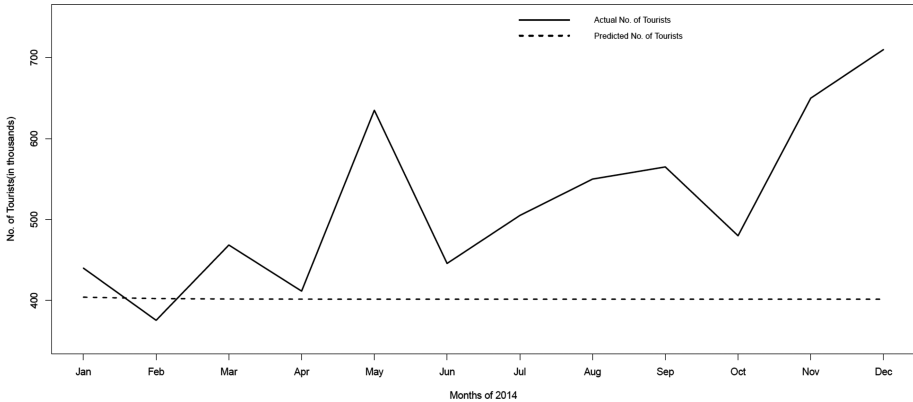
5.4 Results of Forecasting Method IV

This method is based on an ARIMA model with a prediction horizon of 12 months. As in case of Method I, we build the model using time series data for the period Jan 2008–Dec 2013, and use the model to predict the time series value of each month of the year 2014. The function *auto.arima()* defined in the *forecast* package of R is used for estimating the three parameters of the ARIMA model [12]. The obtained values of the parameters for the data are: (i) the *auto regression* parameter (p) = 1, (ii) the *difference* parameter (d) = 1, and (iii) the *moving Average* parameter (q) = 1. Therefore the model ARMA (1, 1, 1) is built for forecast. Using this ARMA model, we call the function *forecast.Arima()* in R with horizon 12 for the purpose of predicting the monthly tourist inflow for the year 2014. The forecasted values are then compared with the actual values and percentage of error for each month is computed. The results of forecasting using Method IV are presented in Table 5. The actual and the forecasted values for each month using Method IV are also plotted in Fig. 6.

As observed from Table 5, Method IV has produced quite high error percentages. This has happened primarily because of the long prediction horizon of 12 months and the time series has exhibited appreciable variations during 2014 because of the presence of a dominant random component. ARIMA applies a moving average technique by smoothening the short-term fluctuations in order to achieve a long-term prediction accuracy. This may sometimes lead to large forecast error in short-term if the time series has a dominant random component, which is the case here. It is also evident that the forecast error increased with time reaching a value of as large as 43% for the month of December 2014. The results also validate the hypothesis that accuracy in ARMA depends on the length of the prediction horizon.

Table 5. Results of forecasting using Method IV

Year	Month	Actual tourist flow	Forecasted tourist flow	Forecast error (%)
	A	B	C	$ C - B /B * 100$
2014	January	440000	404067	8.17
	February	375500	402381	7.16
	March	468500	401838	14.22
	April	411550	401663	2.40
	May	635000	401607	36.75
	June	445760	401588	9.91
	July	505000	401583	20.48
	August	550000	401581	26.99
	September	565000	401580	28.92
	October	480000	401580	16.33
	November	650000	401580	38.22
	December	710000	401580	43.44

**Fig. 6.** Actual vs. Forecasted Tourist Inflow using Method IV (Period: Jan 2008–Dec 2014)

5.5 Results of Forecasting Method V

In Method V, we use ARIMA model with forecast horizon of 1 month for computing the forecasted values of the tourist inflow for each month of 2014. The methodology followed for determining the ARMA model is same as Method IV with the difference that the parameters of ARIMA are evaluated before each time a forecast is being made. The evaluation of the parameters before each forecasting is necessary since the training data set changes constantly due to inclusion of the time series value of the last month. Table 6 presents the results and Fig. 7 depicts the actual and forecasted values for each month of 2014 using Method V.

It is evident from Table 6 that the accuracy of ARIMA model with prediction horizon 1 used in Method V is better than that of prediction horizon 12 used in Method IV. However, due to the presence of strong random components, the error is still significant, with the month of May yielding a high error rate of 35%.

Table 6. Results of forecasting using Method V

Year	Month	Actual tourist flow	Forecasted tourist flow	Forecast error (%)
	A	B	C	$ C - B /B * 100$
2014	January	440000	404067	8.17
	February	375500	420985	12.11
	March	468500	391477	16.44
	April	411550	435423	5.80
	May	635000	413340	34.91
	June	445760	505564	13.42
	July	505000	478071	5.33
	August	550000	488261	11.22
	September	565000	512190	9.35
	October	480000	533759	11.20
	November	650000	512443	21.16
	December	710000	565123	20.41

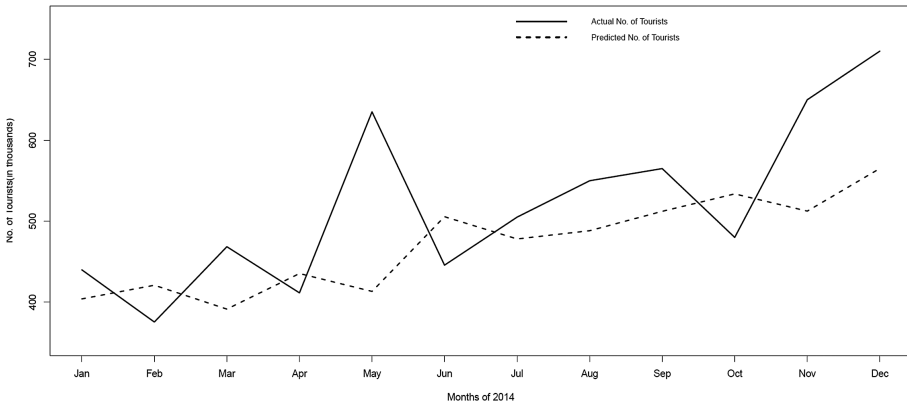


Fig. 7. Actual vs. Forecasted Tourist Inflow using Method V (Period: Jan 2008–Dec 2014)

Table 7. A comparative analysis of the five forecasting methods

Method	Max error	Min error	Mean error	SD of error	RMSE
Method I	42	1	17	14	132.17
Method II	43	1	13	11	98.98
Method III	11	0	4	4	36.44
Method IV	43	2	21	14	154.22
Method V	34	5	14	8	96.91

Table 7 presents the summary of the results of the five methods of forecasting. We have computed five metrics for evaluating the performance of the forecasting approaches: (i) maximum percentage of error, (ii) minimum percentage of error, (ii) mean percentage of error, (iv) standard deviation of error percentages, (v) root mean square error (RMSE).

While a robust forecasting framework should minimize all these parameters, RMSE is usually considered to be the single metric for comparing different predictive models. It may be observed from Table 7 that Method III has produced best results in all parameters and hence it may be considered as the best-fit model for the tourist time series. ARIMA with prediction horizon of 12 months (i.e., Method IV) was found to yield the worst results with an RMSE value of 154.22.

The reason why Method III has produced the best results is not difficult to understand. Actually, the tourist time series has a strong trend component and an almost invariant seasonal component. This enabled us to very accurately predict future seasonality based on the past seasonal behavior. Since we could forecast the trend values quite accurately using *HoltWinters* approach, Method III enabled us to very accurately predict the aggregate of the trend and seasonal component values.

For the same function, prediction horizon of 1 month has produced better forecast accuracy than prediction horizon of 12 months. *HoltWinters* with prediction horizon of 12 months (Method I) and prediction horizon of 1 month (Method II) have produced RMSE values of 132.77 and 98.98 respectively. Similarly, ARIMA with prediction horizon of 12 months (Method IV) and prediction horizon of 1 month (Method V) have produced RMSE values of 154.22 and 96.91 respectively. Smaller prediction horizon has produced increased accuracy. The time series has a dominant random component making forecasting over a longer horizon a challenging task.

6 Conclusion and Future Work

In this paper, we analyzed the domestic tourist inflow time series for the beaches of Digha and Mandarmoni in the state of West Bengal in India, for the period of Jan 2008 to Dec 2014. The decomposition of the time series data into trend, seasonal, and random components provided us with a deeper and clearer insight into the visiting behavior and the seasonal preferences of the tourists about the two beaches. Based on the output of decomposition results, we identified the months of highest and lowest seasonality for tourist inflow in the two beaches. We have also built a robust forecasting framework consisting of five different models for predicting tourist inflow in the two beaches. Extensive results have been provided to demonstrate the effectiveness and accuracy of the forecasting models.

The results of this forecasting would be beneficial for the *destination managing organizations* (DMOs) in planning for the requirements and investments needed for development of beach tourism, and it would also help them in formulating effective marketing plans for the lean seasons with very low tourist inflow.

References

1. Kon, S.C., Turner, L.W.: Neural network forecasting of tourism demand. *Tourism Econ.* **11** (3), 301–328 (2005)
2. Cuhadar, M.: Modelling and forecasting inbound tourism demand to Istanbul: a comparative analysis. *Eur. J. Bus. Soc. Sci.* **2**(2), 101–119 (2014)

3. Akuno, A.O., Otieno, M.O., Mwangi, C.W., Bichanga, L.A.: Statistical models for forecasting tourists' arrival in Kenya. *Open J. Stat.* **5**, 60–65 (2015)
4. Min, W.: A research on tourism market forecast and evaluation based on the application of linear regression model. *Int. J. u- e-Serv. Sci. Technol.* **7**(6), 213–226 (2014)
5. Kusni, A., Kadir, N., Nayan, S.: International tourism demand in Malaysia by tourists from OECD countries: a panel data econometric analysis. *Procedia Econ. Financ.* **7**, 28–24 (2013)
6. Frechtling, D.C.: *Tourism Demand: Methods and Strategies*. Butterworth-Heinemann, Oxford (2001)
7. Ihaka, R., Gentleman, R.: A language for data analysis and graphics. *J. Comput. Graph. Stat.* **5**(3), 299–314 (1996)
8. Sen, J., Datta Chaudhuri, T.: A framework for predictive analysis of stock market indices – a study of the Indian auto sector. *J. Manag. Practices* **2**(2), 1–20 (2016). Calcutta Business School (CBS)
9. Sen, J., Datta Chaudhuri, T.: Decomposition of time series data of stock markets and its implications for prediction – an application for the Indian auto sector. In: *The 2nd National Conference on Advances in Business Research and Management Practices (ABRMP)*, Kolkata, India, 8–9 January 2016, pp. 15–28 (2016)
10. Sen, J., Datta Chaudhuri, T.: An alternative framework for time series decomposition and forecasting and its relevance for portfolio choice: a comparative study of the Indian consumer durable and small cap sectors. *J. Econ. Libr.* **3**(2), 303–326 (2016)
11. Sen, J., Datta Chaudhuri, T.: An investigation of the structural characteristics of the Indian IT sector and the capital goods sector – an application of the R programming in time series decomposition and forecasting. *J. Insur. Financ. Manag.* **1**(4), 68–132 (2016)
12. Coghlan, A.: *A Little Book of R for Time Series*. Release 02 (2015). <https://media.readthedocs.org/pdf/a-little-book-of-r-for-time-series/latest/a-little-book-of-r-time-series.pdf>