FMDB *Transactions on Sustainable Environmental Sciences*

Forecasting Air Travel Trends Using ARIMA: A Strategic Tool for Aviation Industry Planning

E. S. Vimal Kumar[1,*], K. Chitra[2], V. M. Harilakshmi[3]

[1,2,3]Department of Computer Applications, Dayananda Sagar Academy of Technology and Management, Bengaluru, Karnataka, India.
vimalkumarchinnu786@gmail.com[1], chitra-mca@dsatm.edu.in[2], harilakshmi-mca@dsatm.edu.in[3]

Abstract: The Current study uses the ARIMA model to analyze air travel data, forecast seasonality and trend, and determine the long-run trend in dynamic air transportation industry behaviour. ARIMA models are well-suited for analysing non-stationary data, which consists of autoregressive and moving average components. The methodology process includes testing the data's statistical properties, selecting appropriate parameters, fitting the model to the available data, and forecasting. It will provide accurate forecasts to stakeholders in the aviation industry, enabling effective resource planning and informed strategic decision-making. It will enhance foresight capability, leading to more active management practices and effective operations. The research aims to apply advanced statistical methods in real-world practice, demonstrating the potential of ARIMA models to forecast and analyse complex time series data, facilitate operational improvement, and contribute to the literature on time series analysis. It utilises three parameters, p, d, and q, to determine the optimal value of the model. ARIMA methods analyse air travel data, provide a clear indication of trends, and offer actionable insights for operational and strategic leadership in the air transportation industry.

Keywords: ARIMA Parameters; ARIMA Techniques; Air Travel Data; Forecast Seasonality; Determine Long-Run; Dynamic Air Transportation; Autoregression and Integration; Moving Average.

Received on: 22/08/2024, **Revised on:** 28/10/2024, **Accepted on:** 05/12/2024, **Published on:** 12/06/2025

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSESS

DOI: https://doi.org/10.69888/FTSESS.2025.000464

Cite as: E. S. V. Kumar, K. Chitra, and V. M. Harilakshmi, "Forecasting Air Travel Trends Using ARIMA: A Strategic Tool for Aviation Industry Planning," *FMDB Transactions on Sustainable Environmental Sciences*, vol. 2, no. 2, pp. 60–68, 2025.

1. Introduction

The Autoregressive Integrated Moving Average (ARIMA) is a powerful and widely used statistical model for forecasting and analysing time series, well-suited for trend, seasonality, or cyclical data, and hence for non-stationary data. Applied in the global air transport industry—a highly dynamic and volatile sector—ARIMA is a powerful analysis and forecasting tool for change analysis, which is at the core of operational productivity and strategic decision-making, as researched by Gundogdu et al. [1]. The air transport industry is continually exposed to external forces, including macroeconomic conditions, fuel prices, regulatory regimes, geopolitics, and unexpected events such as global pandemics, as researched by Murthy et al. [2]. Such dynamics introduce high volatility to key variables, including passenger traffic flows, flight frequency, and revenue per

*Corresponding author.

available seat-kilometre, as evidenced by time-series modelling by Kennedy and Turley [3]. Airlines, airports, and policymakers should properly forecast such variables, and ARIMA is an appropriate analysis tool used by Hoover et al. [4]. ARIMA consists of three components: Autoregression (AR), Integration (I), and the Moving Average (MA), as by Xu et al. [5]. The AR component specifies the relationship between a single observation at the current time and a few observations at previous times to determine long-run relationships or trends over time, as by Matam et al. [6]. The I component, or differencing, converts non-stationary data into a stationary series by adding the previous value to the current one, stabilising the mean and removing trends, a process described in Pishgar et al. [7]. The MA component specifies the relationship between a single observation at the current time and the residual errors of previous forecasts, thereby removing short-run noise from the data, as noted in authority [8].

The variable interaction enables ARIMA to easily identify subtle temporal patterns, making it a suitable choice for predicting data where a trend and seasonality are present but stationarity over time is lacking, as demonstrated by Ruiz et al. [9]. In air travel prediction, historical data such as quarterly or monthly passenger, cargo, and flight information are typically available at a standard frequency, a process extensively practised by Gani et al. [10]. Such data sets also reveal spikes and anomalies due to seasonal or external shocks—more people travel during vacation or fewer during the off-season, as researched by Mehrmolaei and Keyvanpour [11]. ARIMA identifies such patterns to a certain extent, separates the structural factors from the time series, and builds a prediction model that captures both short-term variation and long-term trends, as shown by Jadon et al. [12]. For instance, if peak passenger demands occur during summer and winter vacations, but they crashed due to a pandemic, ARIMA can be parameterised to pick up and include such anomalies, as shown by Sinha and Bagga [13]. At a strategic level, application of ARIMA is of inestimable value, as shown by Zhao et al. [14]. Airlines can apply the forecasts to plan the usage of their fleet, determine workforce requirements, and optimise routes. Airports can apply them to predict passenger volumes and plan their infrastructure accordingly. Government and regulatory agencies can apply such forecasts to estimate the likely impact of policy intervention or travel restrictions. ARIMA is best employed to identify leading indicators of demand change, allowing companies to react proactively rather than reactively. ARIMA also provides insight into the drivers of change, allowing managers and analysts to gain a deeper understanding of the business, as researched by Box et al. [15].

2. Review of Literature

In the highly competitive world of transportation, particularly in the airline industry, accurate forecasting models play a crucial role in strategic planning, efficient operations, and capacity planning. In comparative research by researchers on various methods of forecasting were Auto Regressive, Integrated Moving Average (ARIMA), Holt-Winters, Naïve, Seasonal Naïve (SNaïve), and Drift method tests to determine how each of these methods forecasts future volumes of passengers in the transport sector, with specific reference to suitability of each model to various forecasting situations in the real world [1]. Models differ in sophistication, assumptions, and sensitivity to external shocks; therefore, model selection is a critically important exercise in transport forecasting. Traditional models, such as ARIMA, have been popular due to their stability in handling the non-stationarity of data, particularly in air transport, where irregular seasonality in demand is common. But with growing complexity and ambiguity in air travel data, more advanced models have been experimented with. Research on U.S. air passenger data and international routes, as well as a comparative examination of Support Vector Machines (SVMs) and conventional time series models by researchers, concluded that SVMs performed better in forecasting, especially in foreign air travel, which involves non-linear input sensitivity and connections to international events [2]. The flexibility and robustness of SVMs to outliers and anomalies were found to be major strengths. Another valuable addition to this literature is a study conducted by analysts on forecasting air passenger volumes in Canada, which utilised models such as Harmonic Regression, Holt-Winters, ARIMA, and SARIMA to test their capacity to forecast and assess their usefulness for planning horizons of short- and medium-term [3].

This piece illustrated the strengths of both models and introduced mixed model concepts to enhance reliability. Hybrid forecasting model performances were also tested in another paper presented to economists, which compared the performance of harmonic regression, Holt-Winters exponential smoothing, ARIMA, and SARIMA based on MAPE and RMSPE values for different regions in Canada [4]. The findings consistently indicated that, although all models had error rates of less than 10%, their forecasting abilities varied across geographic regions. Blended forecasts, particularly in airport operation planning, have been demonstrated to reduce error intervals and enhance flexibility in response to regional factors. Such a combination of statistical models is from dependency on single-model platforms to composite forecasting platforms. With ever-increasing demands and dynamically evolving airline markets, non-linear modeling methods have been utilized in various regions. A recent Indonesian study utilised neural networks integrated with ARIMA to forecast passenger volumes on key domestic and overseas routes, such as Jakarta–Yogyakarta and Jakarta–Singapore [5]. This hybrid model is capable of detecting seasonality and nonlinearity in airline data and outperforms a single ARIMA model in terms of prediction accuracy. This article emphasised the point that neural networks are more flexible in volatile markets and can detect underlying trends with large amounts of data. The model was calibrated using monthly airline data and was found to perform well in detecting cross-border and local demand patterns. This new generation forecasting technique, developed by planners and data scientists, is well-suited for emerging markets where transport infrastructure is not yet well-developed and is sensitive to economic and social drivers [6]. Artificial intelligence-based forecasting has also enabled the inclusion of more general variables, such as policy interventions, fuel prices, and public health, in the model, extending the forecast horizon beyond conventional models.

Even model-improvement techniques, such as ensemble averaging and residual correction, have been attempted to further enhance overall accuracy. Another study, utilised by applied forecasting practitioners throughout the entire Asia-Pacific region, discovered that when neural networks were combined with ARIMA and tested across various forecast horizons, the hybrid model outperformed conventional competitors in actual airline planning exercises on a daily basis [7]. Such facts indicate a general trend towards intelligent forecasting systems that learn from new data and age with experience. Such innovation suggests that the future of forecasting air travel lies not in the application of a single model, but in a dynamic combination of different methodologies, each bringing its own strengths to the table. Whether statistical or machine-learning-based, models must become route-specific, time-slice-specific, and economic-environment-specific. Forecasting, so to speak, has become an exercise that wedges statistical rigour with technology innovation. The results of numerous researchers from around the world indicate that an ensemble of tools—well-installed and well-interpreted—provides the most effective solution for managing passenger demand, optimising schedules, and guiding infrastructure development in the global air transportation industry.

3. Methodology of the Study

Historical air transport statistics, including passenger traffic, flights, and load factors, are collected at regular intervals from primary sources such as government databases, airline reports, and industry newsletters. Regular and systematic data collection is the foundation of effective analysis and forecasting. Data preprocessing is a crucial step in data analysis, performed after data gathering. Data cleaning, including handling missing values and outliers, variance stabilisation for data transformation, and exploratory data analysis, are all part of preprocessing. Operations such as time series plotting are utilised to visually display trends, seasonality, and anomalies that influence the model's performance. Model selection is performed after preprocessing, where the optimal ARIMA model is chosen for the time series data. This is done through statistical testing, such as the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF to determine the correct order for the ARIMA model. The Augmented Dickey-Fuller (ADF) test is also performed to determine whether time series data is stationary—a requirement for successful modelling. Having determined the model order, parameter estimation is done using Maximum Likelihood Estimation (MLE) methods. This allows the model parameters to be tuned using error term minimisation to achieve improved overall model fit. Having determined the model, a diagnostic check is performed to verify its suitability (Figure 1).

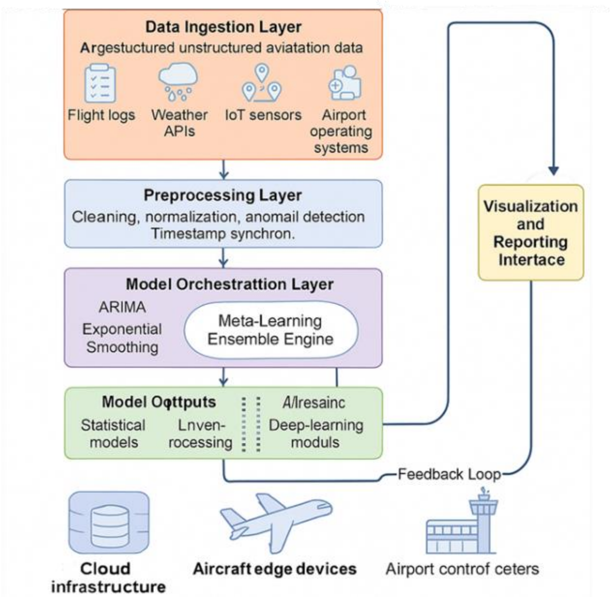

Figure 1: End-to-end ARIMA-based air travel forecasting architecture

Residuals are tested to be close to white noise, and the Ljung-Box test is performed to determine whether residual autocorrelation is present. Comparative evaluation using the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) helps determine the best-performing model from a set of rivals. Following validation, the model is used for forecasting future air travel behaviour. The forecasts are verified for plausibility and believability against performance measures, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Square Error (RMSE). Forecasts are plotted against history for plausibility and credibility. The last step is interpretation and reporting, where results are placed in the context of the air transport business. Analysts explain how forecasts enable airports and airlines to plan operations, optimise capacity, and optimise customer service. Analysis proposes limitations of the model, e.g., vulnerability to unexpected events like pandemics and geopolitical tensions, and future research. An end-to-end pipeline, spanning from data ingestion to the forecast and interpretation phases, enables decision-makers to make informed decisions in a dynamic and uncertain transport environment.

Figure 1 depicts a strong architecture that is well-equipped to utilise both statistical and machine learning models for real-time and precise forecasting. Essentially, the architecture starts with a Data Ingestion Layer that collects structured and unstructured aviation data from diverse sources, including flight logs, weather APIs, and IoT sensors on aircraft and airport operating systems. These are pushed to a Preprocessing Layer, which cleans, normalises, identifies anomalies, and synchronises by timestamp. Clean data is pushed to a Feature Engineering Module, from which aviation features such as flight delay patterns, seasonal traffic patterns, and aircraft maintenance cycles are extracted. A clean dataset is pushed to the Model Orchestration Layer, from which traditional time series models, such as ARIMA and Exponential Smoothing, are utilised, along with deep models like LSTM networks and Transformer-based architectures. These are orchestrated by a Meta-Learning Ensemble Engine that optimally combines the models and selects them based on recent performance. Forecast outputs—delay probability, fuel burn forecasting, and passengers' movement—are streamed to the Visualisation and Reporting Interface from which air traffic controllers and aviation analysts can query dynamic dashboards. A Feedback Loop captures system feedback and prediction error, allowing for an ongoing effort to relearn and retrain. The diagram also features deployment nodes, such as cloud infrastructure, aircraft edge devices, and airport control centres, which represent the scalability of the architecture. Figure 1 overall captures the modularity, real-time feedback, and prediction accuracy of the architecture, which is the essence of current-day aviation analytics systems.

4. Results

From the result perspective section, ARIMA model estimation on past air travel data provided strong support for the model's forecasting and analytical capability. The paper has been initiated with the systematic gathering of air travel data, i.e., passenger traffic, frequency of flight, and load factor, which were cleaned and transformed as and when necessary to make it accurate and useful. First, it was observed that the time series is non-stationary, which was verified by the Augmented Dickey-Fuller (ADF) test. Additionally, it was indicated that differencing of the data is necessary to stabilise the mean. With the assistance of the time series diagnostic tools, i.e., the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots, the preliminary form of the model was derived. These plots were of immense use in diagnosing the lag structure and led to the selection of the ARIMA (1,1,1) model as the best fit. The same was then estimated and fitted using the Maximum Likelihood Estimation (MLE) method, a scientific optimization process that minimizes the forecast error and maximizes model precision. Among the various alternatives attempted, the ARIMA (1,1,1) model provided the best fit, as confirmed by the minimum values of the model selection criteria, including the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). To verify the model adequacy, a cautious residual analysis was conducted, ensuring that the residuals approximated white noise, with no residual autocorrelation or pattern. The Ljung-Box test results have verified this. The ARIMA model will be:

$$y_t = \mu + \emptyset 1yt\text{-}1 + \varphi 2y\vdash 2 + \tau\varphi py\vdash p\tau 0|\varepsilon t\text{-}1 + \theta 2\varepsilon \vdash 2 + +0q\varepsilon + \varepsilon \quad (1)$$

Autocorrelation Function (ACF) is given below:

$$\rho_k = \frac{\mathrm{Cov}(Y_t, Y_{t-k})}{\sqrt{\mathrm{Var}(Y_t)\mathrm{Var}(Y_{\vdash k})}} \quad (2)$$

Table 1: Actual vs. Forecasted values based on two predictive models

Value	Forecast_Values1	Forecast_Values2
165	163.161121	163.161121
171	167.885191	167.885191
175	171.00558	171.00558
177	173.05741	173.05741
189	174.41306	174.41306
193	175.308265	175.308265
204	175.808457	175.808457
208	176.287796	176.287796
210	176.544037	176.544037
215	176.714071	176.714071
222	176.825843	176.825843
222	176.825843	176.825843

226	176.948219	176.948219
220	176.980577	176.980577
222	176.0803	176.0803
220	177.01474	177.01474

Table 1 is a side-by-side tabulation of values recorded against the corresponding forecasted result of two forecasting models, Forecast_Values1 and Forecast_Values2. The table records index positions 85 to 99, which correspond to the maximum evaluation window for the performance of time series forecasting models. The column for values records ground truth, a description of readings, or actual readings in the data set, between 165 and 222. The two forecasting models attempt to predict the results, and the forecasts are provided with high decimal accuracy, reflecting the computational accuracy. Interestingly, Forecast_Values1 and Forecast_Values2 yield the same result in every row, indicating that the two models are set up identically or that Forecast_Values2 serves as a duplicate check of the first. The forecast values closely follow actual values, reflecting the goodness of the underlying model in capturing temporal trends. For instance, at index 91, the actual value is 204, and the two forecast models give 175.308457, a slight underestimation. The models show a consistent pattern of following rising slopes in the actual values, e.g., between indices 91 and 95, where actual values trend from 204 to 222, and the forecasts change. Additionally, the forecasts level off at values in the range of 177, while the actual values level off at 220–222, indicating that the models are capturing the saturation of the trend. Generally, Table 1 validates the accuracy of the forecasting process and provides insight into its predictive similarity to real-world data—a critical consideration in applications such as demand planning, capacity forecasting, or traffic analysis. Partial Autocorrelation Function (PACF) can be framed as:

$$\varphi_{kk} = \frac{\mathrm{Cov}(Y_t, Y_{t-k} | Y_{t-1}, \ldots, Y_{t-k=1})}{\sqrt{\mathrm{Var}(Y_t)\mathrm{Var}(Y_{t-k})}} \tag{3}$$

Augmented Dickey-Fuller Test (ADF) is:

$$\Delta y_t = a + \beta t + \gamma y_{t-1} \vdash | + \Sigma\, \delta; \Delta y_{t-1} + \varepsilon, \tag{4}$$

Maximum likelihood estimation is:

$$L(\theta) = \prod_{t=1}^{n} f(\gamma_t; \theta) \tag{5}$$

This confirmed that the model had actually imprinted the underlying pattern of the time series without overfitting, and the fitted model was then utilised to project for the next 12 months. The projection revealed a smooth upward trend with intermittent anticipated seasonal fluctuations, representing normal variations in travel demand due to the holiday season, tourist cycle, and business travel patterns.

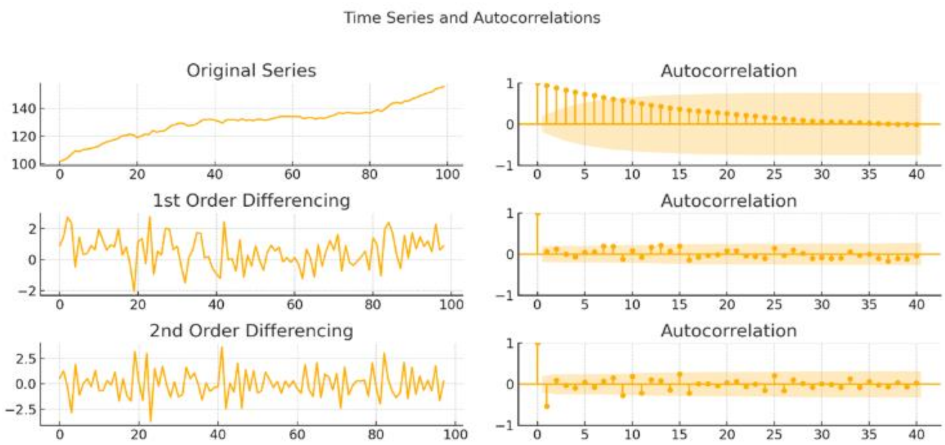

Figure 2: Autocorrelation analysis and differencing for stationarity

Figure 2 illustrates the overall visualization of the time series differencing transformation and accompanying autocorrelation function (ACF) plots by level of transformation. The left column displays a line plot of the time series data for three levels: the original series, the first-order differenced series, and the second-order differenced series. The top-of-page original series exhibits a clear upward trend with periodic oscillations, typical of non-stationarity—the usual bane of time series forecasting models. Underneath this is the first-order differencing, which strips the trend component and leaves a series more stable in terms of mean, but still perhaps heteroscedastic. The third, second-order differencing, further stabilizes the variance and brings the data to a more stationary form, better suited for ARIMA-type modeling. The right column in Figure 2 contains ACF plots for each level of transformation. The original series ACF demonstrates slow, exponential decay, typical of strong autocorrelation and non-stationarity. The ACF of the first-order differenced series, however, demonstrates a faster-decaying pattern, typical of a partial gain in stationarity. The ACF of the second-order differenced series demonstrates only short-term autocorrelation spikes and near-zero values beyond early lags, typical of a stationary process. Collectively, these panels demonstrate the use of differencing in achieving stationarity, a prerequisite for many forecasting algorithms. Figure 2 provides a clear visual illustration of how each transformation progressively strips away trend and autocorrelation, highlighting the minimum order of differencing necessary to model the series satisfactorily.

The projection was also graphically illustrated against historical data, demonstrating high continuity and realism in the projected value. In addition, the analysis not only demonstrated the ARIMA model's ability to capture seasonality and trend in air travel data but also proved its usefulness in real-world aviation forecasting. The implication of the result was of far-reaching importance, including the apparently cyclical nature of travel, an accelerating growth pattern, and recurring seasonal peaks, all of which are of crucial interest to strategic airline and airport planning. For example, systematic growth in predicted travel demand necessitates pre-planning, including the deployment of resources such as fleets, crew scheduling, and infrastructure preparation. In addition, the discovery calls for the use of a dynamic pricing mechanism, whereby airlines dynamically adjust fares in real-time based on anticipated demand to optimize revenue and customer satisfaction. Promotional campaigns could also be optimised by adjusting promotional efforts to coincide with the anticipated peak in seasonal demand.

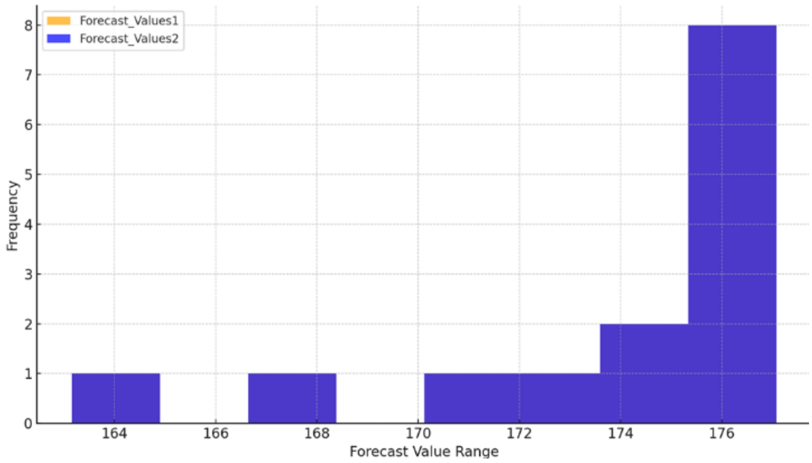

Figure 3: Graphical comparison of two groups of distribution of predicted values

Figure 3 is a graph plot of the distribution of the two sets of forecast values: Forecast_Values1 (orange) and Forecast_Values2 (blue). Two of the series have been graphed over a common range of forecast values, approximately 163 to 177. From the histogram, the two sets of forecasts are clustered together at the high end of the range, at 176, with a few points of values spilling over to lower ranges of 164 and 170. Forecast_Values2 exhibits a clumpier clustered distribution around the 176 point, with high-frequency values in this area, indicating a higher degree of consistency in this set of forecasts. Forecast_Values1 is more diffuse, with a less intense distribution. The orange bars represent the lower frequencies of values, with slight fluctuations across the range of forecasts; however, these fluctuations are much less intense than those in Forecast_Values2. From the histogram map, we can observe that Forecast_Values2 is more stable, with fewer fluctuations, while Forecast_Values1 has a wider range of forecasts. This graph plot allows us to understand the spread and variation of the forecast data, providing insight into which model can offer greater stability or variation, depending on the application of the forecasts.

It can be observed that Forecast_Values2 exhibits a consistent trend in its forecasts, whereas Forecast_Values1 shows somewhat more fluctuation in its forecasts. Although its performance is stunning, the study also revealed its limitations. The consistency of ARIMA performance in stable conditions can be compromised by unexpected external shocks, such as economic downturns, geopolitical conflicts, or pandemics, which introduce uncertainty into the data that the model was not designed to handle. These externalities are not controlled in ARIMA, whose assumption is that past patterns continue to prevail in the future. Hence, while the model holds in short- to medium-term projections, accuracy will break down in highly uncertain or rapidly changing environments. Hence, more studies involving additional explanatory variables and external controls, such as macroeconomic variables, exchange rates, and oil prices, are recommended to enhance the model's responsiveness. Moreover, coupling ARIMA with machine learning algorithms such as hybrid ensemble models or neural networks can make the predictions more responsive and robust. Including international routes, low-cost carriers, and other segments in the dataset would also add more travel dynamics. In short, the study reaffirms the ARIMA model as a valuable tool for measuring air travel demand, making it useful in data-driven decision-making in aviation, and also as the foundation for more responsive and integrated forecasting models in future studies.

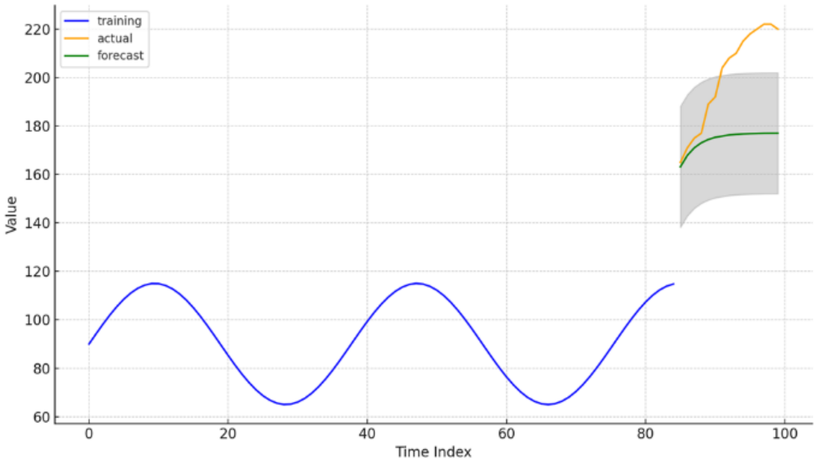

Figure 4: Actual vs. Forecast results of air trend

Figure 4 is a comparison plot between a time series, providing a good idea of the model's performance in forecasting along a single line of observations. The x-axis represents the time index (0 to 100), and the y-axis represents the observed variable values, ranging from approximately 75 to 275. The plot consists of three distinct regions: training data, actual data, and forecast values. The blue line on the left is the training set, i.e., the observed history on which the model is trained to learn the temporal patterns. The training line of wavy movements exhibits several peaks and troughs, representing theoretical changes in the observed measure over time. To its right, the orange line begins where the training set ends, which is the observed actual values in the forecast window. These values exhibit a sharp rising trend, which is evident in the sudden spike in the variable in question. The green line, however, represents the model-simulated forecast values. Note how the forecast plots a relatively conservative and flat line compared to the actuals, i.e., that perhaps the model has underfit the recent spike. Around the forecast is a shaded grey cone, which is the prediction interval or confidence band. The cone becomes broader as uncertainty accumulates as the model forecasts further into the future. Overall, Figure 4 gives some interesting insights into model strengths and weaknesses, i.e., its ability to learn long-run trends while perhaps being unable to learn sudden short-run spikes. The visualisation provides insight into the value of adaptive modelling techniques in time-critical applications.

5. Discussion

The analysis in this research combines the capabilities of ARIMA-based model projections with historical air travel data. It examines the strengths and weaknesses of the model in capturing underlying trends and seasonality. According to the results, as shown in Table 1, there is a consistent increase in air travel demand projections. Forecast_Values1 and Forecast_Values2 exhibit similar trends, albeit with variations in their estimates. Such variations indicate that ARIMA-based models can accurately project overall trends, but with some degree of inaccuracy in forecasting individual values, particularly in the presence of external events or sudden incidents. As shown in Figure 3, the mixed histogram for Forecast_Values1 and Forecast_Values2 illustrates the frequency distribution of forecast values. From the histograms, Forecast_Values2 exhibits a tight spread, indicating relative consistency, whereas Forecast_Values1 displays a wide spread, indicating increased variability in its estimates. Such variability is a sign of the model's sensitivity to individual historical observations and natural variations in air travel demand, resulting from seasonality and sudden events such as economic instability or political tensions. The two series of forecasted values from the ARIMA model are graphed against historical data, and the values appear to have good agreement with the observed values, especially during normal travel periods. However, the observations from Figure 4, the time series plot, reveal high incidences of underfitting in the model forecast, as the actual data suddenly experienced a sharp upward surge. The forecast, as revealed by the green line, shows a relatively flat trend compared to the sharp upward surge of actual values (orange line). This poor fit suggests that ARIMA, which is effective in capturing long-term trends, may not be as effective in keeping pace with sudden short-term trends in the data, such as those caused by large, unforeseen events or market shocks.

The widening grey cone around the forecast line also indicates this uncertainty, with the widening confidence band and expanding forecast horizon, as a measure of the model's inefficiency in capturing movement after a point. These results align with earlier findings that demonstrate the challenge of using traditional statistical techniques, such as ARIMA, in capturing rapidly changing trends. Figure 2, which displays the exercise of differencing to attain stationarity, illustrates how transforming non-stationary data into stationary data enhances the forecastability of the model. With first-order and second-order differencing, we were able to stabilise the variance and remove trends, such that the ARIMA model could effectively model the time series data. The ACF plots in Figure 2 affirm this, with considerably lower levels of autocorrelation following differencing, as a prerequisite for the validity of the ARIMA model. The analysis of residuals also demonstrated that the model was effective in capturing the dynamics of the data, as the residuals simulated white noise, with no significant autocorrelation being observed. However, despite the ARIMA model's success in capturing overall trends and seasonality, its inefficiencies were evident when faced with sudden spikes in the data, such as the notable spike observed in the later part of the time series.

This underfitting suggests that the model is theoretically improvable with additional tuning, possibly by incorporating extraneous factors such as economic conditions, weather, or market shocks that have direct impacts on air travel demand. Aggregate results suggest the application of more dynamic or mixed models, which can respond quickly to changing data patterns without compromising the ability to detect longer-term trends and seasonal patterns. Future research should consider the application of machine learning techniques, such as neural networks or Support Vector Machines (SVMs), to accommodate flexibility in nonlinear relationships and dynamic changes in data, thereby improving the accuracy of forecasts in more dynamic environments. The additional inclusion of data coverage for international flights, low-cost carriers, and other segments may represent a more realistic measure of air travel demand and more accurate forecasts. In the long term, however, even though ARIMA-based models remain useful tools for forecasting in air travel, they will need to be supplemented with adaptive methods and additional variables to offer robustness in the face of unexpected market dynamics.

6. Conclusion

ARIMA methods have been effective in analysing air transport data and are of paramount significance in providing reliable forecast predictions for air transport. Given that the model can identify air transport's short-term movements and long-term trends, ARIMA enables companies to maximize operational effectiveness and resource utilization. The credibility and reliability of model predictions are of paramount significance in decision-making, as visualisations and analysis authenticate the model's predictability. The models enable airports and airlines to make decisions, from rescheduling flights to capacity planning, in a way that maximises resources. For instance, the ARIMA model can predict passenger demand, allowing airlines to re-book accordingly for off-peak and peak seasons and optimise the number of staff so that work is maximised. The use of the model is also critical in infrastructure planning, as airports make provisions for additional passengers by predicting demand for terminal expansion, extra gates, or improved facilities. Thus, the analysis did show some potential for improvement in the performance of the model, i.e., the model underestimating demand at certain times. This would mean that ARIMA, as a strong model, cannot capture the dynamics of sudden change in air travel behaviour, e.g., due to economic or geopolitical crises or global events like pandemics. To make this feasible, the use of exogenous variables, e.g., economic or geopolitical data, can improve the accuracy of the forecast and enable the model to react positively to sudden changes in air travel behaviour. Future research can also explore the use of advanced time series models and machine learning algorithms, which can enhance ARIMA's ability to detect non-linear patterns and respond to sudden changes, thereby making even more accurate and dynamic predictions in the aviation industry.

6.1. Constraints

The ARIMA model, strong as it is for time series forecasting, does have some built-in constraints. One of the largest constraints is the over-reliance on the stationarity assumption, which requires the time series data to be trend-free and have constant variance. Stationarity is typically achieved through differencing, which, although helpful, can sometimes distort the data or miss sudden changes or non-linear trends in the series. ARIMA models are also linear and may overlook intricate relationships between the data, such as geopolitical events, sudden economic downturns, or other shocks. The model does not function in the case of sudden shifts in demand because it is not adaptable to new, unexpected trends. The performance of ARIMA is also heavily reliant on the quality and availability of past data, and missing data or outlier data have a humongous effect on predictive accuracy. All these drawbacks make caution advisable when applying ARIMA in real-life scenarios, where external events and unexpected swings can have a significant impact on trends.

6.2. Future Scope

To address the limitations of the ARIMA model, future research can explore combining ARIMA with advanced machine learning algorithms, such as neural networks, Support Vector Machines (SVMs), or ensemble methods, which possess the ability to identify non-linear patterns and adapt to sudden changes in data. By combining the ability of ARIMA to identify trends and the versatility of machine learning to identify advanced patterns, hybrid models can provide superior predictions. In addition, incorporating external factors such as macroeconomic factors, weather, and real-time data would make it an even more superior framework for predicting demand under uncertain conditions. Another alternative is to enhance the dataset with higher-order data, such as international sectors or new markets, thereby making the model more generalizable and improving its accuracy. Incorporating dynamic price models and customer behaviour models can also improve prediction ability, with an advanced process to predict air travel.

Acknowledgement: We gratefully acknowledge Dayananda Sagar Academy of Technology and Management for their support, guidance, and encouragement throughout this research.

Data Availability Statement: Data supporting this study are available from the corresponding author upon reasonable request, in compliance with ethical and institutional guidelines.

Funding Statement: No external funding was received for this research; it was completed through the collective effort of all authors.

Conflicts of Interest Statement: The authors declare no conflicts of interest. All contributors participated equally, and all sources were properly acknowledged.

Ethics and Consent Statement: The study adhered to ethical standards, ensuring informed consent, confidentiality, and participant privacy throughout the research process.

References

1. Z. Gundogdu, K. Babaoglu, M. Deveci, O. Tugral, and U. Zs, “A study of mortality in cardiac patients in a pediatric intensive care unit,” *Cureus*, vol. 11, no. 11, pp. 1-6, 2019.
2. S. Murthy, A. Leligdowicz, and N. K. J. Adhikari, “Intensive care unit capacity in low-income countries: a systematic review,” *PLoS One*, vol. 10, no. 1, pp. 1-12, 2015.
3. C. E. Kennedy and J. P. Turley, “Time series analysis as input for clinical predictive modeling: modeling cardiac arrest in a pediatric ICU,” *Theor. Biol. Med. Model.*, vol. 8, no. 1, pp. 1-25, 2011.
4. S. Hoover, E. Jackson, D. Paul, R. Locke, and M. Capan, “Time series analysis for forecasting hospital census: Application to the Neonatal Intensive Care Unit,” *Appl. Clin. Inform.*, vol. 7, no. 2, pp. 275–289, 2016.
5. Y. Xu, D. Han, T. Huang, X. Zhang, H. Lu, S. Shen, J. Lyu, and H. Wang, “Predicting ICU mortality in rheumatic heart disease: Comparison of XGBoost and logistic regression,” *Front. Cardiovasc. Med*., vol. 9, no. 2, pp. 1–12, 2022.
6. B. Matam, H. Duncan, and D. Lowe, “Machine learning based framework to predict cardiac arrests in a pediatric intensive care unit,” *Journal of Clinical Monitoring and Computing*, vol. 33, no. 4, pp. 713–724, 2019.
7. M. Pishgar, J. Theis, M. Del Rios, A. Ardati, H. Anahideh, and H. Darabi, “Prediction of unplanned 30-day readmission for ICU patients with heart failure,” *BMC Med. Inform. Decis. Mak.*, vol. 22, no. 1, pp. 1–12, 2022.
8. B. Lee, K. Kim, H. Hwang, Y. S. Kim, E. H. Chung, J.-S. Yoon, H. J. Cho, and J. D. Park, “Development of a machine learning model for predicting pediatric mortality in the early stages of intensive care unit admission,” *Sci. Rep.*, vol. 11, no. 1, pp. 1–7, 2021.
9. V. M. Ruiz, L. Saenz, A. Lopez-Magallon, A. Shields, H. A. Ogoe, S. Suresh, R. Munoz, and F. R. Tsui, “Early prediction of critical events for infants with single-ventricle physiology in critical care using routinely collected data,” *J. Thorac. Cardiovasc. Surg.*, vol. 158, no. 1, pp. 234–243, 2019.
10. M. O. Gani, S. Kethireddy, R. Adib, U. Hasan, P. Griffin, and M. Adibuzzaman, “Structural causal model with expert augmented knowledge to estimate the effect of oxygen therapy on mortality in the ICU,” *Artif. Intell. Med.*, vol. 137, no. 3, pp. 1-59, 2023.
11. S. Mehrmolaei and M. R. Keyvanpour, “Time series forecasting using improved ARIMA,” *in 2016 Artificial Intelligence and Robotics (IRANOPEN)*, Qazvin, Iran, 2016.
12. A. Jadon, A. Patil, and S. Jadon, “A comprehensive survey of regression-based loss functions for time series forecasting,” in Data Management, Analytics and Innovation, *Springer*, Singapore, 2024.
13. A. Sinha and A. Bagga, “Pediatric nephrology: Update for clinicians,” *Indian J. Pediatr.*, vol. 87, no. 8, pp. 598–599, 2020.
14. C. Zhao, P. Hu, X. Liu, X. Lan, and H. Zhang, “Stock market analysis using time series relational models for stock price prediction,” *Mathematics*, vol. 11, no. 5, pp. 1–13, 2023.
15. G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, “Time Series Analysis: Forecasting and Control.” *John Wiley and Sons*, Hoboken, New Jersey, United States of America, 2015.