

Article

# Demand Forecasting Model for Airline Flights Based on Historical Passenger Flow Data

Karina A. Lundaeva <sup>1,\*</sup>, Zakhar A. Saranin <sup>2</sup>, Kapiton N. Pospelov <sup>1</sup> and Aleksei M. Gintciak <sup>1</sup>

<sup>1</sup> Laboratory “Digital Modelling of Industrial Systems”, Advanced Engineering School “Digital Engineering”, Peter the Great St. Petersburg Polytechnic University, 195251 Saint Petersburg, Russia; kapiton.pospelov@spbpu.com (K.N.P.); aleksei.gintciak@spbpu.com (A.M.G.)

<sup>2</sup> Laboratory “Industrial Streaming Data Processing Systems”, Advanced Engineering School “Digital Engineering”, Peter the Great St. Petersburg Polytechnic University, 195251 Saint Petersburg, Russia; midway\_ht@mail.ru

\* Correspondence: karina.lundaeva@spbpu.com; Tel.: +7-(922)-440-80-16

**Featured Application:** The study can serve as a basis for the formation of data-driven passenger transport business models of various transport companies. The approach allows forecasting passenger demand in a wide class of systems.

**Abstract:** This paper addresses the problem of estimating passenger demand for flights, with a particular focus on the necessity of developing precise forecasts that incorporate intricate and interdependent variables for effective resource planning within the air transport industry. The present paper focuses on the development of a model for medium-term flight demand estimation by flight destinations. This is based on the analysis of historical airline data on dates, departure times, and passenger demand, as well as the consideration of the influence of macroeconomic indicators, namely gross regional product (GRP), median per capita income, and population of departure and arrival points. This paper reviews international experience in the development of demand forecasting models and their use for resource planning in the industry. The developed model was evaluated using historical data on demand for a single turnaround flight operated by an airline. The developed model allows for the forecasting of the distribution of potential demand for airline flight destinations in the medium term, utilizing comprehensive historical data on departure times and flight demand by destination.

**Keywords:** Facebook Prophet algorithm; passenger demand forecasting; passenger flow; prediction model; time-series forecasting



**Citation:** Lundaeva, K.A.; Saranin, Z.A.; Pospelov, K.N.; Gintciak, A.M. Demand Forecasting Model for Airline Flights Based on Historical Passenger Flow Data. *Appl. Sci.* **2024**, *14*, 11413. <https://doi.org/10.3390/app142311413>

Academic Editors: Marta Campos Ferreira, João Manuel R. S. Tavares and Teresa Galvão Dias

Received: 3 November 2024

Revised: 27 November 2024

Accepted: 6 December 2024

Published: 8 December 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The attainment of an effective management system within the air transport industry has become increasingly challenging as a result of several factors, including the intrinsic complexity of the systems themselves, the multitude of interrelationships between processes, and the prevalence of external uncertainty. The development of predictive models to assess the operational and financial performance of air transport companies is becoming a topic of considerable interest within the research community [1]. This involves the use of complex mathematical models that consider a multitude of factors influencing performance, including passenger demand dynamics, fuel consumption, airfares, and the condition of the aircraft fleet [2–4]. The development of numerical mathematical models in the air transport industry has been employed to address a range of challenges, including forecasting passenger demand for air travel and short-term flight bookings [5], predicting the distribution of passenger traffic in airport terminals [6], estimating airport delays [7], and estimating passenger load factors for flight service planning and inventory management decisions [8].

Nevertheless, it is the passenger demand estimation model and its accuracy that are of paramount importance for the rational planning of industry resources. The solution to this problem entails the identification of the factors affecting changes in air travel demand, an assessment of their extent of influence, and the selection of appropriate forecasting methods and tools [9]. The factors affecting final demand may be related to the pricing policies of airlines operating in the market under consideration, to a class of social influences, such as the average per capita income of the population of the departure and arrival points, and the purpose of the flights [10], and to geographical features, such as the transport accessibility of airports [11].

Due to economic changes and their impact on air traffic in the air transport industry, a medium-term planning horizon (approximately five years) is considered the optimal horizon for estimating flight demand [12]. Medium-term forecasts provide key data for making decisions on aviation infrastructure planning, in particular, airline route network load planning [13].

The object of the current study is the problem of medium-term forecasting of passenger demand for air transportation, which is caused by the complexity of modeling the relationships between numerous factors influencing demand. This problem covers aspects related to the development of mathematical and numerical models for assessing changes in passenger flows on a planning horizon of up to five years. The study examines the international experience in the development of passenger demand forecasting models in the air transport industry and proposes the creation of an applied solution based on an analysis of the industry's particular characteristics and existing solutions.

This study aims to develop a model for the medium-term estimation of passenger travel demand by flight destination. The study's objectives are to ascertain the functional requirements for the model through an analysis of industry-specific details, select forecasting methods and tools, describe the model's algorithm, test the model on historical data for one flight direction, and evaluate the accuracy of the demand forecast.

## 2. Materials and Methods

The principal challenge inherent in medium-term demand forecasting is the necessity to consider a multitude of intricate and interrelated variables, including country-level economic fluctuations [14,15], changes in consumer preferences [16], airport transport accessibility, and the competitive environment. Concurrently, demand estimation models must be capable of integrating various data types, including historical demand data, economic indicators, airfare data, socio-economic indicators (income and population), and other external factors [17].

In order to ensure the functionality of a demand model, it is essential to consider fluctuations in time and seasonal trends, the flexibility and adaptability of the models themselves, and the capacity to adjust forecasts in response to incoming data regarding changes in the external environment. The highlighted functional requirements ensure the development of efficient and adaptive flight demand forecasting models and should be taken into account in the development of the model proposed in the study.

### 2.1. Description of a Typical Problem of Medium-Term Forecasting of Demand for Flights

The demand for air transportation exhibits a complex nonlinear nature and is characterized by non-stationarity. In recent years, modern machine learning methods, particularly time series analysis and deep learning approaches, have been actively utilized for forecasting passenger demand. A common practice in the development of models for assessing and forecasting air transportation demand is the use of classical methods such as Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA), and Deep Learning Neural Networks (DLNN) [18].

Unlike specific machine learning methods aimed at identifying patterns in data, predictive modeling encompasses the entire process of developing mathematical models to quantitatively assess the accuracy of predictions for future data based on historical data

or training samples [19]. According to the development process of time series forecasting methods, existing approaches are classified into three categories: classical statistical methods, neural networks, and deep learning [20]. Hybrid models can be utilized in several ways. One approach involves combining forecasts, where predictions are first generated using each individual forecasting model, and the results are then integrated through weighted averages or a meta-model. Another approach involves sequential processing, where one method (e.g., ARIMA) is used to model the fundamental components (trend and seasonality), while another method (e.g., neural networks) analyzes the residuals. An additional option for using hybrid models is joint learning with parallel data processing, where different models are trained simultaneously, and their results are integrated [21].

The current focus in the field of time series forecasting is on the development of hybrid models that combine the advantages of statistical methods, neural networks, and deep learning. This approach enables the consideration of both internal dependencies and external influencing factors, which is particularly important in dynamic fields such as air transportation [22,23].

This study analyzes international experience in the development of air passenger demand forecasting models. It examines trends in development, the forecast horizon and modeling objectives chosen, the tools used, and the resulting forecast accuracy. Table 1 presents the findings of an analysis of the experience of the international scientific community in the development of flight demand forecasting models.

**Table 1.** Summary of passengers demand forecasting studies.

Title	Forecast Horizon	Methods Used	Modeling Results
SARIMA damp trend gray forecasting model for the airline industry [24]	Medium-term forecast for 8 routes from 2015Q1 to 2017Q4	Improved DTGM model: SARIMA with a dynamic seasonal damping factor (SDTTM)	The MSE values for the SDTGM model are less than the MSE values for the DTGM model. For the 8 routes analyzed, MAPE metrics are larger for DTGM than for SDTGM. The findings indicate that the proposed SDTGM model is more precise than the DTGM.
Predictive model of air transportation management based on intelligent algorithms of wireless network communication [25]	Medium-term forecast (01.01.2016–31.12.2019): the SARIMA model. Short-term forecast (6 months): stepwise regression. Short-term forecast (2021): combined model	Three forecasting models are combined: the exponential smoothing method, the stationary timeseries forecasting method, and the gray forecasting method	For short-term forecasting: ARIMA has the best accuracy, while the gray forecasting method is the least efficient. It is not necessarily the case that the combined model is superior to the individual models. For medium-term forecasts (2000–2020): the linear combined model demonstrates the greatest accuracy, while the exponential smoothing method exhibits the least efficient performance. The impact of the combined model varies
Forecasting air passenger numbers with a GVAR model [14]	Short-term forecast: one (h = 1) to four (h = 4) quarters ahead	Global vector autoregressive (GVAR) model	The accuracy of the models was assessed using MSE, MAE, and MAPE. The GVAR model demonstrates superior performance to the four benchmark models in the short term for h = 1, 2, 3.
Freight traffic of civil aviation volume forecast based on the hybrid ARIMA-LR model [26]	Long-term forecast for 100 months	ARIMA-LR is a combination of autoregressive integrated moving average (ARIMA) and linear regression (LR)	ARIMA-LR exhibits higher accuracy, as evidenced by lower scores in comparison to the ARIMA model. Specifically, the MAE, MSE, and MAPE metrics demonstrate a reduction of 1.06, 29.02, and 0.03, respectively. In comparison to LR, the indices are reduced by 3.92 and 0.06, respectively
A comparative analysis of the forecasting performance of SARIMA intervention and Prophet models for the number of airline passengers at Soekarno-Hatta International Airport [27]	Short-term forecast from 01.01.2022 to 31.03.2023	Seasonal Autoregressive Integrated Moving Average (SARIMA) and FB Prophet models	The SARIMA model demonstrates the most optimal performance with MAPE 28% and RMSE 433473. The Prophet model demonstrates the most optimal performance with MAPE 37% and RMSE 497154

The aforementioned analysis permits a comparison of existing models and methodologies for forecasting flight demand, with due consideration of the particulars of the data, available resources, and requisite forecast accuracy. This research proposes the utilization of the Prophet package for the purpose of modeling medium-term demand estimation for passenger air travel. Prophet offers a number of advantages, including the ability to account for seasonal variations in demand and the impact of holidays [28], flexibility in handling periodic and non-standard trends, and the capacity to address missing data and uncertainty [29]. The aforementioned advantages are especially pertinent in the context of developing a model for assessing the demand for passenger air travel in the medium term.

### 2.2. Algorithm for Medium-Term Flight Demand Forecasting

A three-stage approach to constructing the demand forecast is proposed at the conceptual level for model development. At the first stage, it is proposed to use historical data pertaining to airline flights over the past three years. This dataset encompasses information such as the date and time of departure, aircraft type, passenger traffic, load percentage, and other relevant details. The data should be subjected to analysis based on the demand training sample collected in order to ascertain the structure of the data as well as to determine some initial forecasting parameters (trend, seasonal index, etc.) [30]. The estimation of demand is based on the flight departure time and the number of passengers carried, which is taken as the final demand for the flight. In the construction of a forecast, it is of great importance to estimate demand at three levels of seasonality: annual, monthly, and weekly. This is due to the fact that demand for air travel is influenced by different factors, which manifest themselves in different time cycles [31].

The subsequent stage entails the examination of the influence of external elements through the utilization of multiple regression models. In this approach, the dependent variable is passenger demand, while the independent variables encompass external economic factors. It is assumed that demand-side factors should be taken into account in relation to the region of flight departure and arrival points. This is because the aggregate demand for a flight is contingent upon the characteristics of the start and end nodes of the flight [32].

The final stage of modeling entails the calibration of the obtained medium-term demand forecast, with due consideration of the overlapping functional dependencies of demand on external factors identified at the second stage. The final demand estimation is represented by the following dependence (see Equation (1)).

$$D(t) = D_0 * f_1(t) * f_2(t) * f_3(t) * f_{external}(t), \tag{1}$$

where  $D(t)$  is the distribution of predicted demand by hour,  $D_0$  is the forecast demand without taking into account seasonality,  $f_{external}(t)$  is the factor of influence of external economic parameters,  $t$  is the time of flight setting,  $f_1(t)$ ,  $f_2(t)$ , and  $f_3(t)$  are the factors of seasonality by year, week, and day of the week, respectively (see Equations (2)–(4)).

$$f_1(t) = 1 + k_m, \tag{2}$$

where  $k_m$  is the coefficient of seasonality by month.

$$f_2(t) = \sqrt{((1 + w_1 * k_w) * (1 + w_2 * \overline{k_w}))}, \tag{3}$$

$$f_3(t) = \sqrt{((1 + w_1 * k_h) * (1 + w_2 * \overline{k_h}))}, \tag{4}$$

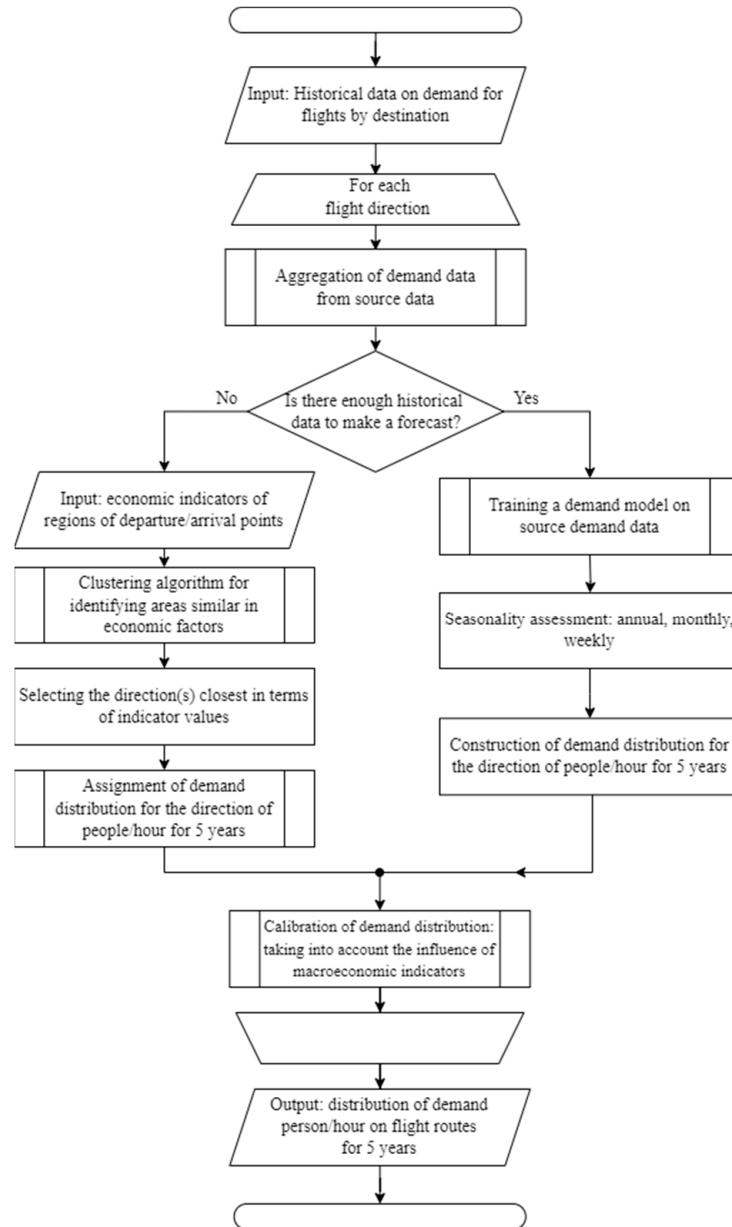
where  $w_1, w_2$  are the weights of the RMS smoothing function,  $k_w$  is the seasonality coefficient by week, and  $k_h$  is the seasonality coefficient by hour in a day.

If the first or second multiplier takes negative values (see Equations (3) and (4)), then only the second multiplier is taken with a weighting factor of one and no square root (see Equations (5) and (6)).

$$f_2(t) = 1 + \overline{k_w}, \tag{5}$$

$$f_3(t) = 1 + \overline{k}_t, \tag{6}$$

In the absence of sufficient historical data on flight destinations, it is proposed that the method of analogy be employed to forecast demand. In such instances, the demand for a flight for which there is insufficient data is assumed to be equal to the demand for the most similar destination in terms of external economic parameters. The issue is addressed by grouping comparable destinations and selecting the nearest destination through the implementation of the *k*-means clustering algorithm [33]. The algorithmic process for constructing a medium-term forecast of flight demand by air travel destination is illustrated in Figure 1.



**Figure 1.** Three-stage algorithm for medium-term flight demand forecasting.

The three-stage algorithm represents a structured approach to demand forecasting, consisting of sequential stages, each aimed at addressing a specific task—from data preparation to obtaining forecast values for the selected forecasting horizon. The presented algorithm forms the foundation of the developed model, where it is proposed that each of the described blocks will be implemented within the model.

The following section will present the mathematical models and tools that are necessary for the implementation of each of the aforementioned model development steps. Time-series models are used to construct medium-term forecasts of air travel demand based on historical data. Such models employ a univariate time series approach, whereby future values are predicted based on past observations. These observations are discrete and occur at equal time intervals. In the context of constructing the demand distribution, observations can be classified as discrete, with a unit of measurement of one hour [16,34]. This study employs the open-source forecasting package Facebook Prophet for time series modeling and the estimation of demand trends. This package is an additive linear model that combines several components to identify different patterns in the data, including trends, seasonality, holidays, and other relevant features [35,36]. The model is expressed by the following relationship (see Equation (7)):

$$y(t) = g(t) + s(t) + h(t) + \zeta(t), \quad (7)$$

where  $g(t)$  is a piecewise linear or logistic growth curve for modeling non-periodic changes in the time series,  $s(t)$  is a function responsible for modeling periodic changes associated with hourly, daily, weekly, and annual seasonality,  $h(t)$  is a function responsible for accounting for the impact of irregular holidays and user-defined events,  $\zeta(t)$  is an error function employed to account for any changes that were not accounted for by the model.

Furthermore, the model can incorporate a multiplicative function to illustrate an exponential trend, whereby the seasonal effect is multiplicative to the trend. This methodology permits the discrepancy between actual and predicted values to be distributed uniformly across the years, thereby enhancing the precision of the forecast. The general equation then assumes the following form [37] (see Equation (8)):

$$\hat{y}(t) = g(t) * (1 + mult.term) + add.term, \quad (8)$$

where  $g(t)$  is a piecewise linear or logistic growth curve for modeling non-periodic changes in time series,  $mult.term$  is the multiplicative component (usually used to model changes that are proportional to the current level of the time series),  $add.term$  is the additive component.

The Prophet model is used as part of machine learning and enables analysis of historical airline passenger demand data in order to identify dependencies and data features and significant statistical characteristics and to model periodic changes associated with weekly and annual seasonality [38]. The method is an efficient means of handling data containing seasonal impacts and is also capable of producing robust forecasts in the face of missing data and emissions. This is a particularly important attribute when analyzing flight demand [36,37]. It can be demonstrated that the utilization of the Prophet package will facilitate the replication of the continuation of the univariate time series of the value of demand for airline flights in the medium term.

In order to analyze the influence of external factors and model their relationship with the demand outcome, the construction of multiple regression models is proposed [39,40]. In order to ascertain the external economic factors influencing demand, GRP (gross regional product), median per capita income, and population size were selected for analysis. Data on the macroeconomic factors previously outlined should be collected in relation to the departure and arrival airport regions of the analyzed route network [9,41]. The objective of constructing a multiple regression model is to identify the coefficients of influence exerted by macroeconomic factors on passenger demand. This is achieved by employing historical data on demand and the dynamics of changes in the aforementioned factors. The multiple regression equation is as follows [42] (see Equation (9)):

$$Y_t = \beta_0 + \beta_{1t} * x_{1t} + \beta_{2t} * x_{2t} + \dots + \beta_{6t} * x_{6t} + \zeta_t, \quad (9)$$

where  $Y_t$  is the value of the total passenger demand for year  $t$ ,  $\beta_0$  is the intercept,  $\beta_{1t}, \beta_{2t}, \dots, \beta_{6t}$  are the regression coefficients determining the level of influence of indepen-

dent variables on the value of the total passenger demand for year  $t$ ,  $x_{1t}, x_{2t}, \dots, x_{6t}$  are the values of selected macroeconomic indicators corresponding to the region of the departure and arrival points of the considered destination by time for year  $t$ :  $x_{1t}, x_{2t}$  correspond to the values of the GRP indicators,  $x_{3t}, x_{4t}$  correspond to the values of the median per capita income,  $x_{5t}, x_{6t}$  correspond to the values of the population of departure and arrival points,  $\zeta_t$  is the model error, and  $t$  is the time index.

Additionally, in order to facilitate the incorporation of the independent values  $x_{1t}, x_{2t}, \dots, x_{6t}$  into the multiple regression model, it is essential to employ time series modeling of macroeconomic indicators for the independent variable  $X$  prior to the forecast period conclusion.

It is then proposed that an average linear model be constructed, equal to the average of the linear annual models over the historical annual periods. In order to achieve this, it is necessary to calculate the averaged regression coefficient for each historical year (see Equation (10)).

$$\bar{\beta}_j = \frac{1}{n} \sum_{i=1}^n \beta_{ji}, \tag{10}$$

where  $\bar{\beta}_j$  is the average coefficient for the  $j$ th independent variable,  $\beta_{ji}$  is the coefficient for  $j$ -th independent variable in a year  $i$ ,  $n$  is the number of years in the forecast period for which data needs to be averaged.

The averaged model  $\bar{Y}_t$  has coefficients averaged over all years (see Equation (11)):

$$\bar{Y}_t = \bar{\beta}_0 + \bar{\beta}_1 * x_{1t} + \bar{\beta}_2 * x_{2t} + \dots + \bar{\beta}_n * x_{nt} + \zeta_t, \tag{11}$$

where  $\bar{Y}_t$  is the predicted value of demand according to the averaged model for the year  $t$ ,  $\beta_0$  is the intercept,  $\beta_1, \beta_2, \dots, \beta_n$  are the averaged regression coefficients (the average values of the coefficients for all periods),  $x_{1t}, x_{2t}, \dots, x_{nt}$  are independent variables (the same as in Equation (9)),  $\zeta_t$  is a model error.

Furthermore, the model data must be normalized in accordance with the historical data from Equation (9), relative to the averaged model. Therefore, the parameter  $f_{external}(t)$  from Equation (1) will be expressed by the following dependence (see Equation (12)):

$$f_{external}(t) = Y_t / \bar{Y}_t, \tag{12}$$

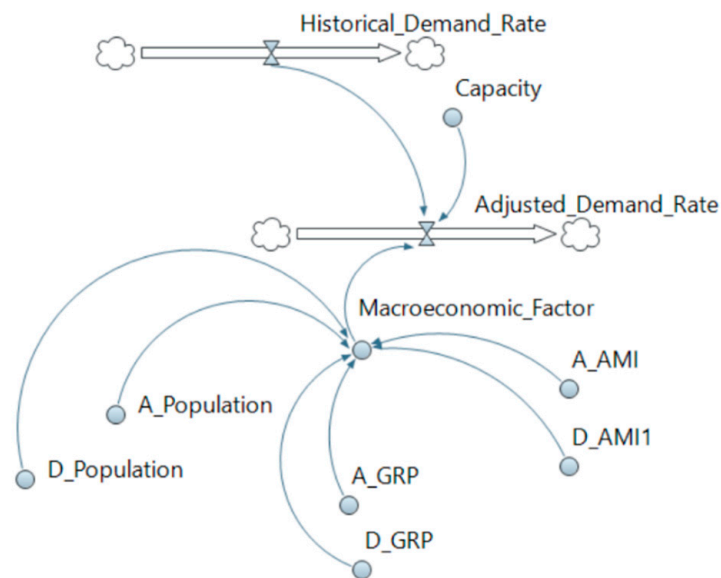
where  $\bar{Y}_t$  is the predicted value of demand according to the averaged model for the year  $t$  (see Equation (11)),  $Y_t$  is the actual value of the final passenger demand for the year  $t$  (see Equation (10)). The result is a series of data reflecting the functional dependence of flight demand on fluctuations in macroeconomic indicators.

The third stage of development proposes the utilization of a system dynamics approach to describe the influence of external factors on demand [43]. System dynamics models provide more reliable forecasts of short- and medium-term trends than statistical models. Additionally, they allow for the calibration of forecast demand distributions by incorporating external factors, thereby enhancing the forecast's informational value in the medium term [43]. The conceptual framework illustrated in Figure 2 depicts the interrelationship between the historical distribution of flight demand, forecast demand, and the impact of macroeconomic variables on the forecast.

The diagram below introduces the following designations:

Historical\_Demand\_Rate is the historical passenger demand; Capacity is the capacity of the aircraft operating the flight; Adjusted\_Demand\_Rate is the forecast demand; Macroeconomic\_Factor is the additive function of the influence of macroeconomic factors; A\_D\_Population is the weighting factor of the influence of the population values of the departure and arrival points, respectively; A\_D\_GRP is the weight coefficient of influence of the gross regional product values of the departure and arrival points, respectively; and A\_D\_AMI is the weight coefficient of influence of the median average per capita income values of the departure and arrival points, respectively.

Based on the modeling results, the forecast of demand distribution should be determined with discreteness for each hour of the forecast period from 2025 to 2029. To ensure the accuracy of this approach, statistical data on the demand for turnaround flight direction will be employed for validation purposes. Therefore, it is recommended that data be collected on the distribution of historical passenger demand and flight departure times for a single airline destination from 2019 to the first half of 2024. Subsequently, the findings on the forecast of demand distribution up to 2029 and an evaluation of the resulting forecast are presented.



**Figure 2.** Interrelationship between historical and forecast demand, taking into account the upper-level calibration.

### 3. Results

The modeling results present the distribution of demand for the airline's turnaround flight direction for each hour of the forecast period from 2025 to 2029. Consequently, two flight directions were considered, with demand determined independently: from A to B and in the reverse direction from B to A. The inputs to the model comprised historical passenger demand data for the A–B and B–A routes, together with historical time slot data, including dates and times of departure for all flights to and from each destination over the specified historical period. The data for developing the forecasting model were obtained from the airline's booking system and collected independently for each route under consideration from 2019 to the first half of 2024. Historical data on the final demand for a flight were equated to the number of tickets sold for each flight on the selected route, with demand calculated as the absolute number of passengers who completed the flight. In order to consider the influence of macroeconomic variables on the distribution of forecast demand, data on GRP, median per capita income, and population for departure and arrival points between 2014 and 2024 were collected. The data were sourced from publications by the Federal State Statistics Service, where each macroeconomic indicator was represented as an annual aggregate value over the 2014–2024 period.

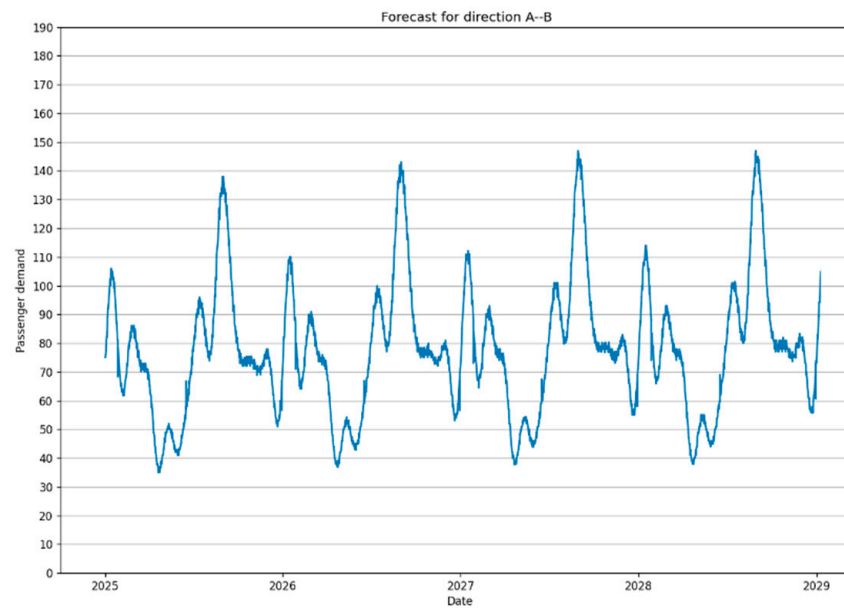
During the model development process, data preprocessing was carried out, including the stages of data collection and cleaning, unification of data formats, extraction of seasonal components at specific levels of seasonality, and determination of rules to account for the geographical specifics of flight routes. Initially, historical passenger demand data for routes A–B and B–A were combined into a single dataset, where each instance described the demand for a specific time slot, including the flight date and time. Outliers and anomalies in the historical data were then removed, specifically excluding records of flights with non-representative passenger demand values (significant deviations from the median demand



for a specific period). Missing values for time intervals were filled using linear interpolation or simulation based on the original time intervals. Economic indicators (regional GDP, income, and population) were normalized and synchronized with the historical demand data over time.

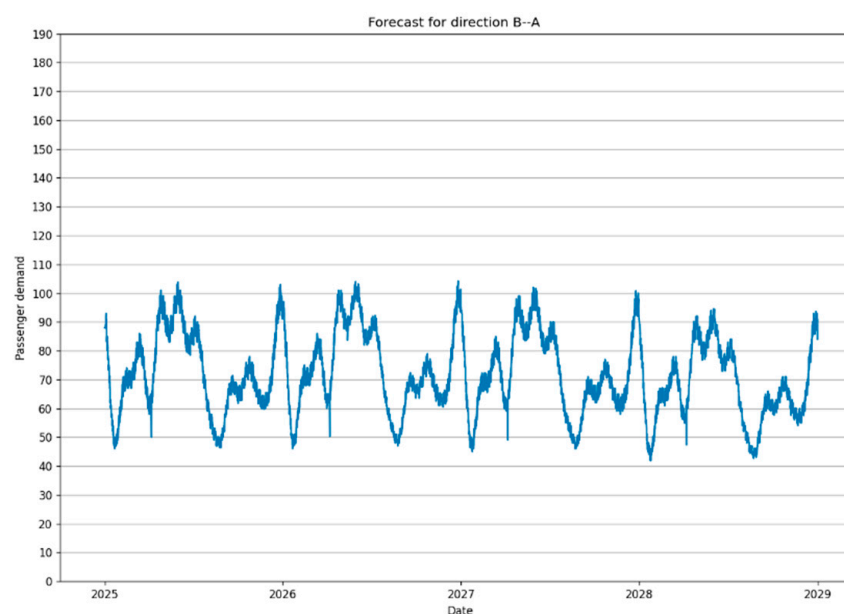
When forming training and testing datasets, data from 2014 to 2023 were used for training the model, while data from 2024 were reserved for testing. To account for the specifics of the routes, data for A–B and B–A were processed independently to eliminate mutual influence and ensure accurate demand modeling for each direction.

The results of the forecast demand estimation for the direction of flights from point A to point B from 2025 to 2029 are presented in the graph (see Figure 3).



**Figure 3.** Distribution of passenger demand forecast for A–B direction.

The results of the forecast demand estimate for the direction of flights from B to A from 2025 to 2029 are illustrated in the accompanying graph (see Figure 4).



**Figure 4.** Distribution of passenger demand forecast for B–A direction.

Table 2 shows the results of the RMSE, MAE, and MAPE forecast accuracy evaluation metrics for two directions: A–B and B–A. The RMSE and MAE metrics represent the mean model prediction error in units of the variable of interest, whereas the MAPE metric expresses the mean absolute error in percentage terms. The metrics were employed to assess the predictive capacity of the developed model by contrasting the actual demand metrics for 2023 with the model's predicted values for that year, based on training on historical data from 2019 to 2022.

**Table 2.** Results of passengers demand forecasting.

Metrics Results	Direction A–B	Direction B–A
RMSE	24.839	27.955
MAE	20.379	25.179
MAPE	35.705	31.435

#### 4. Discussion

The evaluation results indicate that the RMSE and MAE metrics are within an acceptable range. However, to enhance the quality of the modeling results, it is necessary to achieve lower scores, which represents a potential avenue for further research. A MAPE exceeding 30% for both directions indicates a substantial degree of percentage deviation, which may not be deemed acceptable for tasks that necessitate a high degree of accuracy.

RMSE measures the root mean square deviation between actual and predicted values, indicating the absolute average error in units of passenger demand. In this case, the RMSE for route A–B is 24.839, and for B–A it is 27.955. Values below 25–30 may be acceptable for models with large datasets, but in this case, the high RMSE suggests the potential for significant forecasting errors in individual time slots. MAE measures the mean absolute error, reflecting how much the model deviates from actual values on average. The MAE for A–B is 20.379 and for B–A is 25.179, indicating the typical deviation of the model from the actual data. Unlike RMSE, MAE is less sensitive to outliers. A low MAE value may indicate stable forecasting performance, but its increase for the B–A route may point to higher variability in passenger demand or lower model quality for this direction. MAPE expresses the mean absolute percentage error as a percentage of actual values, making it convenient for interpretation. For the obtained forecast, values of 35.705% for A–B and 31.435% for B–A show that, on average, the model deviates by more than 30% from actual demand. MAPE above 30% is considered high for forecasting tasks with strict accuracy requirements, but in the context of forecasting passenger demand for air travel, this could be attributed to anomalies in the data, insufficient consideration of external factors, or limitations of the model itself.

When compared to accuracy metrics from other forecasting methods, it is worth noting that hybrid approaches often outperform classical methods by accounting for both linear and nonlinear components. Research indicates that such models achieve MAPE levels of 10–15%, which is significantly better than the current results [22,44]. Thus, when compared to alternative methods, it was found that the presented model provides a baseline sufficient level of forecasting but falls short in performance compared to modern methods (hybrid approaches, neural network models, and transformers) that can significantly improve forecasting accuracy. For tasks requiring higher precision and shorter forecasting horizons, it is recommended to use alternative approaches to achieve a MAPE below 20% and improve RMSE/MAE.

At this juncture in the model development process, the upper-level calibration of the forecast results is being conducted, with due consideration given to the impact of the dynamics of macroeconomic indicators until the conclusion of the forecast period. In order to enhance the precision of the model, prospective avenues for advancement include the incorporation of the influence of competitor airline flight demand and alternative transportation modalities (public road and rail) to destinations analogous to the airline's route network. The incorporation of supplementary attributes would facilitate the estimation of

the prospective demand for flights, accounting for the probability that passengers may opt to fly with competing companies or alternative modes of transportation [45]. In order to account for the functional dependence of the demand estimate, it is necessary to collect the minimum dataset required for competitor flights and alternative modes of transport. This should include the minimum ticket price, the total passenger demand for competitor flights, and the journey time in hours for the alternative mode of transport.

## 5. Conclusions

The study yielded a predictive model for estimating airline flight demand based on the analysis of historical data to identify the seasonality of demand. The proposed model provides a medium-term forecast using machine learning and multiple linear regression methods, accounting for the impact of changes in macroeconomic factors on the final distribution of passenger demand by flight destination. The developed model was validated using historical data on demand and flight departure times from 2019 to the first half of 2024 for a revolving airline route comprising two flight destinations with independently determined demand: from A to B and from B to A. The values of the metrics for assessing the accuracy of the model results for the demand forecast for the two turnaround flight directions were 24.839 and 27.955, 20.379 and 25.179, 35.705, and 31.435 for RMSE, MAE, and MAPE, respectively.

The key advantages of the model include the consideration of macroeconomic factors, high flexibility for various forecasting horizons with appropriate preprocessing of data for the selected horizon, and demonstrated accuracy compared to baseline methods. In the future, the model can be compared with neural networks and other modern approaches to confirm its competitiveness.

One of the key limitations of the developed model is its dependency on the quality and completeness of the input data. For instance, gaps or errors in historical data on demand and flight schedules can negatively impact forecast accuracy. Another limitation is the model's restricted ability to account for changes in the competitive environment and the potential shift in passenger demand for air travel toward alternative transportation modes. Currently, the model does not include detailed data on competitor actions, such as price reductions, route changes, or increased flight frequencies. Additionally, the model may be less effective in forecasting demand for new routes or in regions with insufficient historical information, as these cases lack adequate data for training.

The potential for enhancement of the presented model lies in the advancement of forecast precision and the incorporation of a function elucidating the impact of the competitive environment (demand for flights by competitor airlines and alternative modes of transport) on the final distribution of passenger demand. The presented model can be used as a decision-making tool for production resource planning in an airline. The integration of the model with internal decision support systems enables the development of a comprehensive system for the estimation and optimization of passenger revenue under a variety of flight schedule planning scenarios.

**Author Contributions:** Conceptualization, K.N.P. and A.M.G.; methodology, Z.A.S.; software, Z.A.S.; validation, K.A.L., Z.A.S. and K.N.P.; formal analysis, Z.A.S.; investigation, K.A.L.; resources, K.N.P.; data curation, A.M.G.; writing—original draft preparation, K.A.L.; writing—review and editing, K.N.P.; visualization, Z.A.S.; supervision, A.M.G.; project administration, A.M.G.; funding acquisition, A.M.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Ministry of Science and Higher Education of the Russian Federation, grant number 075-03-2024-004.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Bakreen, S.; Markovskaya, E.; Merzlikin, I.; Mottaeva, A. Development of the approach to the analysis of aviation industry's adaptation to seasonal disruptions. *Transp. Res. Procedia* **2022**, *63*, 1431–1443. [[CrossRef](#)]
- Seymour, K.; Held, M.; Georges, G.; Boulouchos, K. Fuel Estimation in Air Transportation: Modeling global fuel consumption for commercial aviation. *Transp. Res. Part D Transp. Environ.* **2020**, *88*, 102528. [[CrossRef](#)]
- Yan, C.; Barnhart, C.; Vaze, V. Choice-based airline schedule design and fleet assignment: A decomposition approach. *Transp. Sci.* **2022**, *56*, 1410–1431. [[CrossRef](#)]
- Wei, K.; Vaze, V.; Jacquillat, A. Airline timetable development and fleet assignment incorporating passenger choice. *Transp. Sci.* **2020**, *54*, 139–163. [[CrossRef](#)]
- He, H.; Chen, L.; Wang, S. Flight short-term booking demand forecasting based on a long short-term memory network. *Comput. Ind. Eng.* **2023**, *186*, 109707. [[CrossRef](#)]
- Lin, L.; Liu, X.; Liu, X.; Zhang, T.; Cao, Y. A prediction model to forecast passenger flow based on flight arrangement in airport terminals. *Energy Built Environ.* **2023**, *4*, 680–688. [[CrossRef](#)]
- Güvercin, M.; Ferhatosmanoglu, N.; Gedik, B. Forecasting flight delays using clustered models based on airport networks. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 3179–3189. [[CrossRef](#)]
- van der Walt, A.; Bean, W.L. Inventory management for the in-flight catering industry: A case of uncertain demand and product substitutability. *Comput. Ind. Eng.* **2022**, *165*, 107914. [[CrossRef](#)]
- Wang, S.; Gao, Y. A literature review and citation analyses of air travel demand studies published between 2010 and 2020. *J. Air Transp. Manag.* **2021**, *97*, 102135. [[CrossRef](#)]
- Firat, M.; Yiltas-Kaplan, D.; Samli, R. Forecasting Air Travel Demand for Selected Destinations Using Machine Learning Methods. *JUCS J. Univers. Comput. Sci.* **2021**, *27*, 564–581. [[CrossRef](#)]
- Yirgu, K.W.; Kim, A.M. Airport choices and resulting catchments in the US Midwest. *J. Transp. Geogr.* **2024**, *114*, 103743. [[CrossRef](#)]
- Rodriguez, Y.; Pineda, W.; Olariaga, O.D. Air traffic forecast in post-liberalization context: A Dynamic Linear Models approach. *Aviation* **2020**, *24*, 10–19. [[CrossRef](#)]
- Li, X.; de Groot, M.; Bäck, T. Using forecasting to evaluate the impact of COVID-19 on passenger air transport demand. *Decis. Sci.* **2023**, *54*, 394–409. [[CrossRef](#)] [[PubMed](#)]
- Gunter, U.; Zekan, B. Forecasting air passenger numbers with a GVAR model. *Ann. Tour. Res.* **2021**, *89*, 103252. [[CrossRef](#)]
- Zhang, F.; Graham, D.J. Air transport and economic growth: A review of the impact mechanism and causal relationships. *Transp. Rev.* **2020**, *40*, 506–528. [[CrossRef](#)]
- Banerjee, N.; Morton, A.; Akartunalı, K. Passenger demand forecasting in scheduled transportation. *Eur. J. Oper. Res.* **2020**, *286*, 797–810. [[CrossRef](#)]
- Tolcha, T.D.; Bråthen, S.; Holmgren, J. Air transport demand and economic development in sub-Saharan Africa: Direction of causality. *J. Transp. Geogr.* **2020**, *86*, 102771. [[CrossRef](#)]
- Kanavos, A.; Kounelis, F.; Iliadis, L.; Makris, C. Deep learning models for forecasting aviation demand time series. *Neural Comput. Appl.* **2021**, *33*, 16329–16343. [[CrossRef](#)]
- Gao, Z.; Mavris, D.N. Statistics and machine learning in aviation environmental impact analysis: A survey of recent progress. *Aerospace* **2022**, *9*, 750. [[CrossRef](#)]
- Liu, Z.; Zhu, Z.; Gao, J.; Xu, C. Forecast methods for time series data: A survey. *IEEE Access* **2021**, *9*, 91896–91912. [[CrossRef](#)]
- Zhang, D.; Wang, P.; Ding, L.; Wang, X.; He, J. Spatio-Temporal Contrastive Learning-Based Adaptive Graph Augmentation for Traffic Flow Prediction. *IEEE Trans. Intell. Transp. Syst.* **2024**, 1–15. [[CrossRef](#)]
- Jin, F.; Li, Y.; Sun, S.; Li, H. Forecasting air passenger demand with a new hybrid ensemble approach. *J. Air Transp. Manag.* **2020**, *83*, 101744. [[CrossRef](#)]
- Hajirahimi, Z.; Khashei, M. Hybridization of hybrid structures for time series forecasting: A review. *Artif. Intell. Rev.* **2023**, *56*, 1201–1261. [[CrossRef](#)]
- Carmona-Benítez, R.B.; Nieto, M.R. SARIMA damp trend grey forecasting model for airline industry. *J. Air Transp. Manag.* **2020**, *82*, 101736. [[CrossRef](#)]
- La, J.; Bil, C.; Heiets, I.; Lau, K.A. Predictive model of air transportation management based on intelligent algorithms of wireless network communication. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 1414539. [[CrossRef](#)]
- Chen, B.; Liu, J.; Ruan, Z.; Yue, M.; Long, H.; Yao, W. Freight traffic of civil aviation volume forecast based on hybrid ARIMA-LR model. In Proceedings of the International Conference on Smart Transportation and City Engineering (STCE 2022), SPIE, Chongqing, China, 12–14 January 2022; Volume 12460, pp. 682–689.
- Aziza, V.N.; Moh'd, F.H.; Maghfiroh, F.A.; Notodiputro, K.A.; Angraini, Y. Performance comparison of sarima intervention and prophet models for forecasting the number of airline passenger at Soekarno-Hatta international airport. *BAREKENG J. Ilmu Mat. Dan Terap.* **2023**, *17*, 2107–2120. [[CrossRef](#)]
- Liu, Y.; Feng, G.; Chin, K.S.; Sun, S.; Wang, S. Daily tourism demand forecasting: The impact of complex seasonal patterns and holiday effects. *Curr. Issues Tour.* **2023**, *26*, 1573–1592. [[CrossRef](#)]

29. Svekolnikova, E.A.; Panovskiy, V.N. Review of Open-Source Libraries for Solving Time Series Forecasting Problems. *Model. I Anal. Dannikh Model. Data Anal.* **2024**, *14*, 45–61. (In Russian) [[CrossRef](#)]
30. Pradita, S.P.; Ongkunaruk, P.; Leingpibul, T. Utilizing an intervention forecasting approach to improve reefer container demand forecasting accuracy: A case study in Indonesia. *Int. J. Technol.* **2020**, *11*, 144–154. [[CrossRef](#)]
31. Martin-Rodriguez, G.; Caceres-Hernandez, J.J. Seasonal variations in daily data: An application to air passenger arrivals. *J. Air Transp. Manag.* **2023**, *110*, 102419. [[CrossRef](#)]
32. Tirtha, S.D.; Bhowmik, T.; Eluru, N. Understanding the factors affecting airport level demand (arrivals and departures) using a novel modeling approach. *J. Air Transp. Manag.* **2023**, *106*, 102320. [[CrossRef](#)]
33. Chen, J.H.; Wei, H.H.; Chen, C.L.; Wei, H.Y.; Chen, Y.P.; Ye, Z. A practical approach to determining critical macroeconomic factors in air-traffic volume based on K-means clustering and decision-tree classification. *J. Air Transp. Manag.* **2020**, *82*, 101743. [[CrossRef](#)]
34. Al-Sultan, A.; Al-Rubkhi, A.; Alsaber, A.; Pan, J. Forecasting air passenger traffic volume: Evaluating time series models in long-term forecasting of Kuwait air passenger data. *Adv. Appl. Stat.* **2021**, *70*, 69–89. [[CrossRef](#)]
35. Agyemang, E.F.; Mensah, J.A.; Ocran, E.; Opoku, E.; Nortey, E.N. Time series based road traffic accidents forecasting via SARIMA and Facebook Prophet model with potential changepoints. *Heliyon* **2023**, *9*, e22544. [[CrossRef](#)] [[PubMed](#)]
36. Gull, K.; Kanakaraddi, S.; Chikaraddi, A. COVID-19 outbreak prediction using additive time series forecasting model. *Trends Sci.* **2022**, *19*, 1919. [[CrossRef](#)]
37. Parizad, A.; Hatziadoniu, C.J. Using prophet algorithm for pattern recognition and short term forecasting of load demand based on seasonality and exogenous features. In Proceedings of the 2020 52nd North American Power Symposium (NAPS), IEEE, Tempe, AZ, USA, 11–13 April 2021; pp. 1–6.
38. Chuwang, D.D.; Chen, W. Forecasting daily and weekly passenger demand for urban rail transit stations based on a time series model approach. *Forecasting* **2022**, *4*, 904–924. [[CrossRef](#)]
39. Ma'ruf, A.; Nasution, A.A.R.; Leuveano, R.A.C. Machine Learning Approach for Early Assembly Design Cost Estimation: A Case from Make-to-Order Manufacturing Industry. *Int. J. Technol.* **2024**, *15*, 1037–1047. [[CrossRef](#)]
40. Wang, X.; Cai, J.; Wang, J. A panel data model to predict airline passenger volume. *Digit. Transp. Saf.* **2024**, *3*, 46–52. [[CrossRef](#)]
41. Zachariah, R.A.; Sharma, S.; Kumar, V. Systematic review of passenger demand forecasting in aviation industry. *Multimed. Tools Appl.* **2023**, *82*, 46483–46519. [[CrossRef](#)]
42. Li, C. Combined forecasting of civil aviation passenger volume based on ARIMA-REGRESSION. *Int. J. Syst. Assur. Eng. Manag.* **2019**, *10*, 945–952. [[CrossRef](#)]
43. Suryani, E.; Chou, S.Y.; Chen, C.H. Air passenger demand forecasting and passenger terminal capacity expansion: A system dynamics framework. *Expert Syst. Appl.* **2010**, *37*, 2324–2339. [[CrossRef](#)]
44. Zhao, S.; Mi, X. A novel hybrid model for short-term high-speed railway passenger demand forecasting. *IEEE Access* **2019**, *7*, 175681–175692. [[CrossRef](#)]
45. Lubis, H.A.; Pantas, V.B.; Farda, M. Demand forecast of Jakarta-Surabaya high speed rail based on stated preference method. *Int. J. Technol.* **2019**, *10*, 405–416. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.