Global Demand Forecast Model

Osama Alsalous

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science

In

Civil Engineering

Antonio Trani

Montasir Abbas

Pamela Murray-Tuite

December 3, 2015

Blacksburg, VA

Keywords: air transport demand forecast, regression model, econometric modeling

ABSTRACT

Air transportation demand forecasting is a core element in aviation planning and policy decision making. NASA Langley Research Center addressed the need of a global forecast model to be integrated into the Transportation Systems Analysis Model (TSAM) to fulfil the vision of the Aeronautics Research Mission Directorate (ARMD) at NASA Headquarters to develop a picture of future demand worldwide. Future forecasts can be performed using a range of techniques depending on the data available and the scope of the forecast. Causal models are widely used as a forecasting tool by looking for relationships between historical demand and variables such as economic and population growth. The Global Demand Model is an econometric regression model that predicts the number of air passenger seats worldwide using the Gross Domestic Product (GDP), population, and airlines market share as the explanatory variables. GDP and Population are converted to 2.5 arc minute individual cell resolution and calculated at the airport level in the geographic area 60 nautical miles around the airport. The global demand model consists of a family of models, each airport is assigned the model that best fits the historical data. The assignment of the model is conducted through an algorithm that uses the R^2 as the measure of Goodness-of-Fit in addition to a sanity check for the generated forecasts. The output of the model is the projection of the number of seats offered at each airport for every year up to the year 2040.

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my advisor Dr. Antonio Trani for the continuous support, guidance, and motivation he provided during the course of my studies and research at Virginia Tech.

I would like to thank Dr. Monastir Abbas for enriching my experience at graduate school by being a source of knowledge and inspiration both in classes and research. I would also like to thank Dr. Pamela Murray-Tuite for her genuine support and for all that I learned from her classes.

The research and development effort in this study was carried out as a part of a research project for NASA Langley Research Center. I am grateful for the opportunity to work on this project and would like to acknowledge Ty Vincent Marien, Jeff Viken, Sam Dollyhigh and Tech-Seng Kwa for their guidance and valuable input during the development of the project.

I would like to thank my friends around the world for their encouragement and support. I am also grateful to all the great friends that I made during my stay in Blacksburg, with special thanks to my current and former laboratory mates, my friends in TISE program, my teammates from Salsa Tech and many other friends.

Last but not least, my deepest thanks go to my family for their unconditional support, confidence, and love that helped me accomplish this work.

Table of Contents

List of Figures

List of Tables

1 INTRODUCTION

NASA Langley Research Center employs the Transportation Systems Analysis Model (TSAM) to fulfil the vision of the Aeronautics Research Mission Directorate (ARMD) at NASA Headquarters. ARMD is responsible for establishing a strategic systems analysis capability focused on understanding system-level impacts of NASA's programs, the potential for integrated solutions, and the development of options for new investment and partnerships.

The current version of TSAM has the capability of forecasting long distance travel in the continental U.S. up to the year 2050. TSAM is a nationwide transportation demand model that has an international module that is U.S. centric. The current model predicts international travel from the U.S. to 10 World regions. The development of a global demand forecast model is needed to add the capability of forecasting air traffic at airport level within World regions or between these regions to be able to respond to changes in the budgetary and/or technical targets of ARMD in an effective manner.

Future airport forecasts can be performed using a range of techniques depending on the data available and the scope of the forecast. Examples include Time-series models and Causal models. Time-series models rely entirely on historical values of the variable being projected. Time-series models apply statistical analysis to historical patterns over time and assume that future values will grow in similar patterns. This type of model is simplistic and considered useful for short term planning (Commission 2013).

Causal models are more complex forecasts that look for past relationships between demand and variables such as economy, population, travel costs, market concentration or other variables to explain the historical patterns of the variable being projected (Commission 2013; Chunshui et al. 2012). In the Causal model approach, a picture of future demand can be developed by applying the historical relationships to existing economic and demographic forecasts. Causal models are widely used as a forecasting tool, and the most commonly used causal models are regression-based models (Shen 2006).

The Global Demand Model developed in this report is an econometric regression model that predicts the number of air passenger seats worldwide using the Gross Domestic Product (GDP), population, and airlines market share as the explanatory variables. The forecast horizon of the model is the year 2040. The choice of explanatory variables used to construct the model was based on the variables that are included in economic forecasts.

2 LITERATURE REVIEW

Air transportation demand forecasting is a core element in Aviation Planning. The literature includes numerous papers on air transportation forecasting. Air transportation demand is usually modeled as a function of economic and demographic factors. Gross Domestic Product (GDP) and population are the most commonly used variables as demand driving forces (Grosche et al. 2007; Suryani et al. 2010; Chunshui et al. 2012; Commission 2013).

Linear regression is found to be the most common form of forecast model. For example, The forecasting technique used by The International Civil Aviation Organization (ICAO) includes linear regression using GDP (Chunshui et al. 2012). In another study, Karlaftis et al. developed linear regression models for Miami International (MIA) and Frankfurt International Airports (FRA) using GDP and Gross National Product (GNP) to forecast the international and domestic air travel respectively (Karlaftis et al. 1996). In their study, they highlighted the importance of postfact analysis to test the accuracy of the models. They concluded that R^2 values alone should not be the only criterion for selecting forecasting models, it should be supported by a check of the model results.

In previous forecasting efforts, there are many analyses that focused on air transportation demand forecasting in a specific country or region. For example, the relationship between air transport demand and GDP in Brazil between the years 1966 and 2006 was investigated by Marazzo et al. (2010). The findings of this analysis suggest that there is a strong positive reaction of demand due to a positive change in GDP. Another example of international aviation forecast was performed by Shen (2006). She developed a model to predict international flight operations in the United States. This model forecasts the number of international passenger enplanements at sixty-six U.S. airports, whose combined operations accounted for 99.8% of the total international passenger flight operations in National Airspace System (NAS) in 2004. The forecasted passenger enplanements are distributed among international airports in the U.S. using individual airport market share factors. GDP was used as the predictor variable.

The Terminal Area Forecast (TAF) is the official FAA forecast of aviation activity for U.S. airports. It contains active airports in the National Plan of Integrated Airport Systems (NPIAS). The forecast is based on an econometric model that uses fares, demographics, and economic factors as the independent variables (FAA 2013). Unfortunately, details about model specification are not publicly available. The TAF forecast is updated annually, the most recent TAF report available spans from 2014 to the year 2040.

Global air passenger forecasts are available from organizations such as The International Air Transport Association (IATA), The International Civil Aviation Organization (ICAO) and aircraft manufacturers such as Boeing and Airbus. IATA published three 20-year air passenger forecasts extending from 2014 up to 2034, each based on a specific growth scenario (Pearce 2015). The forecast is an aggregate model to estimate passengers for all World regions. Under the most optimistic growth scenario, the forecast suggests a 280% increase in passenger demand by 2034. Conversely, the more conservative growth scenario predicts a 180% increase in passenger demand. ICAO, Boeing, and Airbus also provide air passenger forecasts that aggregate across World, regional and country levels (Gillen 2009; Chunshui et al. 2012).

The Global Demand Model combines the global scope with airport level analysis. Based on the reviewed previous studies, we did not find any study that matches the characteristics of our scope. In the analysis we adopted regression models with GDP, population, and airline market share as predictor variables.

3 MODEL DEVELOPMENT

3.1 Overview

The Global Demand Model is an econometric model that produces future estimates of passenger demand Worldwide. The model covers a large scope of 3,017 airports around the world. The global scope and the ability to produce forecast at the individual airport level are unique features of this project. The Global Demand Model consists of four steps shown in Figure 1.

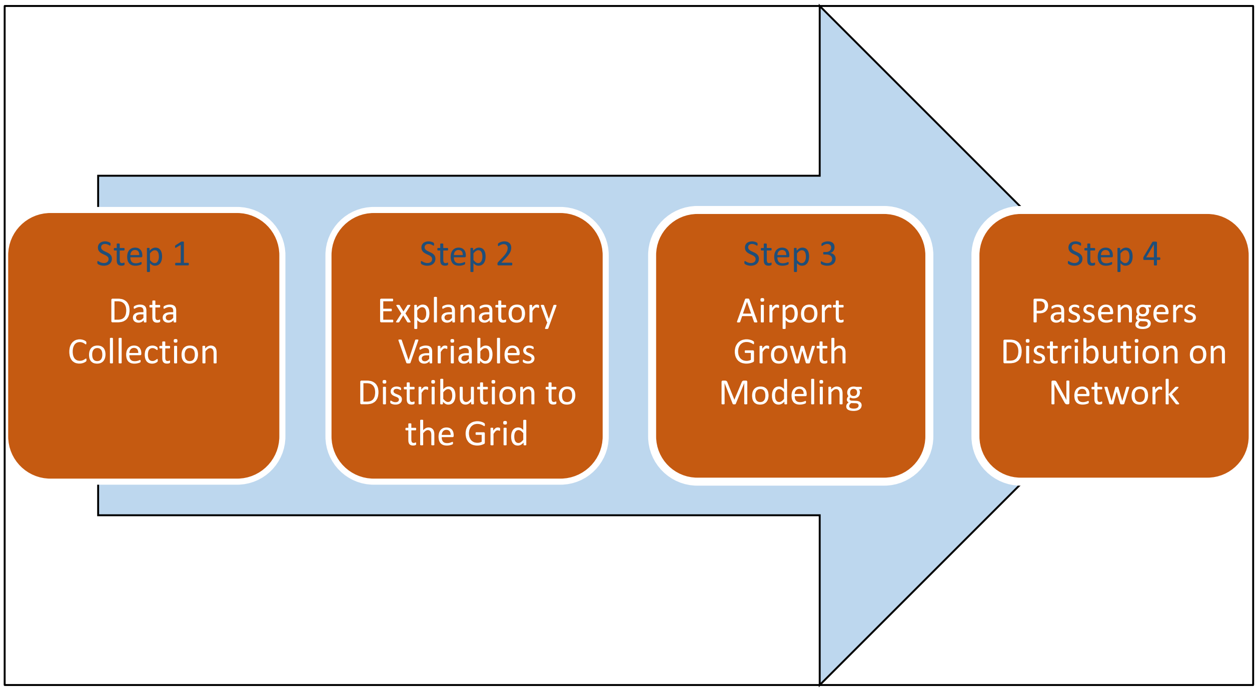

Figure 1: Global Demand Forecast Model Framework

3.2 Data Collection and Processing

The development of the Global Demand Model requires historical passenger data as well as data to produce explanatory variables to be used in the forecast process. Data collected from different sources include aggregate country-level data such as Gross Domestic Product (GDP), GDP per capita, population, exports, employment ratio, and others. Due to the spatial nature of the model, we also employ more detailed data sources that include a global gridded data with a resolution of 2.5 minutes (2.5 nm at the equator). Gridded data is intended to give more accurate modeling results due to the disaggregate spatial distribution of the data. Gridded data sets are available for some indicators such as GDP, population, and purchasing power parity (PPP).

3.2.1 ***Population Data***

The population data sources used in the Global Demand Model include the following:

1. Gridded Population of the World (GPW) v3 (Balk et al. 2010).

GPWv3 dataset consists of gridded population estimates for the World for years 1990, 1995, and 2000. In addition, “forecasts” for 2005 and 2010 are available. The level of detail is 2.5-minute individual cell resolution (2.5 nm at the equator) which is a large amount of data to handle (over 29 million cells). Note that the data is originally adjusted to meet the United Nations (UN) national totals.

2. United Nations estimates by country (UN 2013).

In the Global Demand Model, GPW is considered the main data source for analysis. GPW has information up to the year 2010. The UN totals for the following years were distributed using similar patterns as the historical data from GPW. This maintained data consistency for all years. It is assumed that the distribution of population on the grid has the same pattern for all cells on the grid as the historical data up to the year 2040.

The distribution of the country level population to the 2.5-minute grid was a time consuming process due to the computer processing power it requires. Firstly, The MATLAB script that we developed calculates and saves the proportion of the national total population contained within each cell based on the GPW patterns. This was done for each individual cell for all countries. In the second step, the script calculates the population in each 2.5-minute cell by multiplying the cell proportion by the UN total for the country where the cell is located. It then assigns the population value of the cell to the corresponding cell in the global grid. It takes 0.032 milliseconds to perform the calculations for a single cell using a computer with a 3.60GHz Quad Core processor and 32 GB of RAM. That is equal to 2,777 seconds for all cells for one year only. This process had to be executed for all years from 2005 to the year 2040, and that is equal to about 27.77 hours.

3.2.2 Economic Data

Economic data sources used in the Global Demand Model include the following:

1. Geographically based Economic data (G-Econ data) (Nordhaus et al. 2006).
2. U.S. Department of Agriculture (USDA) International GDP forecast (USDA 2014).

The latest G-Econ dataset consists of GDP estimates for the World for the year 2005 with data resolution of one degree of Latitude by one degree of Longitude. The G-Econ datasets are available up to the year 2005. The country level totals from other sources for the following years are distributed on the grid in similar patterns as the historical data was distributed to keep the data consistent and still maintain the same resolution for all years. An assumption is made that the distribution of economic indicators on the grid has the same pattern for all cells on the grid as the historical data throughout the project horizon up to the year 2040.

The one-degree resolution data was converted to a 2.5-minute resolution level to match the level of detail available in the GPW population data. Typically, the one-degree by one-degree cell is converted to 576 (24 x 24) cells of the 2.5-minute resolution since one degree of Lat or Lon is equal to 60 minutes. Each cell of the grid in the one-degree by one-degree resolution covers a large area, some cells cross country borders and cover parts of the ocean. This issue was taken into consideration when the value of each cell was distributed to the finer 2.5-minute resolution level by dividing the total GDP in the cell by the number of the 2.5-minute cells that actually are located within the respective country.

Figure 2 shows an example of the data distribution process from one-degree to 2.5-minute resolution for a cell located in the United Kingdom with coordinates (54 Lat, -1 Lon) of the southwest corner. The cell in the example is shown in Figure 2 with yellow borders and it can be seen that it covers a part of the ocean. The total GDP in this particular cell was 6.460 US$ billions in the year 2005 according to the G-Econ dataset. The 6.460 US$ billions were divided by 224 cells of the 2.5-minute resolution instead of the typical 576 cells because 352 cells fall into the ocean. This way we were able to distribute the GDP value to cells that are actually located in the country (i.e., hatched area in Figure 2) without losing any data.

The cells within each country were identified using the Gridded Population of the World, Version 3 (GPWv3) National Identifier Grid. This grid is derived from the land area grid to create a raster surface where cells that cover the same country or territory have the same identifier value. The cells in the oceans have a unique identifier as well (Center for International

Earth Science Information Network - CIESIN - Columbia University and Centro Internacional de Agricultura Tropical - CIAT 2005).

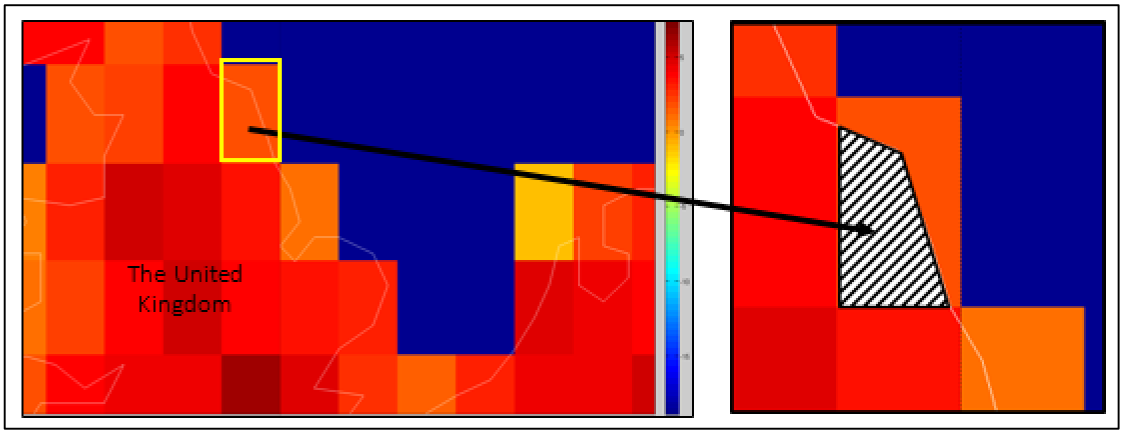

Figure 2: Distribution of economic data to 2.5-minute resolution

Figure 3 shows the framework of the distribution of country level forecast to 2.5-minute grid level that is used in the model. The national level GDP values were converted from country level to cell level using a similar procedure to the one described in section 3.2.1. This process also takes about 27.77 hours to be executed for all years from 2005 to the year 2040 using a computer with the same specifications described in section 3.2.1.

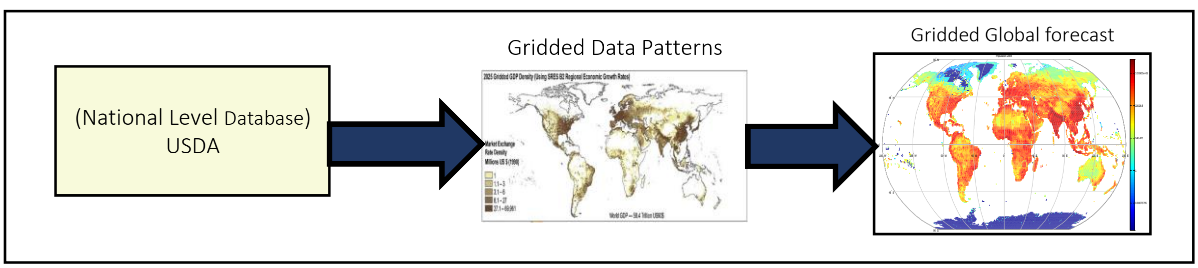

Figure 3: Distribution of country level forecast to grid level

3.2.3 Airport Passengers Data

The Global Demand Model requires a Worldwide historical passenger data at the individual airport level. The data is used as a set of dependent variables using regression analysis in the model. Data sources include:

1. ICAO DATA+

   A dataset from the International Civil Aviation Organization (ICAO) was acquired for the project. It covers historical data for 808 major airports Worldwide. The data includes:

   - Number of passengers Embarked
   - Number of Passengers Disembarked
   - Transit Passengers

2. OAG Data

   A dataset from the Official Airlines Guide (OAG) was used in the project. It includes historical data for 3,661 airports, with 60 data fields such as:

   - Flight origin and destination airports.
   - Available seats per flight
   - Aircraft type
   - Flight frequency
   - Flight distance
   - Operating Airlines
   - Airport geographic location

   Airports in the OAG dataset belong to different geographic regions. OAG categorized airports by location based on the world region in which they are located. Table 1 lists the 17 world regions defined by OAG.

Table 1: OAG world regions

	Region	Description
1	AF1	Africa : North Africa
2	AF2	Africa : Southern Africa
3	AF3	Africa : Central/Western Africa
4	AF4	Africa : Eastern Africa
5	AS1	Asia : South Asia
6	AS2	Asia : Central Asia
7	AS3	Asia : South East Asia
8	AS4	Asia : North East Asia
9	EU1	Europe : Western Europe
10	EU2	Europe : Eastern/Central Europe
11	LA1	Latin America : Caribbean
12	LA2	Latin America : Central America
13	LA3	Latin America : Upper South America
14	LA4	Latin America : Lower South America
15	ME1	Middle East
16	NA1	North America
17	SW1	Southwest Pacific

The two sources mentioned above can be used together to cover all World airports considered (3000+). The ICAO passenger data was organized in a MATLAB data structure to be used in the model. However, The ICAO data has many gaps in the records for some airports due to missing years. Inconsistent level of detail of the available records was another issue. The data consists of fields sorted by flight type and broken down into Embarked/Disembarked passengers in addition to total passengers. Some airports have only total passenger counts whereas others have them as breakdown of Embarked/Disembarked passengers. Figure 4 shows the procedure that we initially developed to use the OAG data to complement the ICAO data. However, The OAG data source was used as the main source due to the lack of quality in the ICAO dataset.

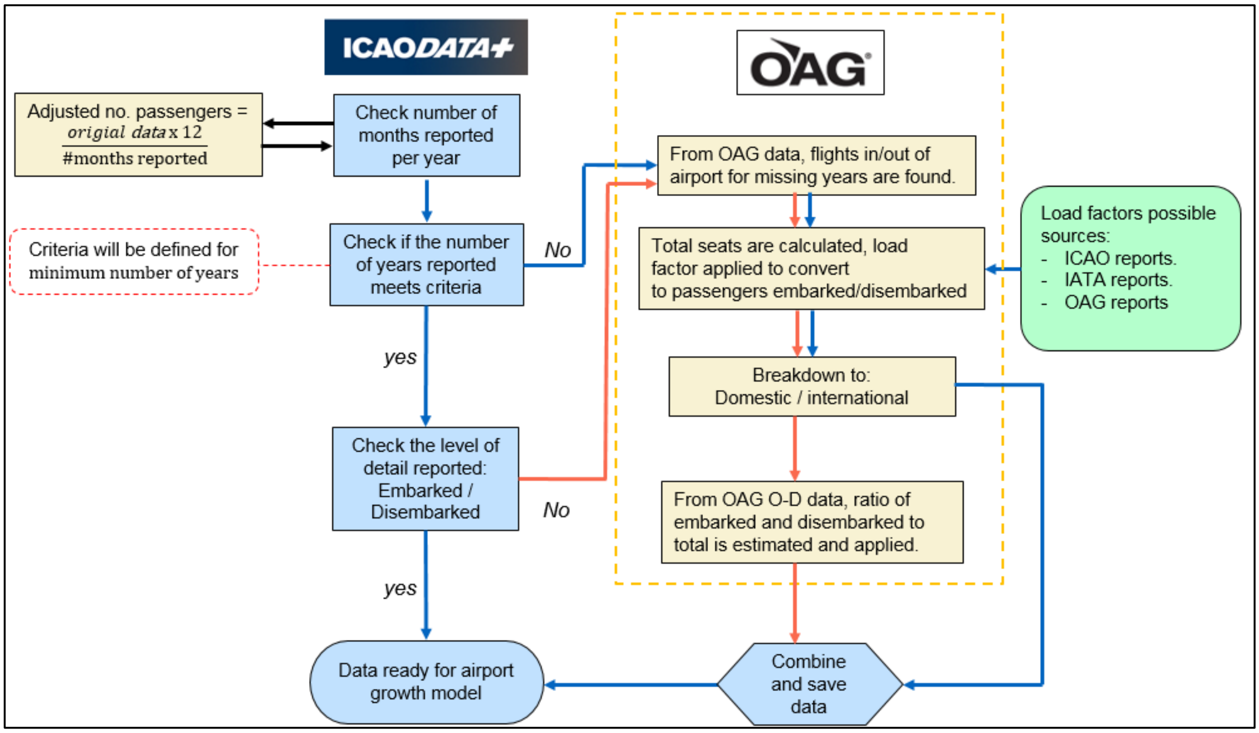

Figure 4: Use of OAG and ICAO passenger data sources to estimate demand for 3000 World airports

4 AIRPORT GROWTH MODEL

The third step of the model shown in Figure 1 is intended to forecast the future air traffic demand using airport level information. This model is called The Airport Growth Model (AGM). The Airport Growth Model is a statistical regression model that uses several socioeconomic variables to forecast the number of passengers at each airport in various World regions.

4.1 Model Variables

4.1.1 Dependent Variable

Historical data of passenger seats per flight is used as the model dependent variable. Data is extracted from the OAG dataset for 3,661 airports around the world. The dataset covers 10 years, from the year 2005 to the year 2014 for the majority of the airports. The Global Demand Model predicts the number of seats offered at the airport. The number of seats can be converted to number of passengers by applying load factors.

4.1.2 Explanatory Variables

Explanatory socioeconomic variables include population, GDP, and airline market share distribution. The model forecasts the number of seats offered at an airport, for that purpose all of the explanatory variables are related to the airport level activities. For example, population is interpreted as the population living within a catchment area around the airport (i.e., 60 nm radius). Similarly, GDP is the local GDP activity associated with a geographical region around the airport of interest.

The relationship between GDP and historical demand was observed by examining the historical trends in the data. Figure 5 shows an example of a linear relationship between the two variables. On the contrary, Figure 6 shows more scatter in the points, the behavior of the dependent variable in this type of cases is affected by regional or operational factors that cannot be captured by the economic growth trends only.

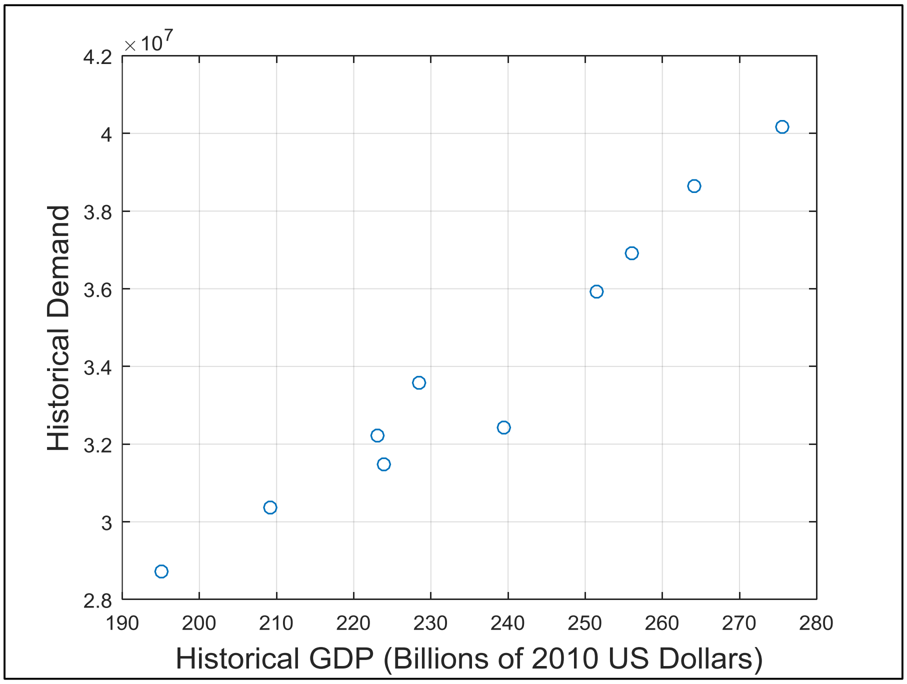

Figure 5: Scatterplot of the historical number of seats vs. the Local GDP (60 nm) for HKG

Figure 6: Scatterplot of the historical number of seats vs. the Local GDP (60 nm) for LHR

Population and GDP are two of the most commonly used variables in regression models (Shen 2006; Karlaftis et al. 1996). It is not recommended to use both variables in the model at the same time due to high correlation. Therefore, the two variables were combined in one variable. This new variable is calculated by dividing the GDP by the population in each cell in the grid to create a GDP per capita at the 2.5-minute cell resolution. To associate this at the airport level, the total GDP per capita was calculated for catchment areas 60, 90, and 120 nm radii around the airport and named Local GDP per capita. The Local GDP per Capita for 60, 90, and 120 nm was used in the regression analysis and the resulting R^2 values were very close. We used the 60 nm values in the model because using 90 or 120 nm is more time consuming and intensive data processing without significant improvements to the model results.

The identification of catchment areas was a time consuming process due to the intensive processing power required to perform it. Using the computer described in section 3.2.1, the MATLAB script that we developed takes 12.5 seconds to identify the cells in the grid that are located within a specified radius from an airport. This means that it takes 12.71 hours to run the script for all 3,661 airports in the OAG dataset.

Creating the population variable at the airport level was done by a MATLAB script that calculates the sum of population values within the catchment area by reading these values from the global grid. The global grid was created according to the procedure described in section 3.2.1 and using the same computer specifications. It took 0.8 seconds to calculate the population in the 60 nm catchment area for a single airport for one year. The total time required for creating the population variable for all airports from the year 2005 to the year 2040 would be 29.2 hours. Similarly, the time to calculate GDP within 60 nm catchment areas would be 29.2 hours as well. Therefore, multiple computers were used in parallel processing to reduce processing time.

The processing time increases drastically for larger catchment areas. The calculation of population or GDP for 90 and 120 nm catchment areas takes 1.8 and 3.4 seconds per airport for one year, respectively. That means that it would take 65 hours to calculate a variable within 90 nm for all airports from the year 2005 to the year 2040 and 124 hours (over 5 days) to calculate within 120 nm catchment areas.

Local GDP per capita is the main variable used in our model. Generally, GDP per capita has been used as a main predictor variable in global air passenger forecasts such as the International Air Transport Association (IATA) forecast (Pearce 2015). Figure 7 shows how GDP per capita is considered a major driver of the air travel demand according to IATA.

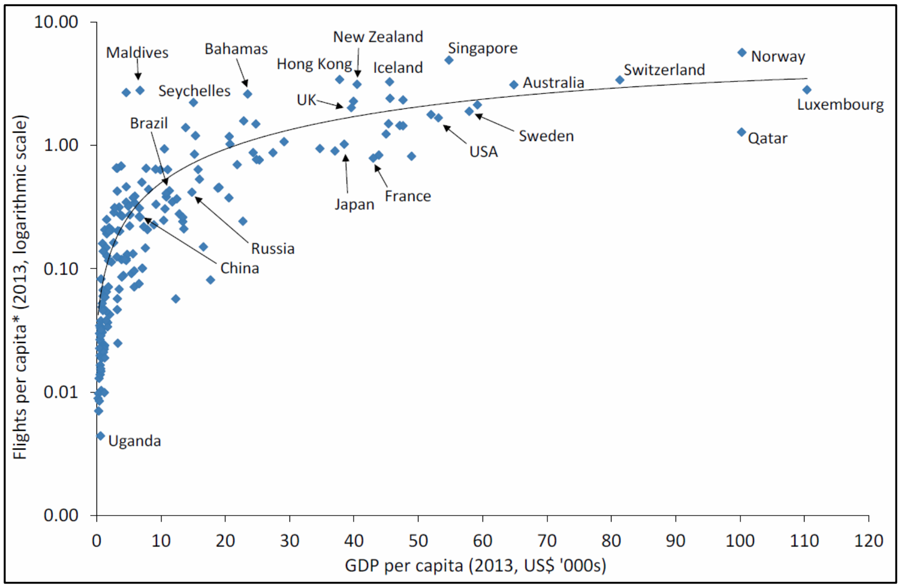

Figure 7: GDP per capita as a major driver of air travel (Pearce 2015)

Airline operations at an airport is an important factor that contributes to the passenger demand trends especially for airports in mature markets where the correlation between demand and economic growth alone is not enough to forecast demand. The relationship between market share changes and demand at an airport is presented in the model through the Herfindahl Index (H). The Herfindahl Index is a measure of the size of firms in relation to the industry and an indicator of the amount of competition among them (Gillen and Mantin 2009). In the Global Demand Model, it represents the competition between airlines operating at the airport. The value of (H) is calculated using Equation (1):

$$H = \sum_{i=1}^{N} s_i^2 \qquad (1)$$

Where:

H: Herfindahl Index (dimensionless)

s_i: Market share of airline (i) (fractional equivalent)

N: Total number of airlines operating at the airport

A high value of Herfindahl Index indicates less competition. For example, if one airline fully dominates a market, the value of H would be exactly 1.0. If two carriers serve a market with equal number of flights, the value of H would be 0.5.

4.2 Model Specification

The Airport Growth Model is a regression-based model. The model assumes that the operations at an airport are mainly driven by activities in the geographical region around the airport of interest. The relationship between variables is found through statistical analysis and is used to forecast future values of the dependent variable.

Linear regression is the most commonly used form in travel forecast models (Karlaftis et al. 1996) due to being simple and easy to apply for future projection. In our analysis, it was not practical to have a single model that fits the data well for all airports due to the global scope. The scope of the Global Demand Model covers airports with a variety of historical demand trends and includes multiple markets in the world with different socioeconomic characteristics. Figure 8 shows the trends for Hong Kong airport (HKG) which is an example of high correlation between the demand and the Local GDP per capita.

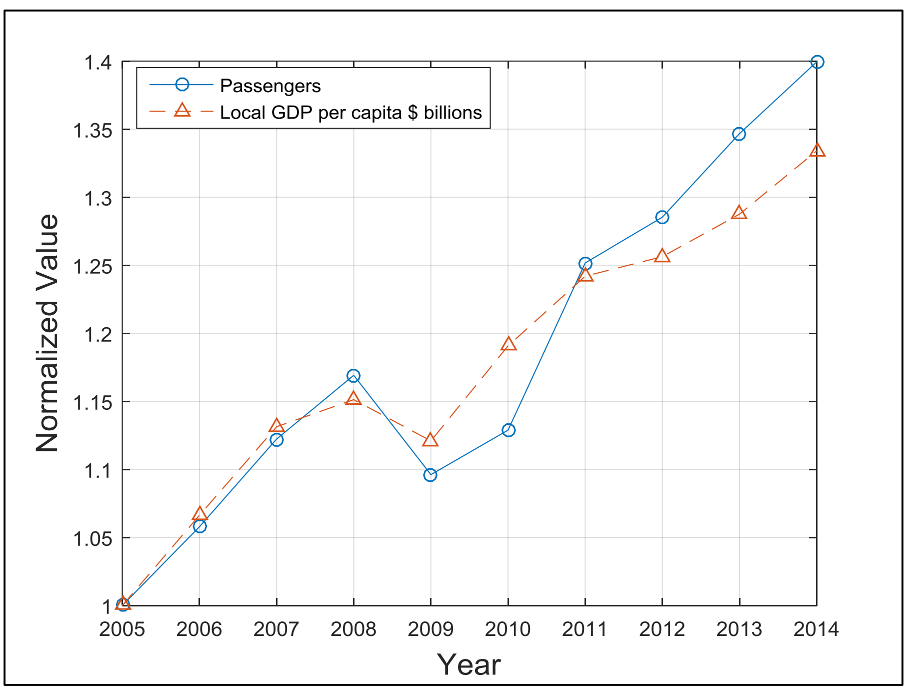

Figure 8: Historical demand and socioeconomic trends for Hong Kong airport (HKG)

On the other hand, for Cincinnati airport (CVG), Figure 9 shows an example of divergent trends that make it difficult for any model to find a relationship with a strong explanatory power. This last trend is attributed to unpredictable factors such as airline market decisions, political issues, financial crises, or other unforeseen factors. In this specific case of CVG, the airport used to be a hub for Delta Airlines until Delta decided to reduce the number of flights at CVG starting from the year 2005 due to financial reasons. This had a negative effect in the number of total enplanements at the airport.

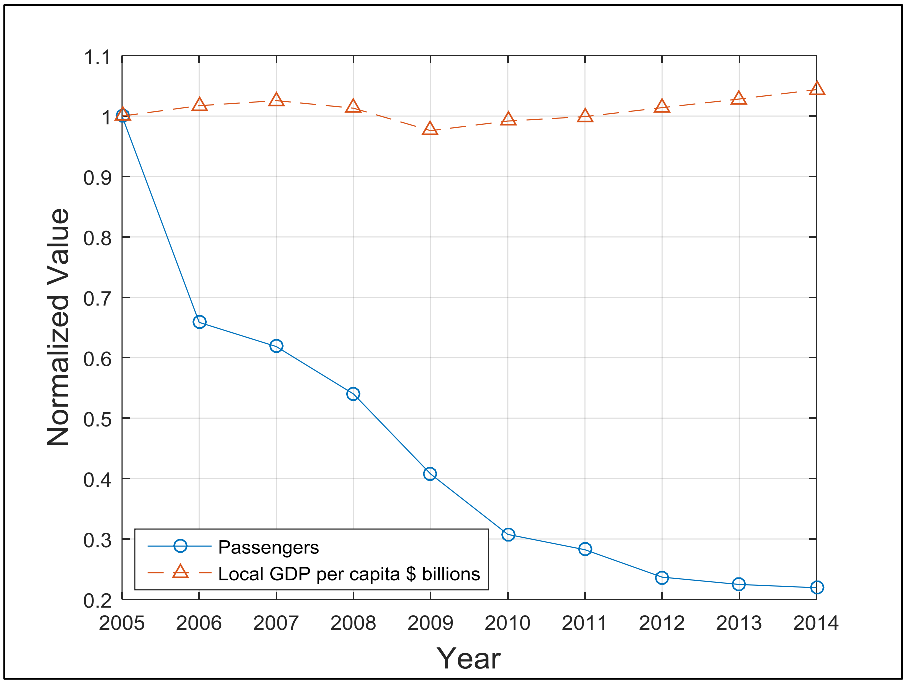

Figure 9: Historical demand and socioeconomic trends for Cincinnati airport (CVG)

The Global Demand Forecast Model consists of a group of linear and nonlinear sub-models as shown in Figure 10. The purpose of having a group of sub-models is to maximize the predictive power of the model using one or more of the available predictor variables mentioned in Section 4.1. Each airport in the analysis is individually assigned the model specification that best fits the historical data according to R^2 value as a measure of Goodness-of-Fit. Each of the sub-models is identified by a number, Table 2 lists the number corresponding to each sub-model.

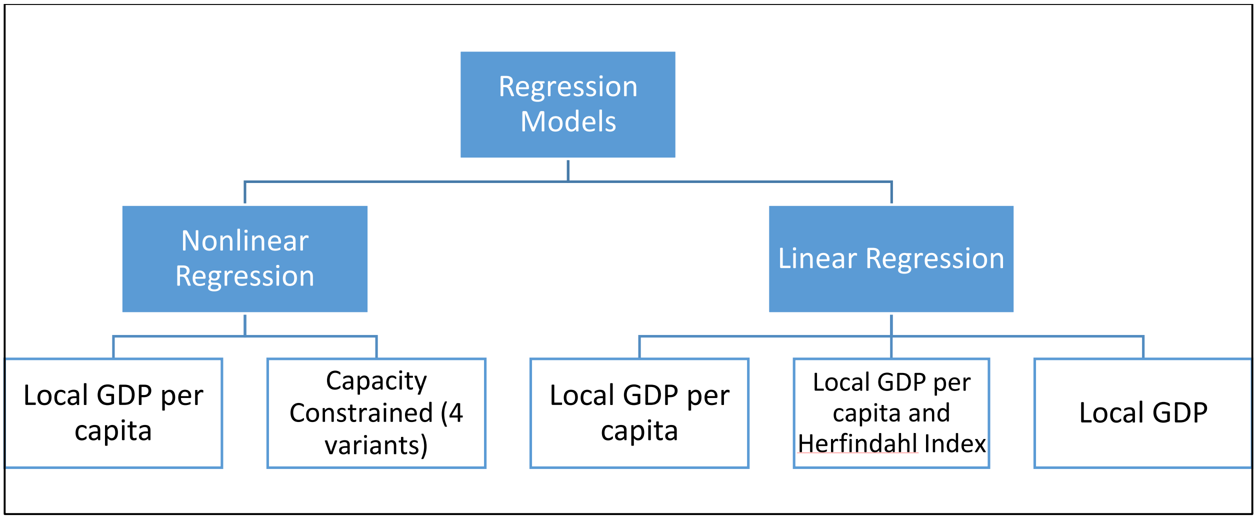

Figure 10: Regression models used to forecast airport growth

Table 2: List of sub-models in the Global Demand Model

Model ID	Description
1	Linear (Local GDP per capita)
2	Nonlinear (Local GDP per capita)
3	Linear (Local GDP per capita, H)
4	Capacity Constrained (10%)
5	Capacity Constrained (20%)
6	Capacity Constrained (30%)
7	Capacity Constrained (40%)
8	Linear (Local GDP)

Airports that have a minimum of 5 years of data and an increasing historical demand trend for at least the last 5 years of data are covered in the model. The minimum number of years of data was chosen to have a reasonable amount of data for a regression analysis taking into consideration that there are airports in the data set with a very low number of data years. The total number of airports that pass the above-mentioned criteria is 1,875 out of 3,661 airports.

4.3 Linear Regression Models

4.3.1 Linear Regression with Local GDP per capita

This model uses the local GDP per capita as the explanatory variable described earlier in Section 4.1. We use the following model specification:

$$Y = a + b * GDP \quad (2)$$

Where:

Y: Number of seats offered (seats)

a: Intercept

b: Coefficient

GDP: GDP per capita in the region bounded by a 60nm radius around airport (US$ billions)

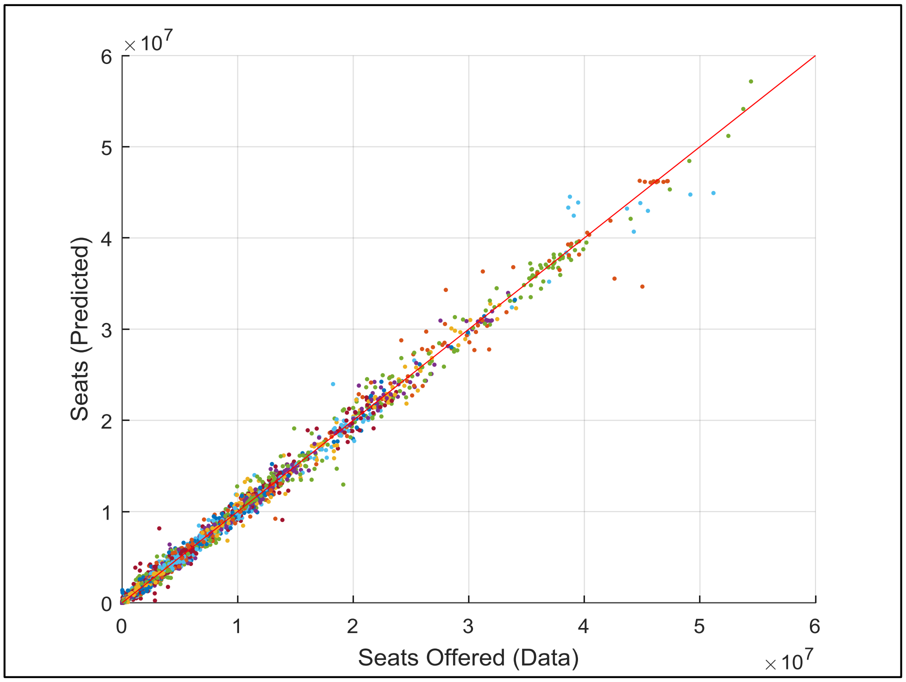

Figure 11: Linear model results (GDP)

Results of this model are compared to the historical data and plotted against the 45-degree line. Figure 11 shows the results for all 1,875 airports. It can be observed that the majority of the points plotted are close to the 45-degree line except for some points where the model overestimated or underestimated the demand by a relatively large amount.

The coefficient of determination, R^2, is selected as the statistical measure of Goodness-of-Fit and was calculated for all airports. Figure 12 shows the distribution of R^2 values for this model specification. The figure shows that many airports have low values of R^2 below 0.5. Section 5 contains more analysis of the results and Goodness-of-Fit.

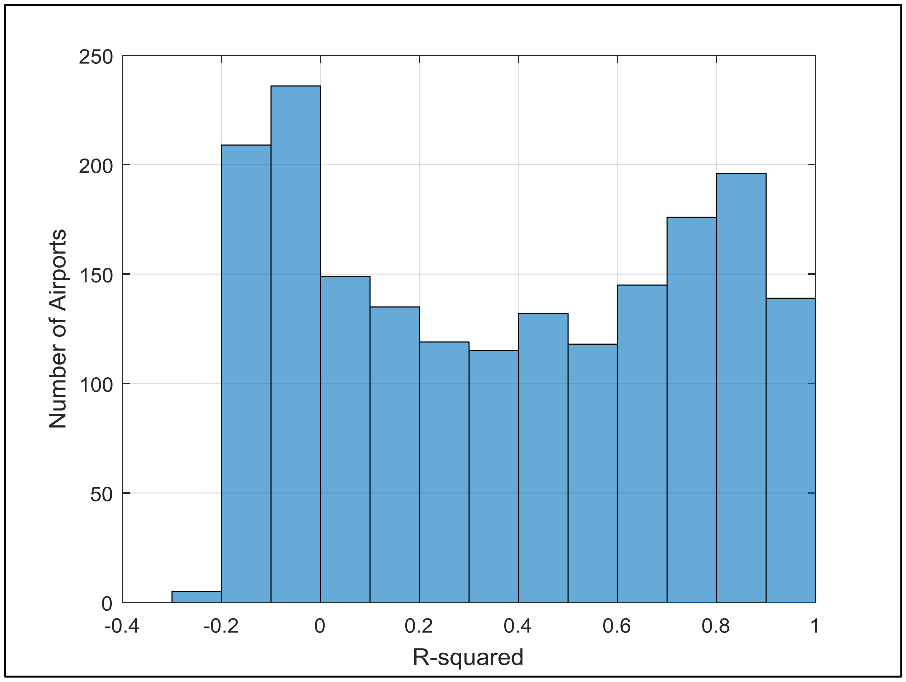

Figure 12: R-squared distribution for Linear model (GDP per capita)

4.3.2 Linear Regression with Local GDP per capita and Herfindahl Index

The Herfindahl Index described in Section 4.1 was added to the model in an attempt to improve the results by capturing the effect of the airline competition at each airport. The model specification is as follows:

$$Y = a + b * GDP + d * H \qquad (3)$$

Where:

Y: Number of seats offered (seats)

a: Intercept

b, d: Coefficients

GDP: GDP per capita in the region bounded by a 60nm radius around airport (US$ billions)

H: Herfindahl Index

The results of model specified in Equation *(3)* are compared to the historical data. Figure 13 shows a plot of the data versus the predicted values using a 45-degree line for comparison. It can be seen that the majority of the points plotted are close to the 45-degree line except for some points where the model overestimated or underestimated the demand by a relatively large amount.

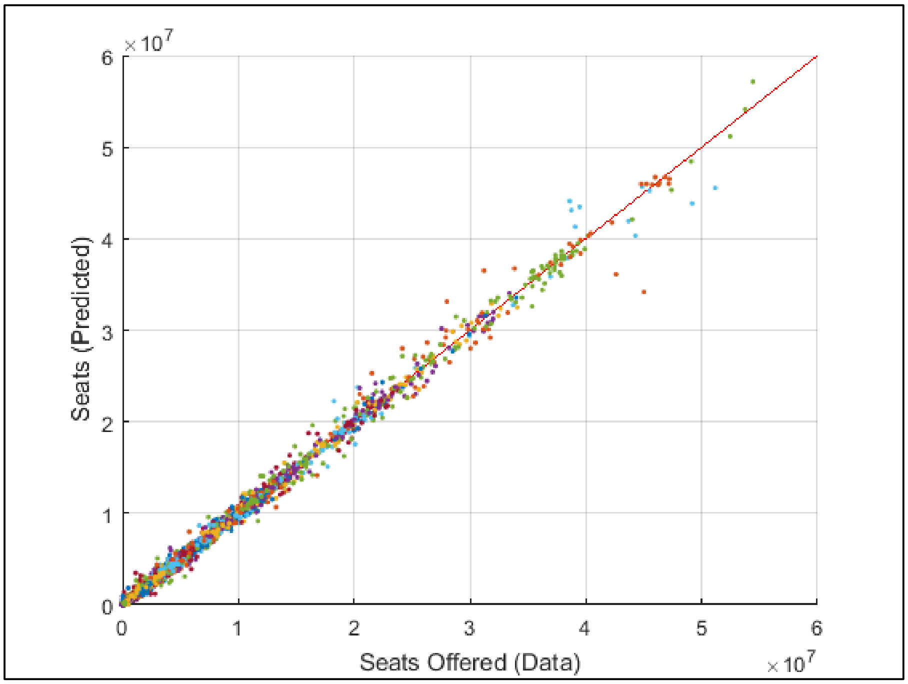

Figure 13: Linear model results (GDP, H)

Figure 14 shows the distribution of R^2 values for the new model specification. The figure shows that many airports have low values of R^2 below 0.5. However, there is an improvement compared to the model presented in Section 4.3.1. Figure 14 shows an improvement in the number of airports with high R^2 values with the introduction of Herfindahl Index (H). This helps increase the explanatory power of the model beyond using the Local GDP Per Capita as the only explanatory variable. Section 5 contains more analysis of the results and Goodness-of-Fit.

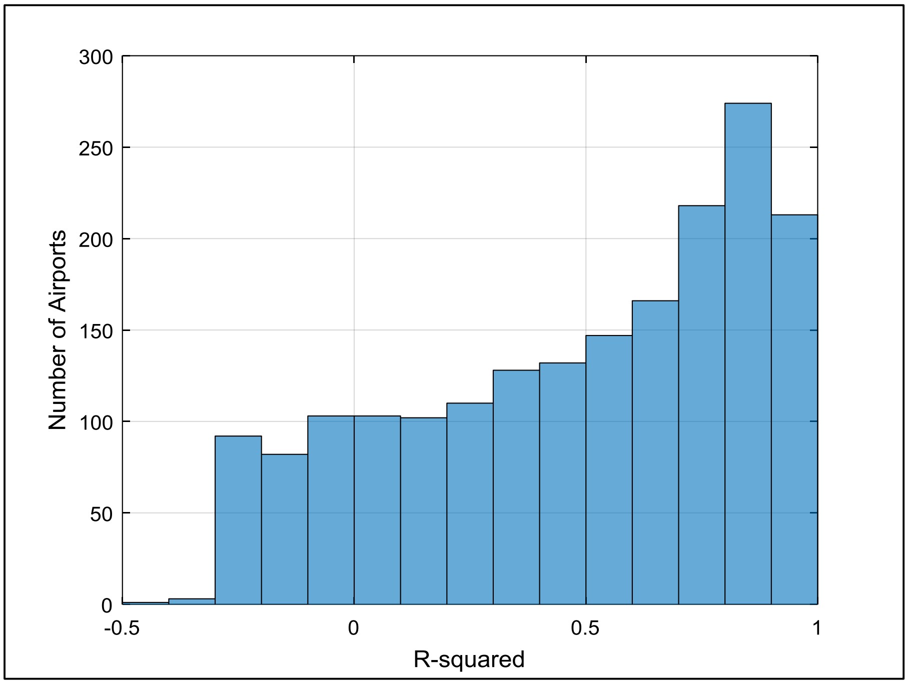

Figure 14: R-squared distribution for linear model (GDP, H)

4.3.3 Linear Regression with Local GDP

This model uses the local GDP around the airport of interest as the explanatory variable. This model is considered the simplest version of all sub-models included in the Global Demand Model. In our analysis, we used the following model specification:

$$Y = a + b * GDP \tag{4}$$

Where:

Y: Number of seats offered (seats)

a: Intercept

b: Coefficient

GDP: GDP in the region bounded by a 60nm radius around airport (US$ billions)

The results of this model specification were plotted compared to the historical data. Figure 15 shows the plotted points against the 45-degree line to present the overall Goodness-of-Fit for this model.

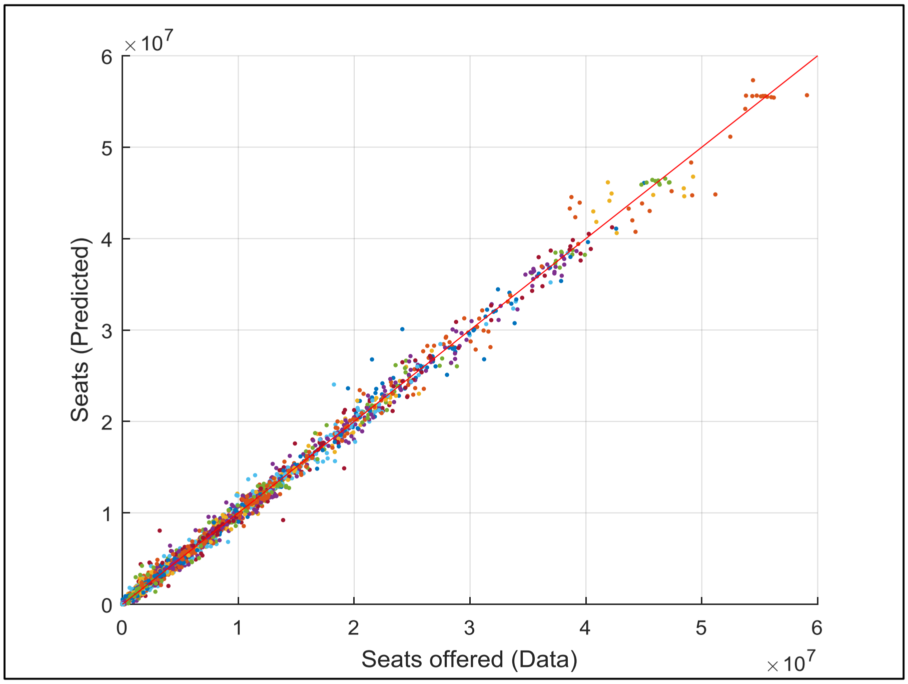

Figure 15: Linear model results (GDP)

Similar to the other models presented previously in Section 4.3, the values of R^2 resulted from applying this model varied from low to very high values. The distribution of R^2 values is shown in Figure 16.

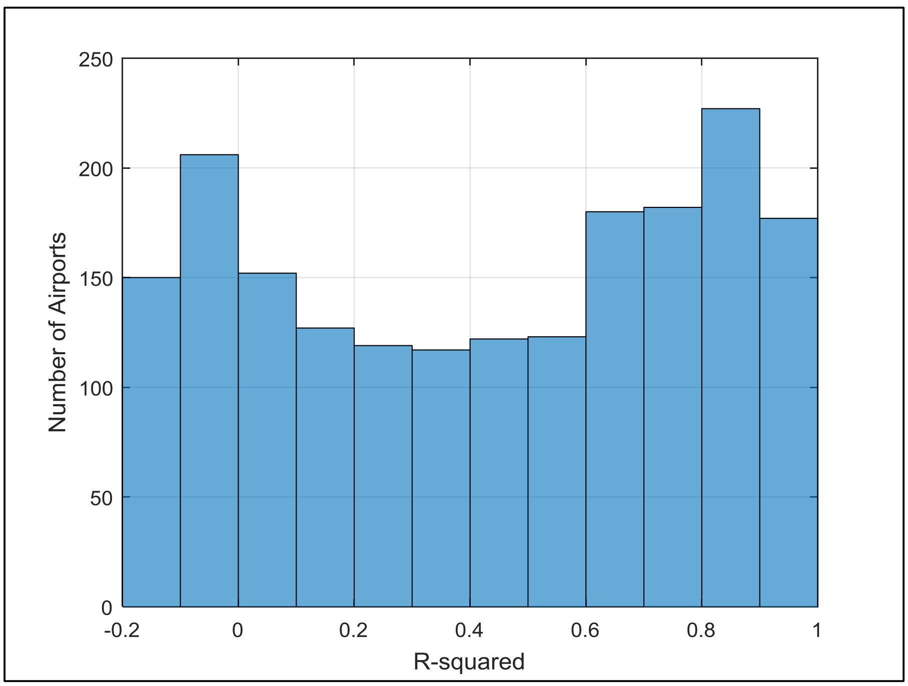

Figure 16: R-squared distribution for linear model (GDP)

4.4 Nonlinear Regression Models

4.4.1 Nonlinear Regression with Local GDP per capita

This model uses the local GDP per capita as the explanatory variable. Equation (5) shows the model specification:

$$Y = a + b * (GDP)^d \qquad (5)$$

Where:

Y: Number of seats offered (seats)

a: Intercept

b, d: Coefficients

GDP: GDP per capita in the region bounded by a 60nm radius around airport (US$ billions)

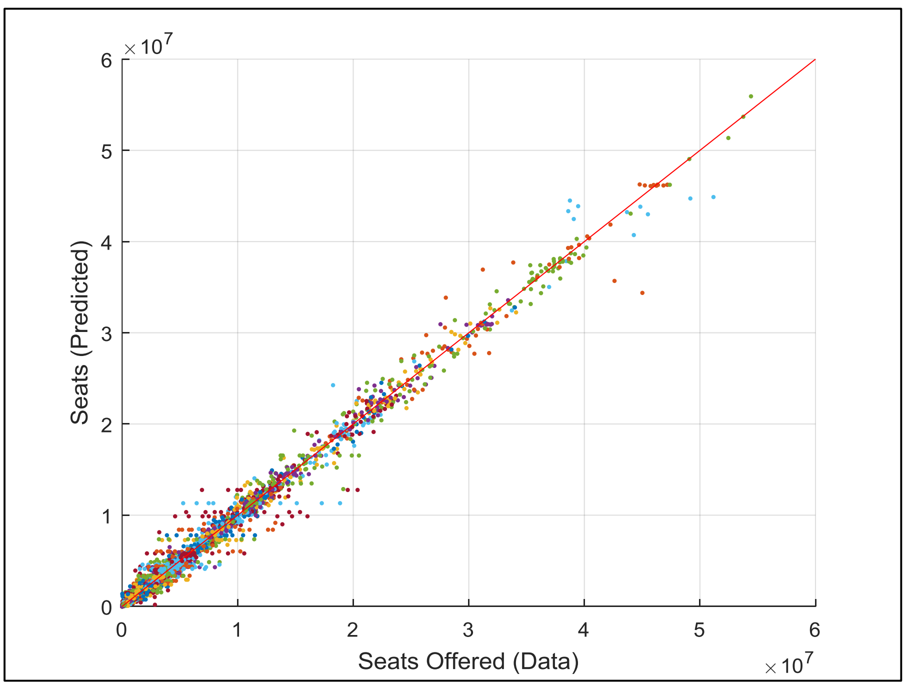

Figure 17: Nonlinear model results (GDP)

Figure 17 shows the predicted values produced by the model and the historical data points plotted with the 45-degree line for comparison. The majority of the airports modeled yield points close to the line. However, the model overestimated or underestimated the demand by relatively large amount for some airports. The distribution of R^2 values using this model is shown in Figure 18. The results show a modest improvement to the explanatory power of the model compared to the linear specification using Equation *(2)* with Local GDP Per Capita.

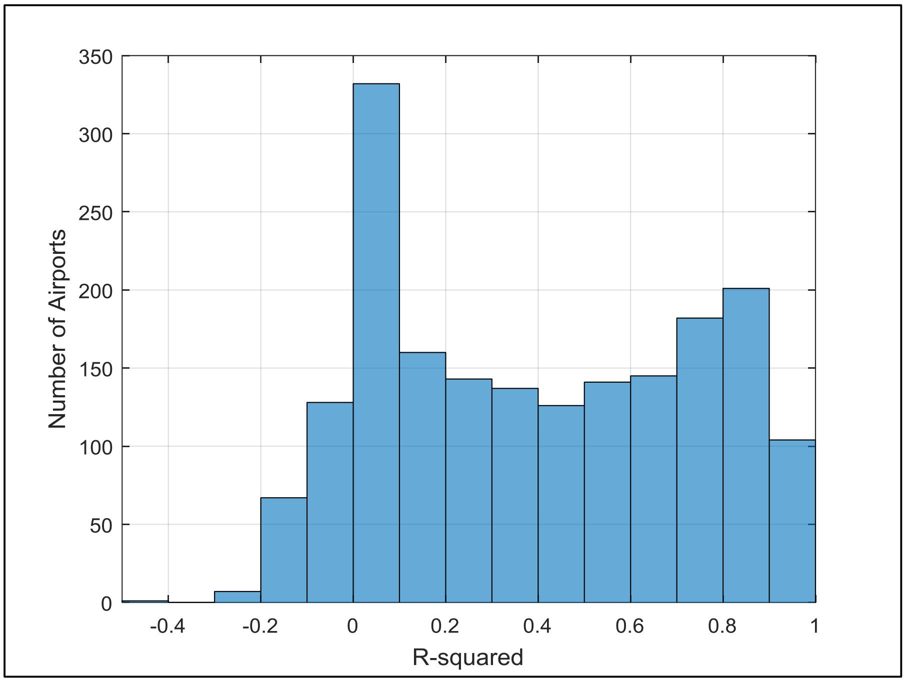

Figure 18: R-squared distribution for nonlinear model (GDP)

4.4.2 Nonlinear Regression with Local GDP per capita (Capacity-Constrained)

All of the models presented so far predict the future demand under the assumption of unrestricted growth. For some airports, this assumption can be realistic because they operate at low volumes compared to their estimated throughput capacities. However, in mature markets where some airports operate at traffic volumes close to their estimated throughput capacities, it is more realistic to introduce a capacity limitation into the model in an attempt to explain the relationship between historical demand trends and socioeconomic variables. This model is expected to have reasonable long-term projection.

The capacity-constrained model is based on the Logistic Function. There are many applications of the logistic equation in modeling where growth limitations are considered, population and market growth are two examples. The simple Logistic function is defined by Equation (6) (Wooldridge 2012). Figure 19 shows an example of the standard “S” shape curve of the logistic function with curve’s maximum value $L = 1$.

$$Y = \frac{L}{1 + e^{-x}} \qquad (6)$$

Where:

Y: The dependent variable

L: The curve's maximum limit

x: The independent variable

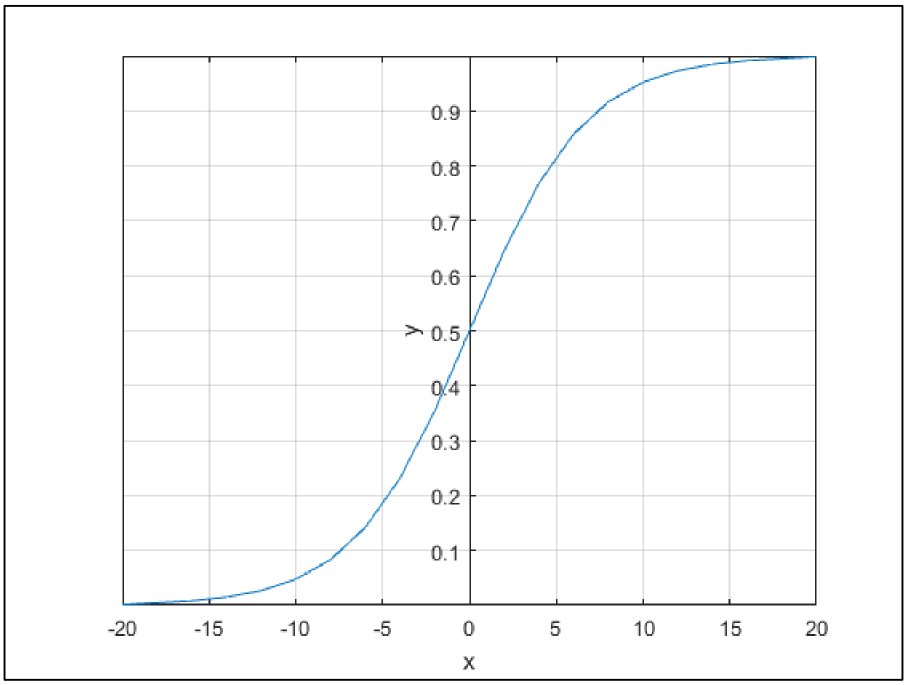

Figure 19: The logistic function

In our model, Equation *(7)* shows the general capacity-constrained model specification. It is based on the logistic equation with the term C as the saturation capacity of the airport.

$$Y = \frac{C}{1 + b * e^{-d*GDP}} \qquad (7)$$

Where:

Y: Number of seats offered (seats)

C: Estimated capacity of the airport (passengers)

b, d: Coefficients

GDP: GDP per capita in the region bounded by a 60nm radius around airport (US$ billions)

A crude method to estimate the airport capacity (C) in Equation (7) was used. The capacity is assumed to range from 10% to 40% above the highest value of passenger demand in the last 5 years of data. This provided four capacity-constrained model variants with increments of 10%.

The four variants of the capacity-constrained model were executed for all World airports. Figure 20 shows the results of all variants plotted against the data with a 45-degree line for comparison.

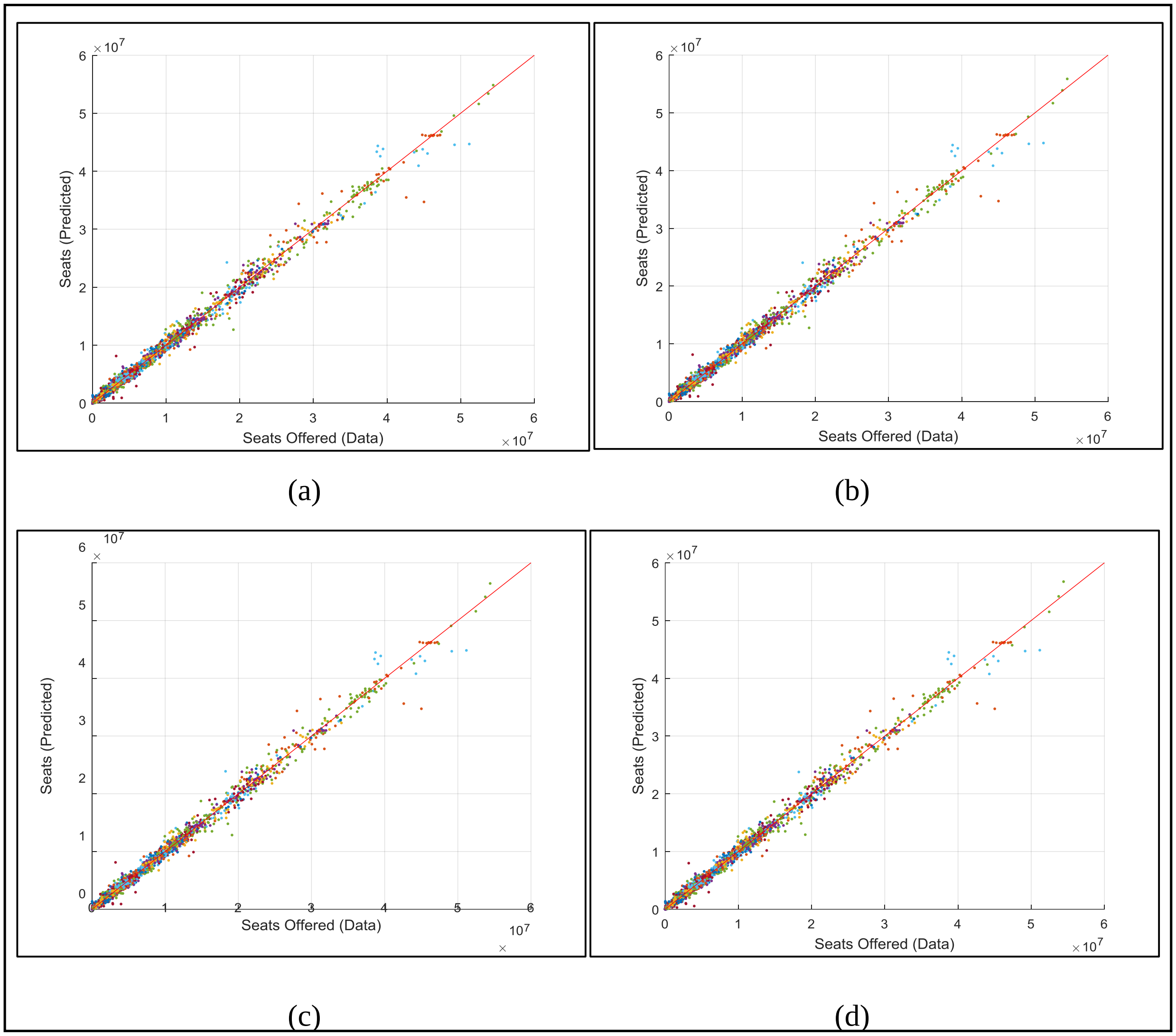

Figure 20: Capacity-constrained model results (a) 10% (b) 20% (c) 30% (d) 40%

Figure 21 presents the distribution of R^2 for the four variants of the capacity-constrained model. Having multiple variants is intended to improve the model results by achieving higher R^2 values. The utilization of each model variant is presented by how frequent it was selected as the best fitting model for airports in the analysis. This is presented in Section 5.1.

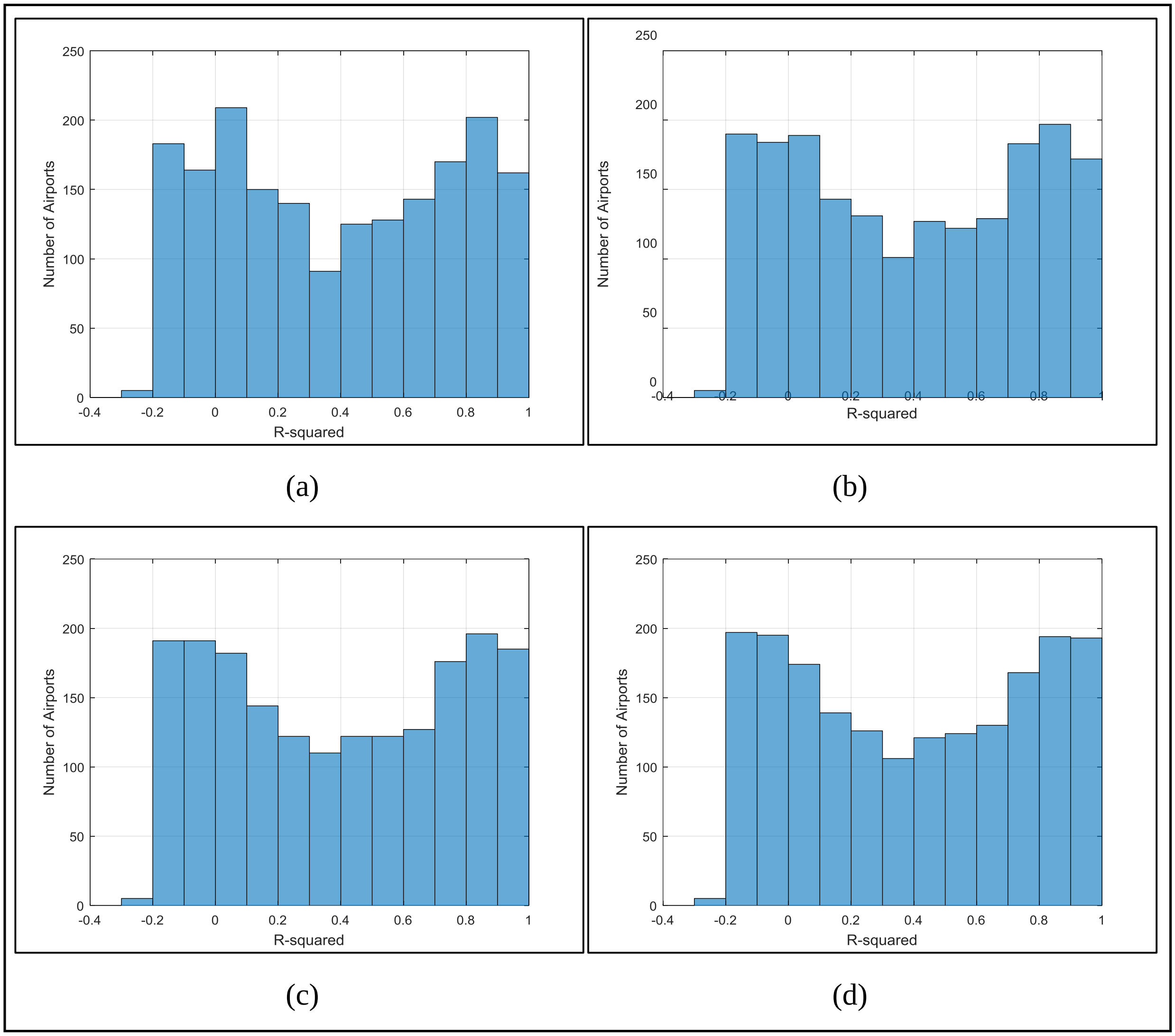

Figure 21: R-squared distribution for capacity-constrained model (a) 10% (b) 20% (c) 30% (d) 40%

4.5 Model Selection

The methodology in selecting the best model for each airport starts by executing all models shown in Figure 10 for all airports in the dataset. The results of each model are compared in order to select the best model that provides the highest correlation for each airport individually. The model selection decision is made based on selecting the highest R^2 value among all results of the group of models shown in Figure 10.

As a sanity check, the future projection resulted from the selected model for each airport was examined to avoid negative future trends. This check was necessary because for some airports the historical trends of the local GDP per capita do not reflect the actual economic trends. This issue resulted from the fact that the available data is limited to the period between the year 2005 and the year 2014 where some economies suffered from economic crisis at different levels, and that creates an issue since population continued to growth at a faster rate than GDP for a number of years. For example, Dubai International airport (DXB) population and Local GDP trends are shown in Figures 22 and 23, respectively. When the Local GDP per capita is used, the trend is shown in Figure 24 and producing an inaccurate model.

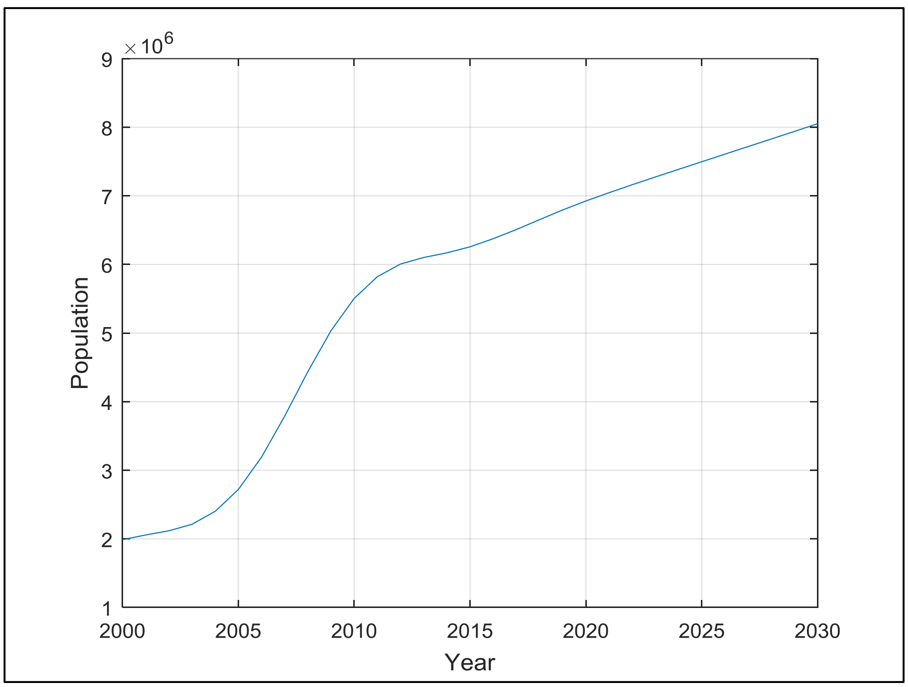

Figure 22: Population 60 nm around DXB

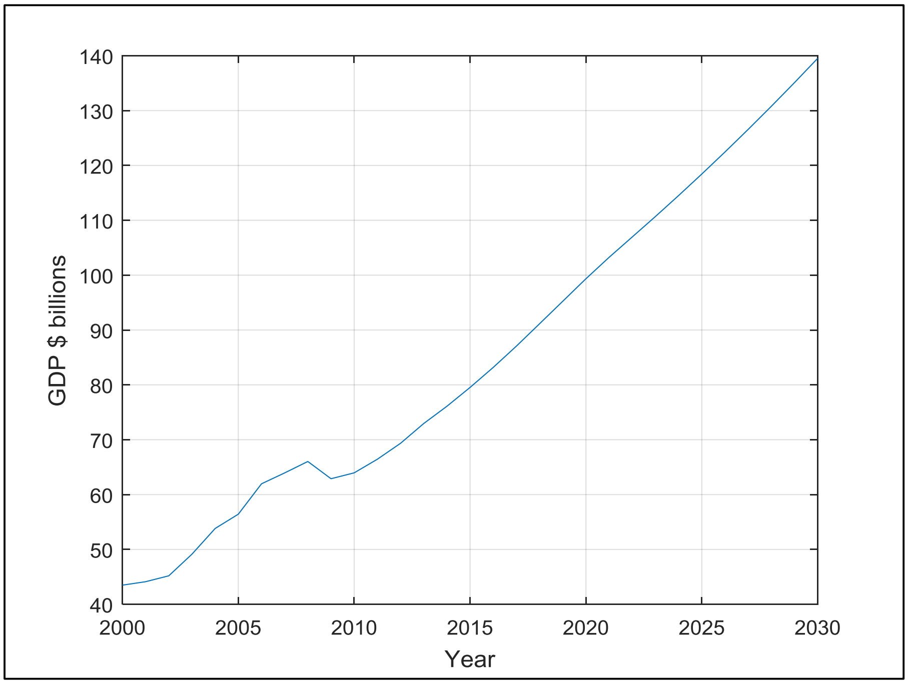

Figure 23: GDP 60 nm around DXB

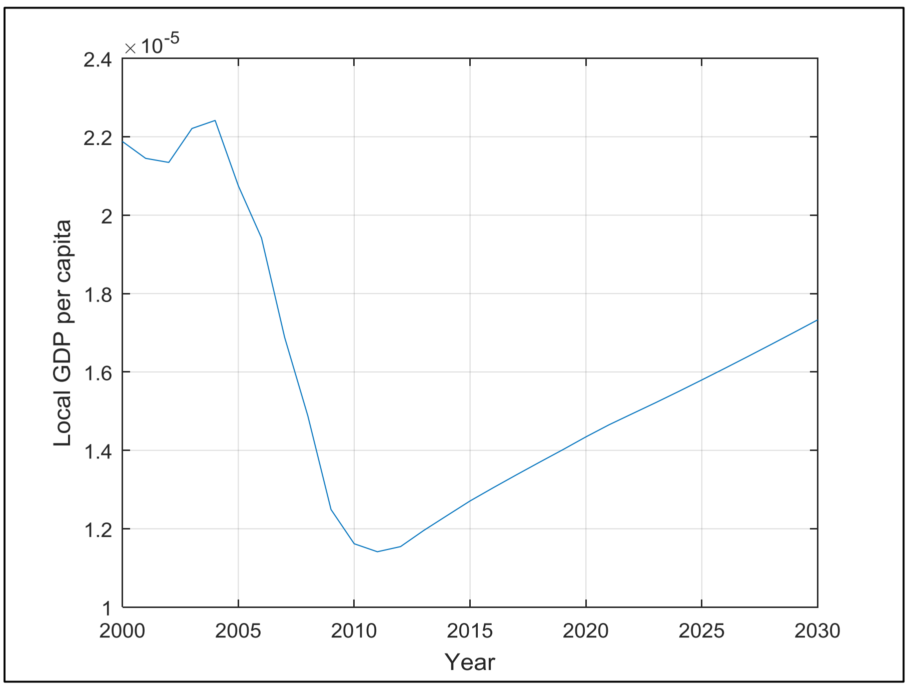

Figure 24: Local GDP per capita 60 nm around DXB

In order to overcome this problem, the model selection algorithm was modified to identify airports with decreasing future projections and reassign a different model. The selected new model is the (Local GDP) model that avoids the misleading trends for airports with cases similar to the described above by using the economic growth (GDP) only as the predictor variable.

Figure 25 shows the result of the (Local GDP) linear regression model for Dubai International airport (DXB). The Model Selection procedure takes into consideration that the highest R^2 does not necessarily provide the most accurate forecasting model all the time (Karlaftis et al. 1996).

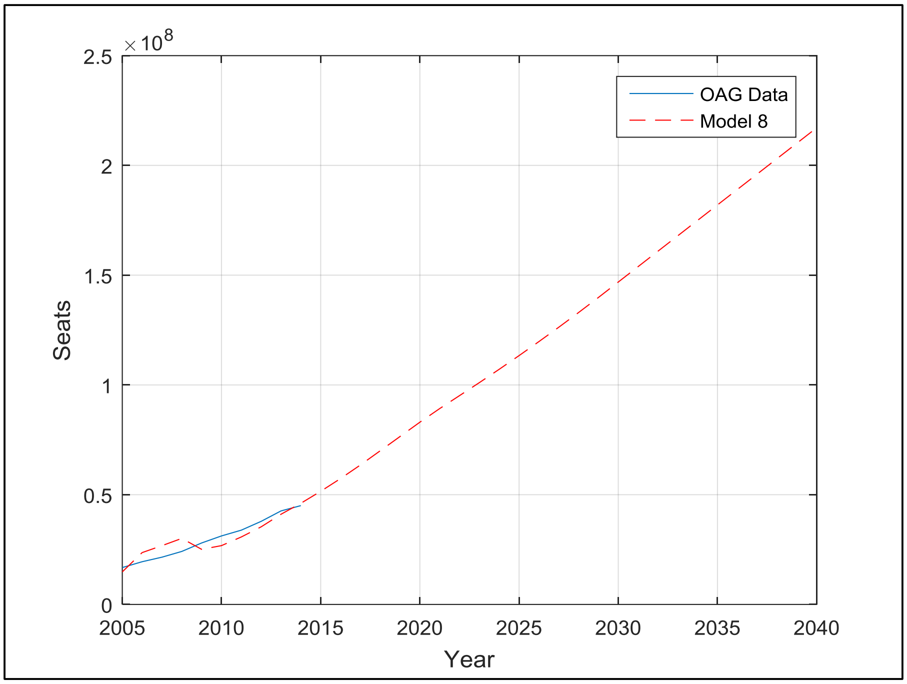

Figure 25: Regression Model (Local GDP) results for DXB airport

For airports having a decreasing demand trend over time, similar to the trend shown in , the modeling procedure varies because a regression model using the same explanatory variables will produce negative numbers over time which is not a realistic outcome for most airports showing this behavior. The logic of modeling for this group of airports is described in Section 4.5.

4.6 Forecast for Airports with Decreasing Trends

This group of airports consists of 1,142 airports, equivalent to 31% of the airports in the dataset studied. Figure 26 shows the breakdown of airports by historical demand trend. For airports with decreasing trends, Using the regression models with the socioeconomic variables mentioned in Section 4.1 results in negative coefficients that lead to unrealistic results.

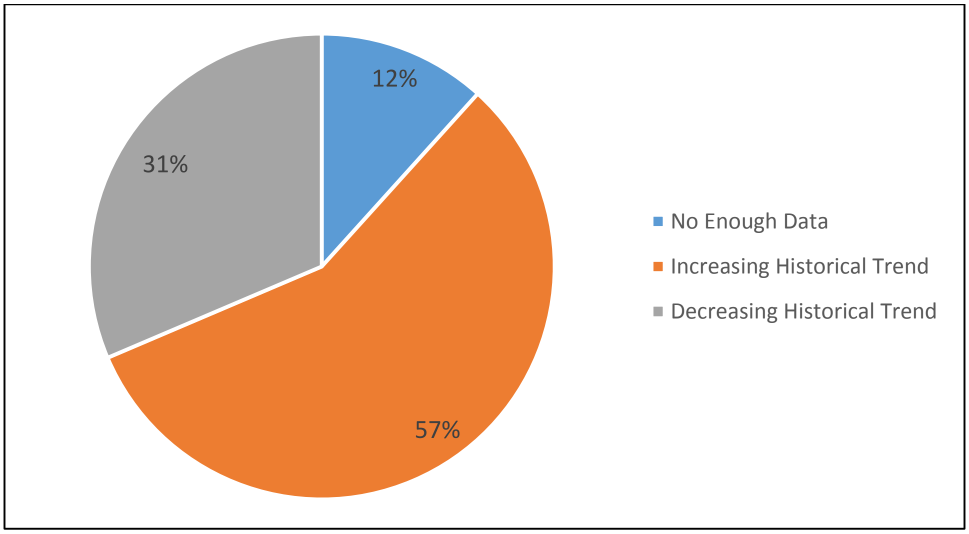

Figure 26: Breakdown of airports by historical demand trend

In this project, it is assumed that each airport will recover and start growing in the future. This approach is based on the long-term assumption that economic growth will drive more demand in the future. The Federal Aviation Administration (FAA) Terminal Area Forecast (TAF) (FAA 2013) uses a similar approach. Figure 27 shows an example of the TAF forecast for Cincinnati/Northern Kentucky International Airport (CVG).

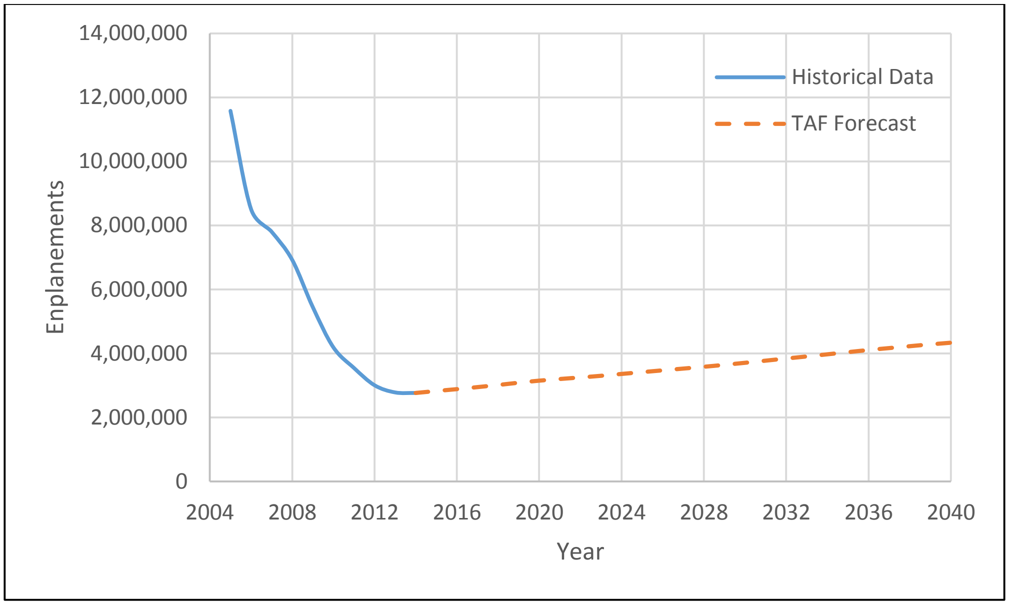

Figure 27: TAF forecast for Cincinnati/Northern Kentucky International Airport (CVG)(FAA 2013)

The TAF values of future growth rates are derived from the Table 3, the Annual Compound Growth for an airport is selected based on the forecasted growth rates of airports with similar size.

Table 3: Summary of enplanements according to Table S5 in Terminal Area Forecast 2013 (FAA 2013)

Enplanements at Towered Airports

	Count	2012	2040	Annual Compound Growth Rate 2012-2040
Large Hubs	29	516,724,328	872,685,971	1.88
Medium Hubs	34	126,633,945	207,105,899	1.77
Small Hubs	77	63,871,759	96,428,368	1.48
Non Hub Towers	375	19,225,942	32,253,515	1.86
Total	515	726,455,974	1,208,473,453	1.83

According to the TAF model and report, an airport is considered a large hub if it serves a total of 1% or more of total passenger enplanements in the U.S. A medium hub airport serves between 0.25% and 0.99%, a small hub from 0.05% to less than 0.25% and a non-hub airport enplanes less than 0.05% of total U.S passengers (FAA 2013).

In our model, Annual Compound Growth rates are applied to airports by geographic region. A default value of 1.8% is set in the model. This value represents a suggested modest growth rate. However, the user can change the value for any of the 17 World regions mentioned in Table 1 in Section 3.2.3.

5 GLOBAL DEMAND MODEL RESULTS AND ANALYSIS

5.1 Model Selection Results

In order to achieve the best results, all of the sub-models mentioned in Sections 0 and 4.4 were executed for the 1,875 airports and the model specification for each airport was individually assigned according to procedure in Section 4.5. Figure 28 shows the frequency at which each model specification was selected as the best fitting model before the addition of Model 8. All Models are identified by numbers listed in Table 2.

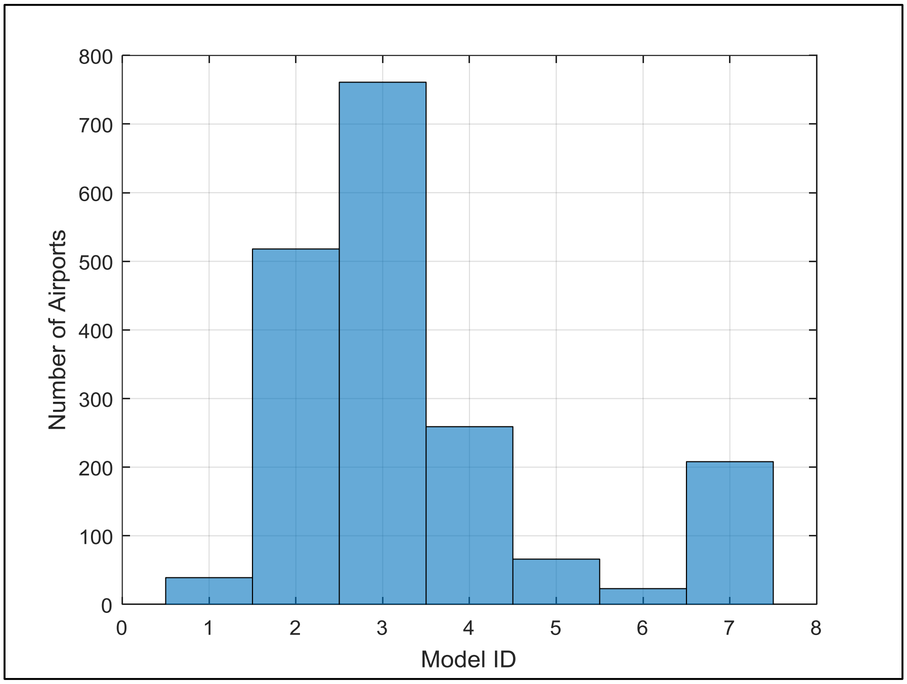

Figure 28: Model assignment before adding Model 8

Figure 28 shows that the most frequently assigned models were Model 3 with 41% followed by Model 2 with 32% and Model 7 with 12% of all modeled airports. The analysis was executed again after adding Model 8 to the group of sub-models. The results of the modified model selection algorithm are shown in 29.

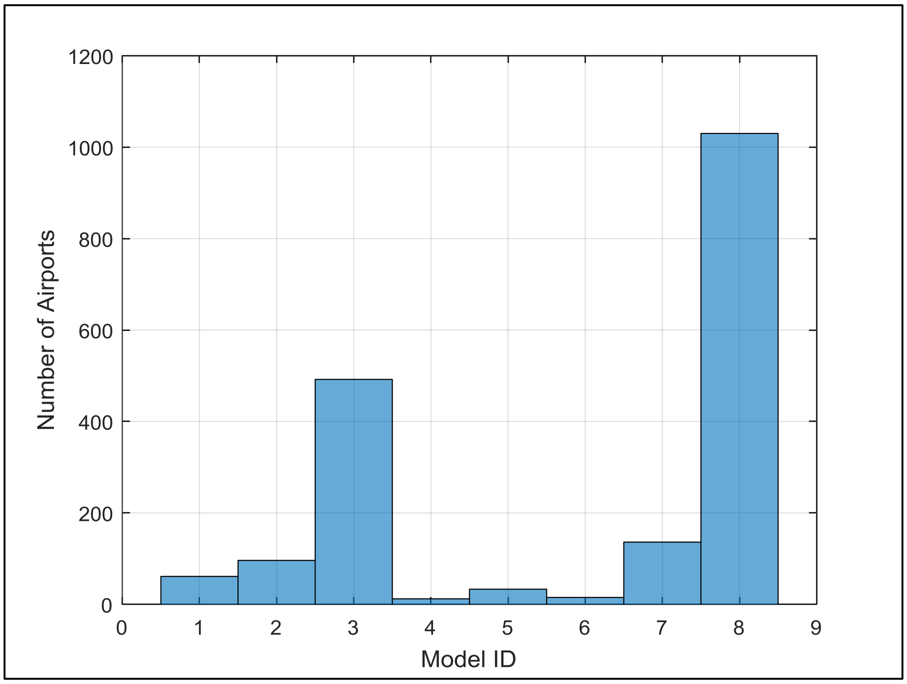

Figure 29: Model Assignment after adding Model 8

29 shows that the top three assigned models were Model 8 with 55%, Model 3 with 26%, and Model 7 with 7% of all modeled airports. In other words, the simplest model specification with Local GDP as the main demand driver has the largest share of the total airports in the analysis. The GDP is considered a key predictor variable according to previous similar analyses (Marazzo et al. 2010; Gillen 2009). When the results in Figures 28 and 29 are compared, it is evident that the majority of airports that initially used Model 2 migrated to Model 8. Airports from Models 2,3, and 4 make up the majority of airports that were assigned Model 8, The breakdown of airports by their initial model assignment is shown in Figure 30.

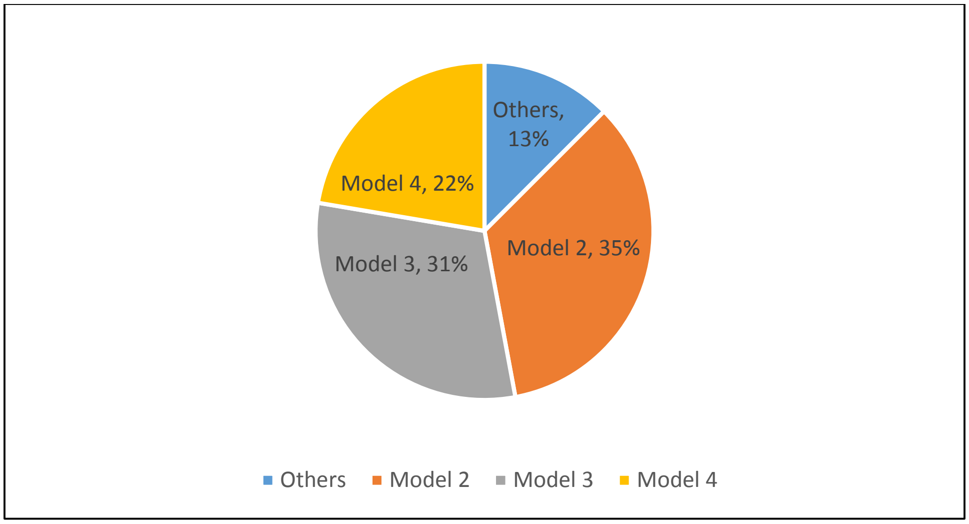

Figure 30: Initial model assignment for airports that switched to Model 8

Countries around the world showed different model assignment trends. For example, the model assignment distribution for airports in the United States is shown in Figure 31. The trend of the model selection in the United States is similar the global trend shown in Figure 29, the most frequently selected models are models 8 and 3 in that order.

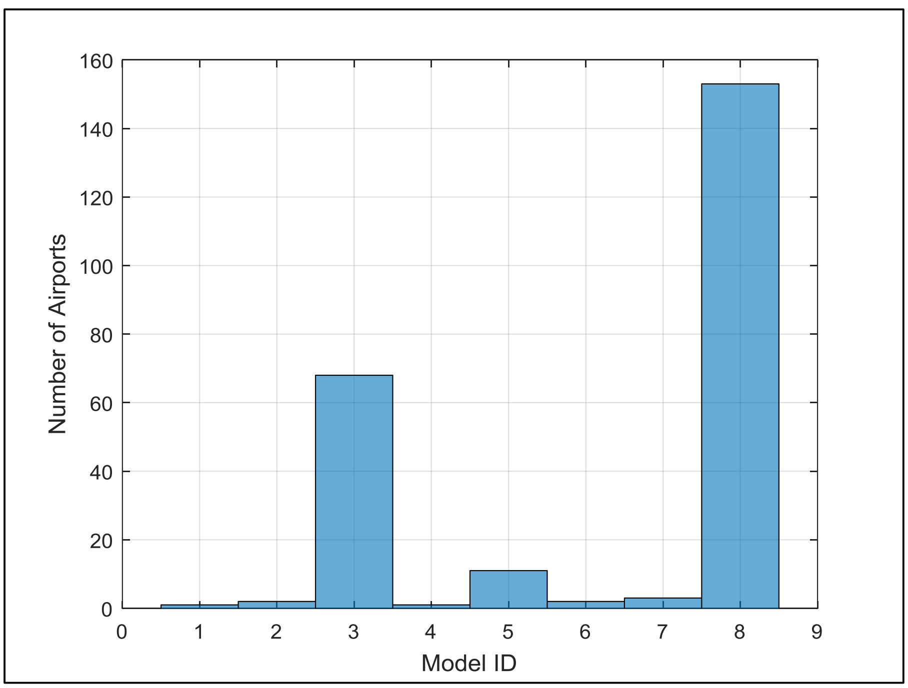

Figure 31: Model assignment distribution for the United States

Another trend example is China, an economy with a recent aggressive growth. Figure 32 shows that the most frequently selected model in China was model 7, the capacity-constrained (40%), which is different from the global results shown in Figure 29.

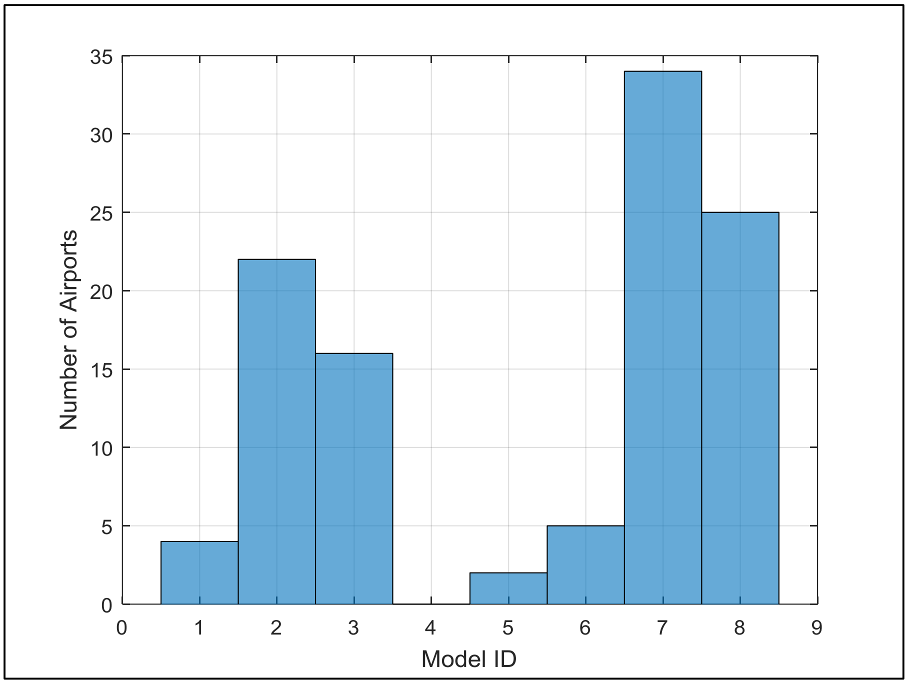

Figure 32: Model assignment distribution for China

A possible explanation to the contrast in trends between the United States and China is that both economic and demand growth rates were aggressive in China. The steep growth trend makes the capacity-constrained models fit the data better because this behavior matches the general shape of the logistic function shown in Figure 19. It is reasonable to select the capacity-constrained models for the aggressively growing trends because it helps avoid overestimation for the long term forecast. Figure 33 shows the forecast produced by model 7 for Changchun Longjia International Airport (CGQ) in China.

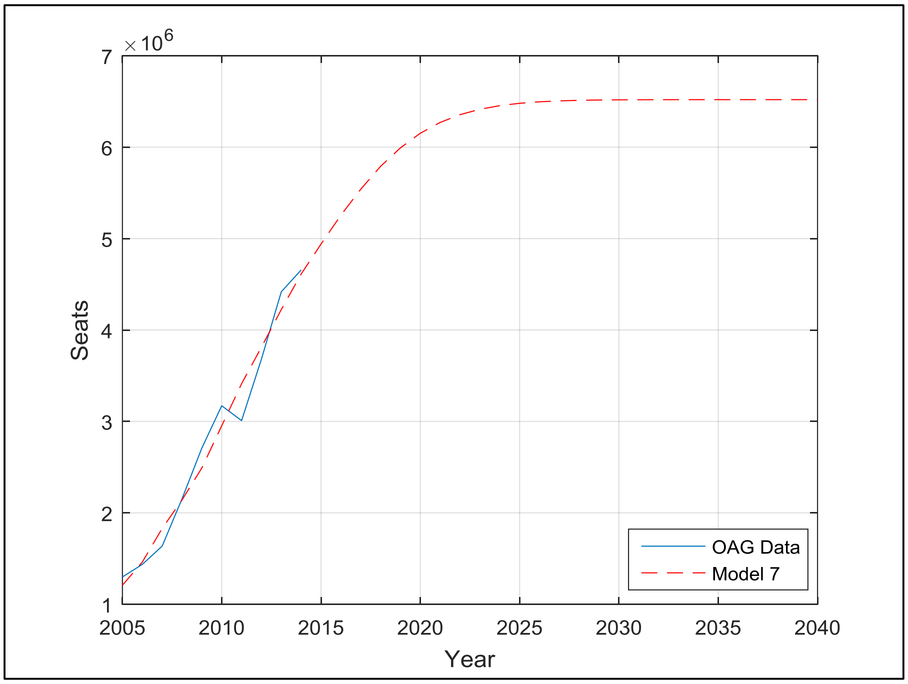

Figure 33: Forecast for CGQ

5.2 Goodness-of-Fit

The results of all models combined are plotted to show the overall Goodness-of-Fit for the model. Figure 34 shows the Data versus the predicted number of seats plotted against the 45-degree line.

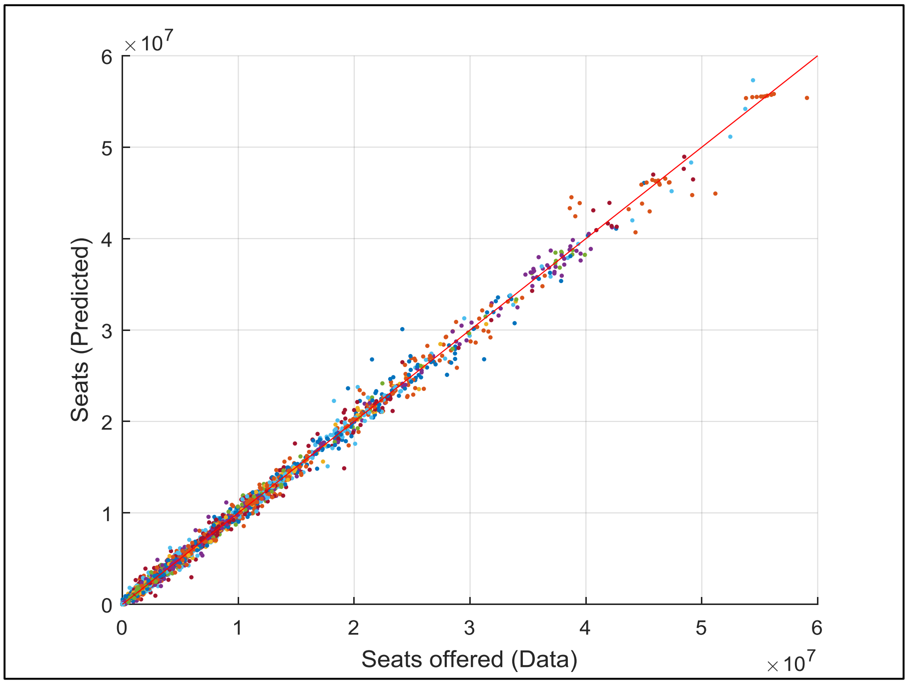

Figure 34: Combined Model Results

The R^2 results of model selection algorithm are presented in Figure 35. Although there is a number of airports with low R^2 values, the overall distribution has improved compared to any individual sub-model presented in Sections 0 and 4.4.

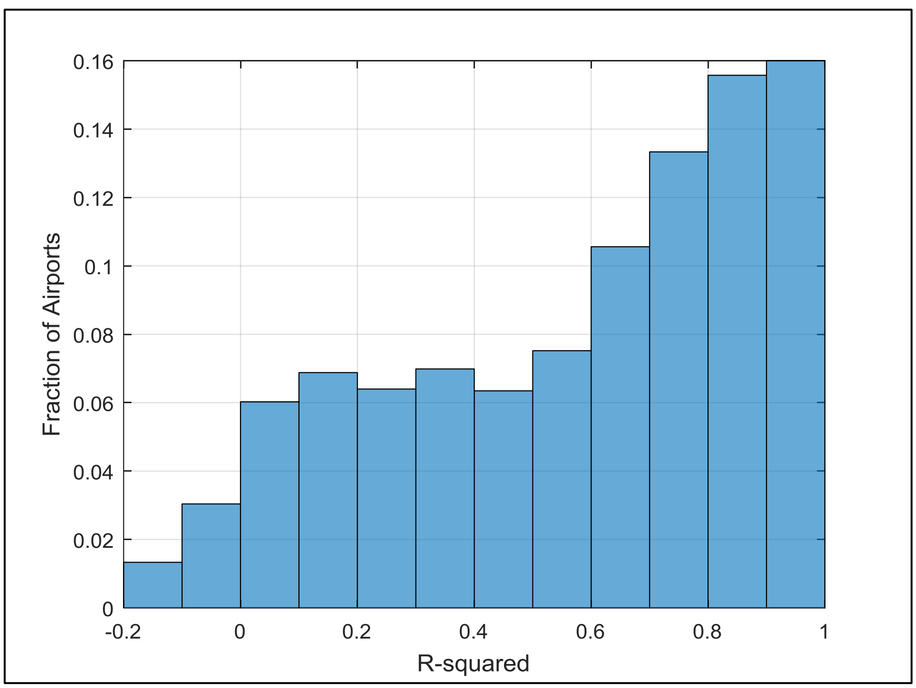

Figure 35: R-squared distribution for combined model results

As an additional measure of Goodness-of-Fit in the model, the Prediction Error was calculated for each airport and saved with the results for reference. This parameter measures how far the predicted values of a model are deviated from the original data. Prediction Error is calculated using Equation (8).

$$PE = \frac{|\hat{y} - y|}{y} \tag{8}$$

Where:

PE: Prediction Error (Fraction)

$\hat{y}$: Predicted value

y: Observed value

For example, Logan International Airport at Boston (BOS) has a low R^2 value of 0.116. However, the deviation of results from the original data is within +/- 3%, which can be considered a reasonable range for forecasting purposes. Therefore, the model is still considered acceptable with the current R^2 value. Figure 36 shows an example of applying a suggested threshold of +/-5% prediction error. It is noted that for such threshold, all predicted points are within the acceptable range. The reason of the low R^2 value is the presence of outliers in the data and the limited number of data points.

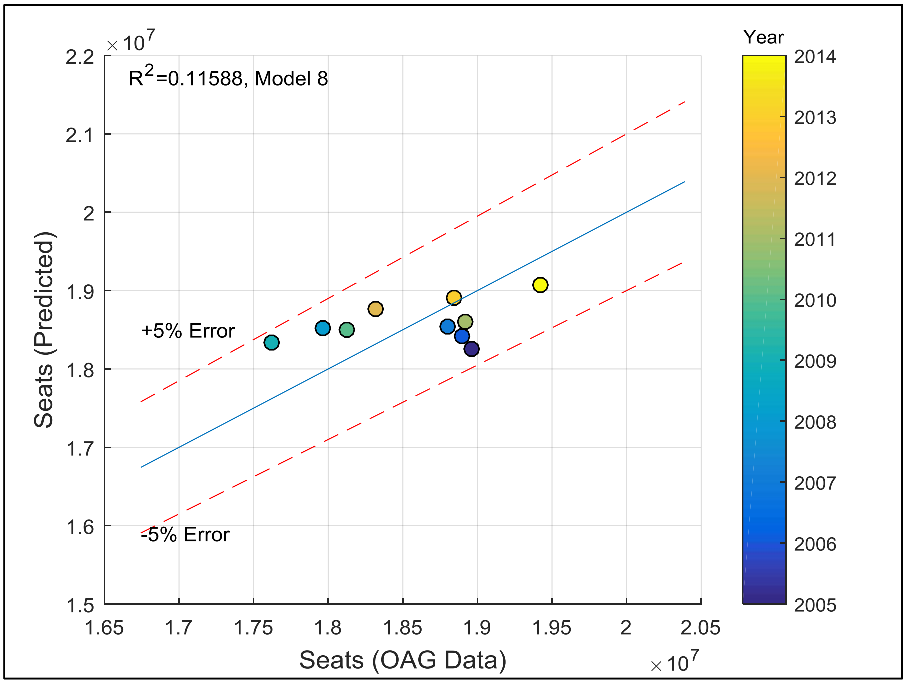

Figure 36: Model results vs data for BOS with upper/lower limits +/-5% Prediction Error

Many airports have relatively low values of R^2. However, these values did not necessarily correspond to high prediction errors. A number of airports had relatively low prediction errors despite having low R^2 values. These airports are shown in the lower left quadrant of Figure 37. Conversely, there were instances where high R^2 values corresponded to high prediction errors as shown in the upper right quadrant of Figure 37. This shows that though R^2 can be informative about the model performance, it is far from conclusive.

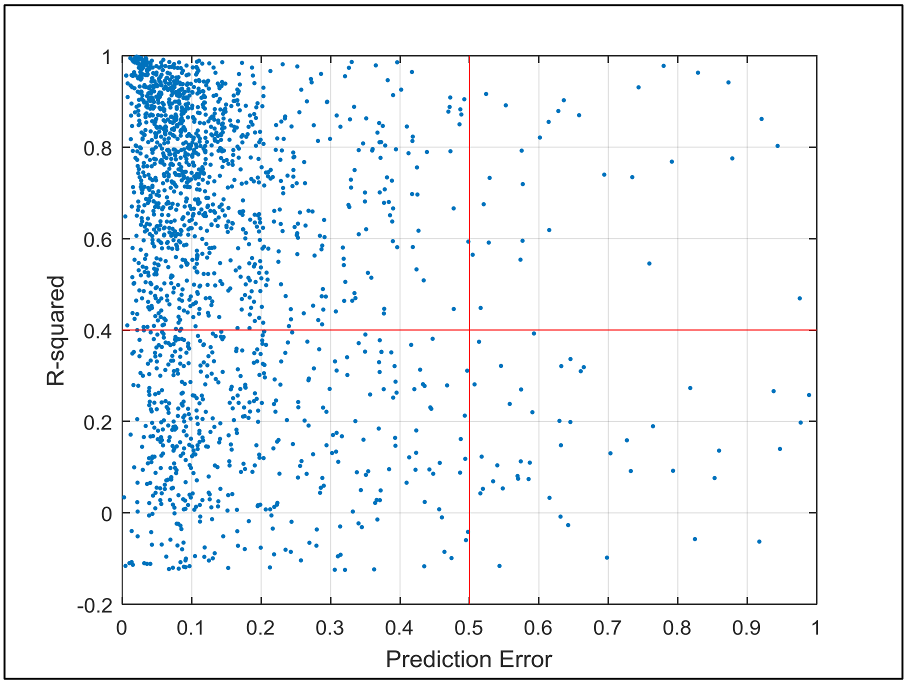

Figure 37: R-squared vs. Prediction Error

5.3 Sample Forecast Output

Below are a few examples of forecast generated by the model for multiple airports around the world. The figures show how the model tracks the historical data and estimates the future demand using the sub-model assigned by the Model Selection algorithm described in Section 4.5. The model ID is indicated on the graph of each airport. The description corresponding to each Model ID can be found in Table 2.

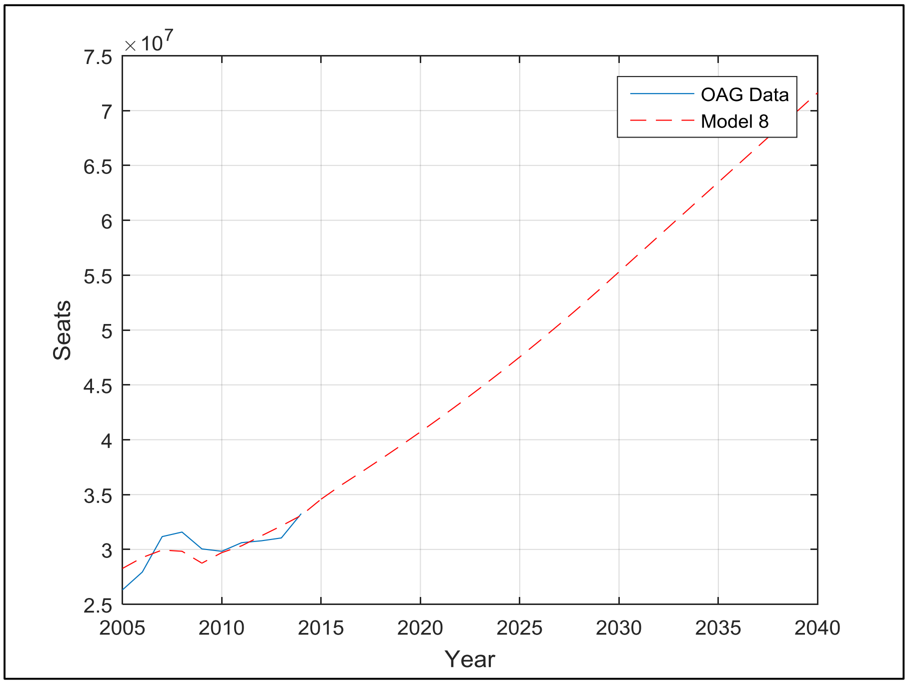

Figure 38: Forecast model for JFK

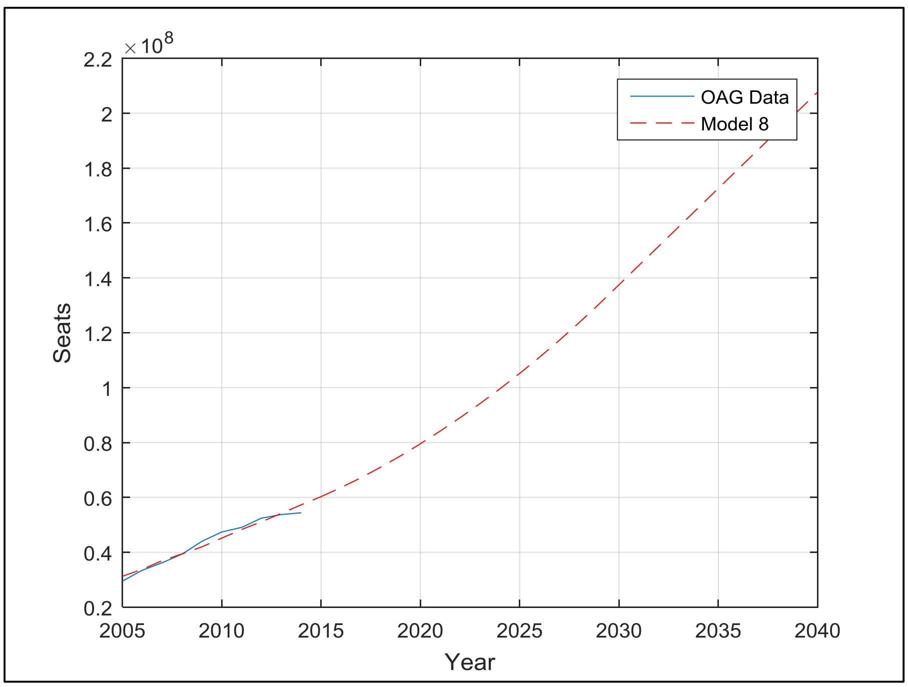

Figure 39: Forecast for PEK

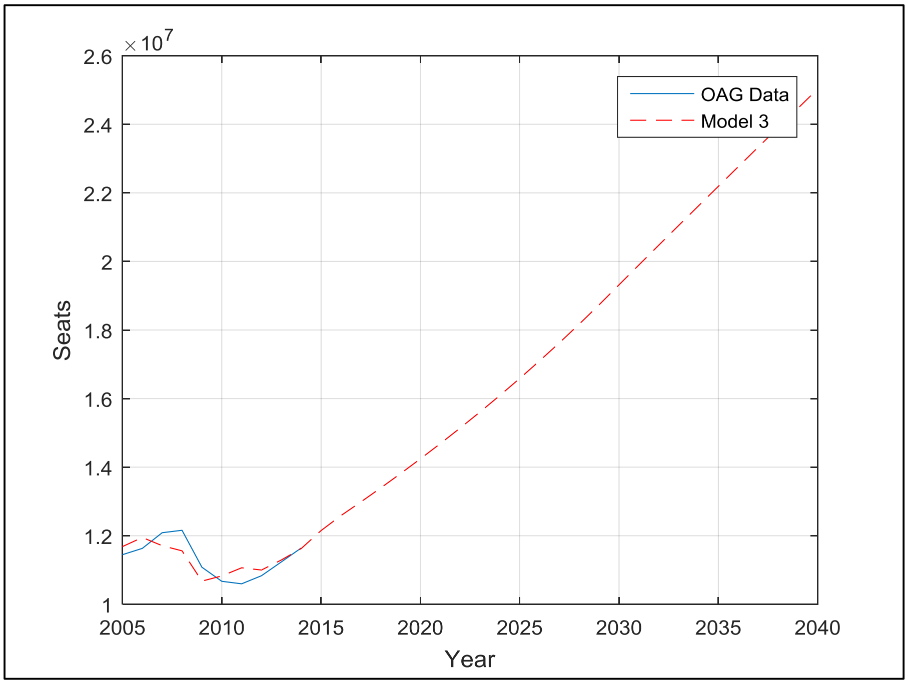

Figure 40: Forecast for SAN

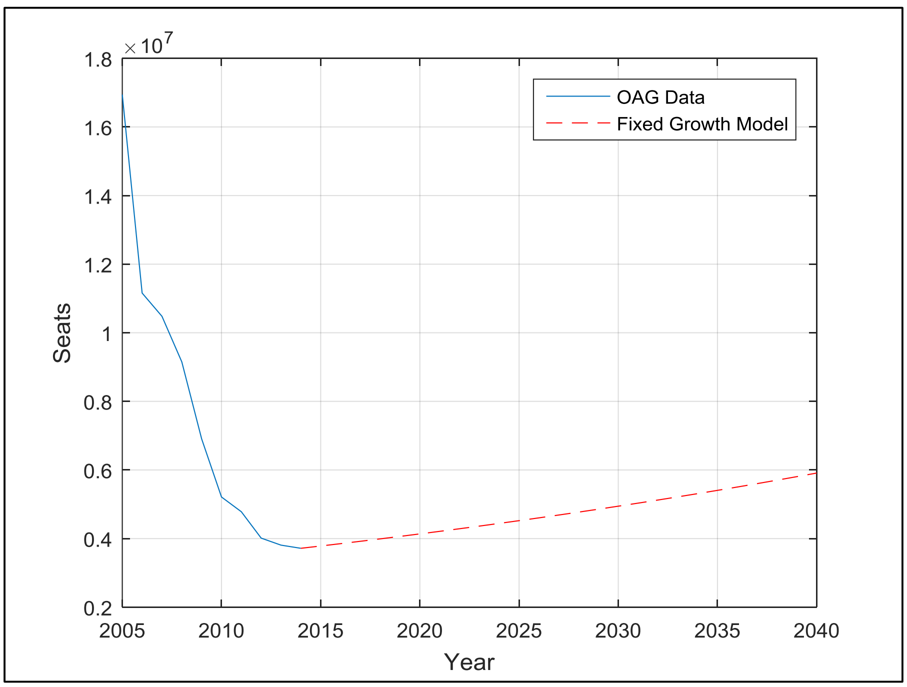

Figure 41: Forecast for CVG

On the other hand, the model provides the freedom to the user to exclude the choice of a capacity constrained model for airport in the analysis if only unrestricted growth forecasts are desired. Figure 42 shows the forecast for capacity-constrained model (40%) for JFK compared to the unrestricted linear regression.

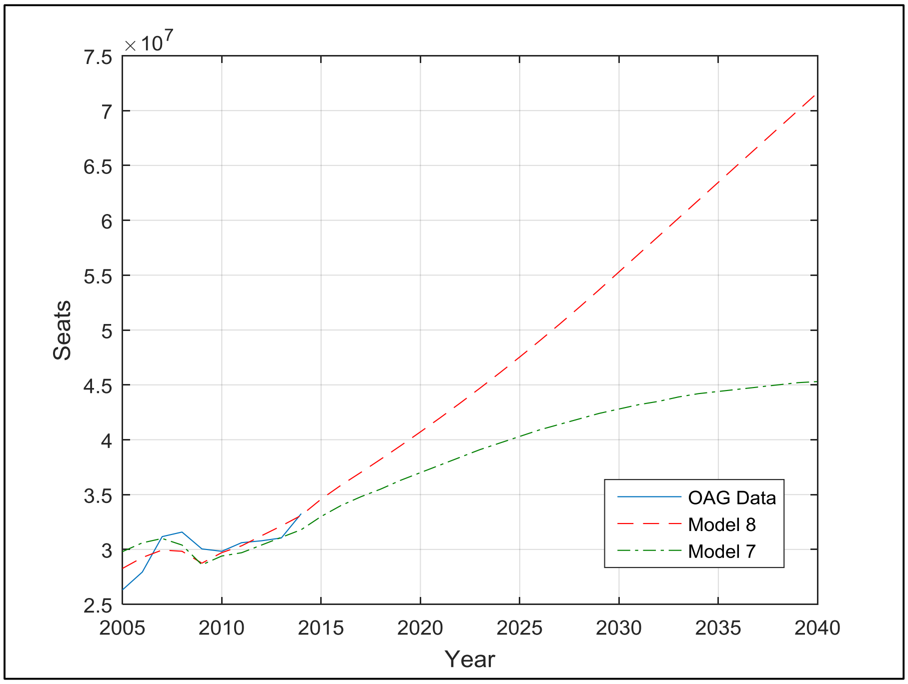

Figure 42: Capacity-constrained model for JFK

6 CONCLUSIONS

The Global Demand Model developed in this project estimates the number of seats offered at airports Worldwide. The model applies regression using GDP, population, and Herfindahl Index for airline market share at the airport as explanatory variables. The Global Demand Model utilizes a family of regression models that includes 3 linear models, 5 nonlinear models, and a fixed growth rate model. Each airport is individually assigned the model that best fits the historical demand data. The model assignment is conducted through an algorithm that first considers the R^2 value as its initial criterion. It then checks the assignment against underlying assumptions to see if the assignment is valid and reassigns if necessary.

The Global Demand Model covered 3,017 airports in the forecasting efforts. Regression analyses were performed for 1,875 airports that exhibited a positive historical trend for the last 5 years of data. For the remaining 1,142 airports, a fixed annual growth rate method was applied.

The most commonly assigned model, Model 8, was also the simplest. Model 8 was assigned to 55% of the airports in the regression analyses. It solely used GDP 60 nm around the airport as the explanatory variable for the number of seats.

The R^2 values were used to measure the model performance. However, when compared against the respective prediction errors, it was demonstrated that the R^2 was an imperfect measure of model performance. This conclusion resulted from the fact that a number of airports had relatively low prediction errors despite having low R^2 values.

The output of the model consists of three tables that contain model selection results, forecast generated by regression, and forecast generated by fixed growth rates all. In addition, the model produces an individual plot of the forecast for each airport.

7 RECOMMENDATIONS AND FUTURE WORK

The Airport Growth Model developed in this thesis can serve as the basis for modeling the distribution of passengers on the global air transportation network. This represents the fourth step of the Global Demand Forecast Model as shown in Figure 1.

For the sub-model described in Section 4.3.2, future Herfindahl Index values are necessary to be able to produce future passengers forecast. The Global Demand Model currently assumes that the market shares are going to stay the same beyond the year 2014. Researching the availability of airline market shares forecast is important to replace the current numbers in the model in order to have even more enhanced results.

The rough estimation of capacity in the capacity-constrained models can be replaced with a procedure that uses airport data such as the number of runways, runways configuration, and the length of runways. This procedure can make the model selection procedure and the forecast results more accurate.

Prediction error can be utilized in model assignment. As shown in our analysis, R^2 is not always representative of the model performance. The model selection algorithm can be modified to include the prediction error in the process along with R^2 values.

Investigate additional variables that might improve the regression models performance. The cost of travel can also be used. Airport specific characteristics such as the number of runways and the number of nearby airports can also be used to enhance the model.

Investigate other model formats. Future works can research additional model specifications to be added to the family of models to rather maximize the ability of the regression models to capture the relationships between the dependent and independent variables.

Acquiring more data years for the model is recommended to improve the regression model performance. Updating the data used in the model whenever new releases of population or GDP datasets are available is recommended to keep the model variables up-to-date with the latest forecasts.

REFERENCES

1. Balk, D, G Yetman, and A de Sherbinin. 2010. "Construction of gridded population and poverty data sets from different data sources." In *e–Proceedings of European Forum for Geostatistics Conference, Tallinn, 5–7 October 2010*.
2. Center for International Earth Science Information Network - CIESIN - Columbia University, and Centro Internacional de Agricultura Tropical - CIAT. 2005. "Gridded Population of the World, Version 3 (GPWv3): National Identifier Grid." In. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).
3. Chunshui, Jiang, Yu Haiyang, and Anwar Zubair. 2012. "On the ICAO system of air traffic forecasting." In *Proceedings of 2012 9th IEEE International Conference on Networking, Sensing and Control*.
4. Commission, Airports. 2013. "Discussion Paper 01: Aviation Demand Forecasting " In. UK: Airports Commission - An independent commission appointed by Government.
5. FAA. 2013. "Terminal Area Forecast Summary, Fiscal Years 2013-2040." In. United States: Federal Aviation Administration.
6. Gillen, David. 2009. 'The Future for Interurban Passenger Transport'.
7. Gillen, David, and Benny Mantin. 2009. 'Price volatility in the airline markets', *Transportation Research Part E: Logistics and Transportation Review*, 45: 693-709.
8. Grosche, Tobias, Franz Rothlauf, and Armin Heinzl. 2007. 'Gravity models for airline passenger volume estimation', *Journal of Air Transport Management*, 13: 175-83.
9. Karlaftis, M., K. Zografos, J. Papastavrou, and J. Charnes. 1996. 'Methodological Framework for Air-Travel Demand Forecasting', *Journal of Transportation Engineering*, 122: 96-104.
10. Marazzo, Marcial, Rafael Scherre, and Elton Fernandes. 2010. 'Air transport demand and economic growth in Brazil: A time series analysis', *Transportation Research Part E: Logistics and Transportation Review*, 46: 261-69.
11. Nordhaus, William, Quazi Azam, David Corderi, Kyle Hood, Nadejda Makarova Victor, Mukhtar Mohammed, Alexandra Miltner, and Jyldyz Weiss. 2006. 'The G-Econ database on gridded output: methods and data', *Yale University, New Haven*.
12. Pearce, Brian. 2015. "Challenges of high growth: Global aviation outlook." In.: IATA.
13. Shen, Ni. 2006. 'Prediction of International Flight Operations at Sixty-six U.S. Airports', Virginia Tech.
14. Suryani, Erma, Shuo-Yan Chou, and Chih-Hsien Chen. 2010. 'Air passenger demand forecasting and passenger terminal capacity expansion: A system dynamics framework', *Expert Systems with Applications*, 37: 2324-39.
15. UN. 2013. "World Population Prospects: The 2012 Revision, DVD Edition." In.: The United Nations.
16. USDA. 2014. 'Projected GDP Per Capita Dataset', U.S. Department of Agriculture Economic Research Service (ERS) Website. http://www.ers.usda.gov/data-products/international-macroeconomic-data-set.aspx.
17. Wooldridge, Jeffrey. 2012. *Introductory econometrics: A modern approach*.

APPENDIX A – SCREENSHOT OF MODEL SELECTION RESULTS MATLAB STRUCTURE FILE

Model_selection

1x1875 struct with 17 fields

Fields	ID	Rsquared	Model_Type	Model_ID	Initial_Model_ID	Data	DataYears	Predicted	Intercept	PE	X1	X2	C	b	d	a	R
1	'AAE'	0.3000	'Linear (GDP to...	8	8	[275043,245...	[2005,2006,2...	[2.4308e+05,...	1.9083e+05	[-0.1162,0.00...	1.0154e+04	[]	[]	[]	[]	[]	[]
2	'AAL'	0.7287	'Linear (GDP to...	8	3	[409604,456...	[2005,2006,2...	[7.8808e+05,...	3.9208e+06	[0.3718,-0.15...	-3.0517e+04	-3.2864e+06	[]	[]	[]	[]	0.0036
3	'AAN'	0.5390	'Linear (GDP on...	1	1	[88256,1061...	[2005,2006,2...	[9.4209e+04,...	-1.2054e+04	[0.0675,-0.16...	3.6380e+09	[]	[]	[]	[]	[]	[]
4	'AAQ'	0.3805	'Linear (GDP an...	3	3	[299996,308...	[2005,2006,2...	[2.7663e+05,...	-8.2200e+05	[-0.0779,-0.0...	7.7749e+11	3.3011e+06	[]	[]	[]	[]	[]
5	'ABA'	0.1703	'Linear (GDP to...	8	7	[173813,100...	[2005,2006,2...	[1.2363e+05,...	2.8522e+05	[-0.3278,0.06...	-7.0308e+04	[]	1.6216e+05	0.0127	-8.8343e+05	[]	0.1039
6	'ABD'	0.0336	'Linear (GDP to...	8	3	[154812,194...	[2005,2006,2...	[1.7191e+05,...	1.0775e+05	[0.0666,-0.09...	481.9335	-2.9130e+05	[]	[]	[]	[]	-0.0860
7	'ABI'	0.6089	'Linear (GDP an...	3	3	[136136,159...	[2005,2006,2...	[1.4239e+05,...	-7.8337e+04	[0.0459,-0.06...	1.8924e+09	-1.9280e+05	[]	[]	[]	[]	[]
8	'ABJ'	0.0395	'Linear (GDP to...	8	3	[1235632,11...	[2005,2006,2...	[1.0749e+06,...	1.0901e+06	[-0.0920,-9.7...	-4.3034e+03	9.6589e+06	[]	[]	[]	[]	-0.1240
9	'ABT'	0.4670	'Linear (GDP on...	1	1	[159041,160...	[2005,2006,2...	[1.4497e+05,...	1.4736e+04	[-0.0885,-0.0...	3.0145e+10	[]	[]	[]	[]	[]	[]
10	'ABV'	0.8819	'Linear (GDP to...	8	8	[1180008,17...	[2005,2006,2...	[1.1844e+06,...	-1.1705e+06	[0.0037,-0.18...	1.8165e+05	[]	[]	[]	[]	[]	[]
11	'ABY'	0.2628	'Linear (GDP to...	8	8	[70250,6850...	[2005,2006,2...	[6.2089e+04,...	1.9913e+05	[-0.1162,-0.1...	-2.2728e+03	[]	[]	[]	[]	[]	0.2628
12	'ABZ'	0.8512	'Linear (GDP to...	8	8	[1789060,19...	[2005,2006,2...	[1.7733e+06,...	-4.0163e+06	[-0.0088,-0.0...	4.2218e+04	[]	[]	[]	[]	[]	[]
13	'ACC'	0.9800	'Linear (GDP to...	8	8	[1156024,11...	[2005,2006,2...	[1.0759e+06,...	-1.5156e+05	[-0.0693,-0.0...	3.1724e+05	[]	[]	[]	[]	[]	[]
14	'ACE'	0.4604	'Linear (GDP to...	8	7	[1816005,19...	[2005,2006,2...	[2.4620e+06,...	3.2469e+06	[0.2990,0.104...	-1.0228e+05	[]	4.2471e+06	0.0010	-1.3938e+05	[]	-0.1220
15	'ACY'	0.0525	'Linear (GDP to...	8	7	[502472,492...	[2005,2006,2...	[5.9195e+05,...	-1.1008e+05	[0.2798,0.261...	2.5308e+04	[]	1.1489e+06	0.0095	-1.2703e+06	[]	-0.0475
16	'ADA'	0.9280	'Linear (GDP to...	8	8	[656664,963...	[2005,2006,2...	[5.6303e+05,...	-3.9075e+06	[-0.1426,-0.0...	1.9838e+05	[]	[]	[]	[]	[]	[]
17	'ADB'	0.9255	'Linear (GDP to...	8	8	[1486704,21...	[2005,2006,2...	[1.3797e+06,...	-1.1662e+07	[-0.0720,0.05...	5.0041e+05	[]	[]	[]	[]	[]	[]
18	'ADD'	0.9966	'Linear (GDP to...	8	8	[1347020,17...	[2005,2006,2...	[1.4262e+06,...	-1.0105e+06	[0.0588,-0.03...	3.4908e+06	[]	[]	[]	[]	[]	[]
19	'ADE'	0.7618	'Linear (GDP to...	8	3	[269631,243...	[2005,2006,2...	[2.6216e+05,...	-6.6407e+05	[-0.0870,0.05...	3.5645e+05	-3.2484e+05	[]	[]	[]	[]	0.1482
20	'ADK'	0.3776	'Linear (GDP to...	8	8	[2652,2802,...	[2005,2006,2...	[5.1845e+03,...	-4.8054e+04	[0.9550,1.357...	1.6080e+04	[]	[]	[]	[]	[]	[]
21	'ADL'	0.9213	'nonlinear (cap...	5	5	[3638319,40...	[2005,2006,2...	[3.7560e+06,...	[]	[0.0323,-0.01...	[]	[]	6.1983e+06	6.9573	6.0204e+05	[]	[]
22	'ADQ'	0.0441	'Linear (GDP to...	8	8	[115199,103...	[2005,2006,2...	[1.3807e+05,...	-8.8355e+04	[0.1985,0.398...	8.7858e+03	[]	[]	[]	[]	[]	[]
23	'ADZ'	0.8340	'Linear (GDP to...	8	8	[410643,465...	[2005,2006,2...	[4.1552e+05,...	2.0938e+05	[0.0119,-0.05...	1.2134e+05	[]	[]	[]	[]	[]	[]
24	'AEP'	0.7949	'Linear (GDP to...	8	8	[3849908,37...	[2005,2006,2...	[3.2847e+06,...	-2.6799e+06	[-0.1468,-0.0...	1.3441e+06	[]	[]	[]	[]	[]	[]
25	'AER'	0.8210	'nonlinear (cap...	7	7	[717012,826...	[2005,2006,2...	[6.1382e+05,...	[]	[-0.1439,-0.0...	[]	[]	2.4890e+06	125.2878	2.1534e+06	[]	[]
26	'AES'	0.7732	'Linear (GDP to...	8	8	[597070,607...	[2005,2006,2...	[5.3202e+05,...	-1.8618e+06	[-0.1089,-0.0...	1.2129e+05	[]	[]	[]	[]	[]	[]
27	'AEY'	0.6198	'Linear (GDP to...	8	8	[117460,129...	[2005,2006,2...	[1.1442e+05,...	-1.1894e+05	[-0.0259,-0.0...	6.4020e+04	[]	[]	[]	[]	[]	[]
28	'AFA'	0.2007	'Linear (GDP to...	8	7	[38186,3769...	[2005,2006,2...	[3.5831e+04,...	7.9406e+04	[-0.0616,-0.1...	-3.2455e+03	[]	4.8787e+04	0.0016	-8.1136e+05	[]	0.0860
29	'AFZ'	0.3845	'Linear (GDP an...	3	3	[24024,1749...	[2005,2006,2...	[2.2914e+04,...	-4.5515e+04	[-0.0462,-0.2...	6.6164e+09	2.6207e+04	[]	[]	[]	[]	[]
30	'AGA'	0.7916	'Linear (GDP an...	3	3	[549596,598...	[2005,2006,2...	[4.9865e+05,...	-1.1037e+06	[-0.0927,-0.0...	1.0209e+12	2.2294e+06	[]	[]	[]	[]	[]
31	'AGH'	0.1451	'Linear (GDP to...	8	3	[254526,188...	[2005,2006,2...	[2.3541e+05,...	-1.4272e+05	[-0.0448,0.22...	8.3547e+03	-5.4495e+05	[]	[]	[]	[]	-0.0483
32	'AGN'	0.0915	'nonlinear (cap...	5	5	[7390,5875,...	[2005,2006,2...	[3.6302e+03,...	[]	[-0.5088,-0.2...	[]	[]	9.1008e+03	5.0840e+07	913.8312	[]	[]
33	'AGP'	0.3785	'Linear (GDP an...	3	3	[7037264,73...	[2005,2006,2...	[7.1223e+06,...	-8.0262e+06	[0.0121,0.026...	6.7151e+11	7.3375e+07	[]	[]	[]	[]	[]
34	'AGR'	0.3098	'Linear (GDP an...	3	3	[5950,2975,...	[2005,2006,2...	[7.5164e+03,...	1.1520e+03	[0.2633,1.691...	1.1315e+10	6.8948e+04	[]	[]	[]	[]	[]
35	'AGS'	0.5829	'Linear (GDP an...	3	3	[272697,231...	[2005,2006,2...	[3.0627e+05,...	2.3306e+05	[0.1231,-0.02...	8.1821e+09	-1.2550e+06	[]	[]	[]	[]	[]
36	'AGU'	0.2802	'Linear (GDP an...	3	3	[335222,412...	[2005,2006,2...	[3.4728e+05,...	-1.2590e+06	[0.0360,-0.08...	2.3278e+11	1.8391e+06	[]	[]	[]	[]	[]
37	'AHB'	0.7416	'Linear (GDP to...	8	3	[841440,900...	[2005,2006,2...	[7.3057e+05,...	-1.8676e+05	[-0.0715,-0.1...	9.6573e+04	-3.9689e+06	[]	[]	[]	[]	0.7242
38	'AHO'	0.6701	'Linear (GDP to...	8	5	[730530,646...	[2005,2006,2...	[9.6207e+05,...	5.1802e+06	[0.2043,0.255...	-4.4558e+04	[]	1.3774e+06	1.2248e-06	-1.5639e+05	[]	0.3982
39	'AHU'	0.5803	'Linear (GDP an...	3	3	[15602,1858...	[2005,2006,2...	[1.0949e+04,...	-3.3488e+04	[-0.2982,-0.3...	2.8292e+10	1.5215e+04	[]	[]	[]	[]	[]
40	'AIN'	0.0772	'nonlinear (cap...	6	6	[24563,1693...	[2005,2006,2...	[1.3213e+04,...	[]	[-0.4621,-0.1...	[]	[]	16393	9.5244e+13	177.7404	[]	[]
41	'AIT'	0.7721	'Linear (GDP to...	8	4	[29941,2984...	[2005,2006,2...	[3.1853e+04,...	2.4626e+04	[0.0200,0.075...	3.2568e+03	[]	4.4912e+04	1.8594	0.0025	[]	0.6605
42	'AIU'	0.3160	'Linear (GDP to...	8	8	[5840,5824,...	[2005,2006,2...	[5.8129e+03,...	2.0013e+04	[-0.0046,-0.1...	-1.1106e+03	[]	[]	[]	[]	[]	0.3160
43	'AJA'	0.8688	'Linear (GDP an...	3	3	[631288,601...	[2005,2006,2...	[6.1247e+05,...	6.2950e+05	[-0.0298,-0.0...	3.6354e+09	-2.7684e+06	[]	[]	[]	[]	[]
44	'AJF'	0.1655	'Linear (GDP an...	3	3	[168268,199...	[2005,2006,2...	[1.8125e+05,...	-5.4805e+05	[0.0771,-0.03...	4.1328e+09	3.4347e+05	[]	[]	[]	[]	[]
45	'AJI'	0.8210	'Linear (GDP to...	8	8	[13068,1587...	[2005,2006,2...	[-2.7574e+0...	-2.8629e+05	[-1.2110,0.05...	9.3775e+03	[]	[]	[]	[]	[]	[]
46	'AJL'	0.6521	'Linear (GDP to...	8	4	[46650,7643...	[2005,2006,2...	[7.2544e+04,...	1.0709e+04	[0.3958,-0.03...	4.2761e+03	[]	1.4132e+05	33.8119	0.0019	[]	0.5211
47	'AJR'	0.7001	'Linear (GDP to...	8	8	[40560,4550...	[2005,2006,2...	[4.1706e+04,...	-4.3830e+03	[0.0282,-0.03...	1.1472e+03	[]	[]	[]	[]	[]	[]
48	'AJU'	0.8867	'Linear (GDP an...	3	3	[460503,564...	[2005,2006,2...	[4.7728e+05,...	-7.4522e+05	[0.0364,-0.11...	6.0382e+11	-2.3037e+06	[]	[]	[]	[]	[]
49	'AKF'	0.1666	'Linear (GDP to...	8	4	[36356,2564...	[2005,2006,2...	[1.6513e+04,...	-1.0478e+04	[-0.6268,-0.3...	6.0366e+03	[]	7.2014e+04	46.6367	2.2551e-05	[]	0.0364
50	'AKI'	0.3521	'Linear (GDP to...	8	8	[9251,4062,...	[2005,2006,2...	[8.5172e+03,...	-8.0333e+04	[-0.0793,1.68...	1.0694e+03	[]	[]	[]	[]	[]	[]
51	'AKL'	0.8023	'Linear (GDP to...	8	8	[7808971,79...	[2005,2006,2...	[7.8398e+06,...	-3.1927e+06	[0.0039,0.024...	1.2977e+06	[]	[]	[]	[]	[]	[]
52	'AKU'	0.9766	'nonlinear (cap...	6	6	[56120,5019...	[2005,2006,2...	[1.6400e+04,...	[]	[-0.7078,-0.4...	[]	[]	6.4782e+05	3.0932e+03	2.2636e+05	[]	[]
53	'AKV'	0.6619	'Linear (GDP to...	8	8	[13321,1409...	[2005,2006,2...	[1.3228e+04,...	-9.8321e+04	[-0.0070,0.14...	4.7737e+04	[]	[]	[]	[]	[]	[]
54	'AKX'	0.9712	'Linear (GDP an...	3	3	[100264,101...	[2005,2006,2...	[9.6439e+04,...	-7.0198e+04	[-0.0381,0.09...	5.0269e+10	-1.8056e+05	[]	[]	[]	[]	[]
55	'ALA'	0.9690	'Linear (GDP on...	1	1	[1113621,13...	[2005,2006,2...	[1.1091e+06,...	-2.7882e+06	[-0.0040,0.08...	5.5497e+12	[]	[]	[]	[]	[]	[]
56	'ALC'	0.4526	'Linear (GDP an...	3	3	[4704224,48...	[2005,2006,2...	[4.9200e+06,...	-2.7988e+06	[0.0459,0.057...	4.1317e+11	3.6589e+07	[]	[]	[]	[]	[]
57	'ALF'	0.4066	'Linear (GDP to...	8	4	[284482,278...	[2005,2006,2...	[2.7273e+05,...	2.9116e+05	[-0.0457,0.00...	-398.2668	[]	3.1429e+05	3.2141e+06	1.8127e-05	[]	-0.1230
58	'ALG'	0.9559	'Linear (GDP to...	8	8	[2867026,28...	[2005,2006,2...	[2.7141e+06,...	-2.7655e+06	[-0.0533,-0.0...	3.8053e+06	[]	[]	[]	[]	[]	[]
59	'ALH'	0.6513	'nonlinear (GD...	2	2	[36900,3934...	[2005,2006,2...	[3.3470e+04,...	-29.4444	[-0.0930,-0.0...	[]	[]	[]	1.0999	[]	3.0004	[]
60	'ALS'	0.3447	'Linear (GDP to...	8	4	[20786,1894...	[2005,2006,2...	[1.7873e+04,...	-6.3382e+03	[-0.1021,0.02...	475.7180	[]	2.5059e+04	1.5778e+04	1.9371e-05	[]	0.2475
61	'AMD'	0.8861	'Linear (GDP an...	3	3	[1319621,18...	[2005,2006,2...	[1.5309e+06,...	2.4287e+05	[0.1601,-0.01...	3.7255e+12	-1.1759e+07	[]	[]	[]	[]	[]
62	'AMH'	0.4697	'Linear (GDP to...	8	6	[4200,7130,...	[2005,2006,2...	[6.2762e+03,...	-5.4384e+04	[-0.1458,-0.5...	805.7152	[]	2.6503e+04	6.4698e-05	-1.0665e+06	[]	-0.0521
63	'AMM'	0.9442	'Linear (GDP to...	8	8	[2381891,28...	[2005,2006,2...	[2.5026e+06,...	-3.3332e+06	[0.0507,-0.02...	1.4016e+05	[]	[]	[]	[]	[]	[]
64	'AMQ'	0.9628	'Linear (GDP to...	8	4	[165424,182...	[2005,2006,2...	[1.3676e+05,...	-7.9055e+05	[-0.2117,-0.0...	3.0001e+05	[]	8.2386e+05	1.2890e+04	0.0022	[]	0.9343
65	'AMS'	0.2410	'Linear (GDP an...	3	3	[25198371,2...	[2005,2006,2...	[2.6867e+07,...	-2.7069e+07	[0.0662,0.013...	1.1698e+12	1.8871e+08	[]	[]	[]	[]	[]
66	'ANF'	0.6746	'nonlinear (cap...	7	7	[864495,963...	[2005,2006,2...	[7.0388e+05,...	[]	[-0.1858,-0.1...	[]	[]	2.0909e+06	177.7468	5.9615e+05	[]	[]
67	'ANM'	0.3260	'Linear (GDP to...	8	3	[13563,7545...	[2005,2006,2...	[9.9256e+03,...	4.5566e+04	[-0.1920,0.08...	-5.0267e+04	-2.5241e+04	[]	[]	[]	[]	0.2047
68	'ANS'	0.8609	'Linear (GDP to...	8	4	[10200,6950...	[2005,2006,2...	[1.3772e+03,...	-5.9178e+04	[-0.8172,-0.4...	7.4984e+03	[]	4.0046e+04	6.8389e+06	0.0016	[]	0.7566
69	'ANX'	0.8953	'Linear (GDP to...	8	3	[60377,6034...	[2005,2006,2...	[5.6704e+04,...	-4.7734e+04	[-0.0080,-0.0...	4.8571e+03	-1.5385e+05	[]	[]	[]	[]	0.4813
70	'AOI'	-0.0138	'nonlinear (cap...	4	4	[344653,315...	[2005,2006,2...	[3.1489e+05,...	[]	[-0.0864,-0.0...	[]	[]	4.6555e+05	0.0062	-9.7663e-05	[]	[]
71	'AOK'	0.4249	'Linear (GDP to...	8	3	[87141,7344...	[2005,2006,2...	[7.9102e+04,...	1.0142e+05	[-0.0013,0.11...	-3.2750e+03	-1.6564e+05	[]	[]	[]	[]	-0.0746
72	'AOR'	0.8393	'Linear (GDP to...	8	2	[235744,219...	[2005,2006,2...	[2.6390e+05,...	-6.2078e+05	[0.0919,0.032...	1.3081e+04	[]	[]	-0.8677	[]	579.3312	-0.0632
73	'APL'	0.6098	'Linear (GDP to...	8	8	[147822,144...	[2005,2006,2...	[1.2385e+05,...	-7.0100e+05	[-0.1622,0.01...	9.2573e+03	[]	[]	[]	[]	[]	[]
74	'APO'	0.2282	'Linear (GDP to...	8	8	[87038,9543...	[2005,2006,2...	[8.1586e+04,...	2.0140e+04	[-0.0626,-0.1...	3.5046e+03	[]	[]	[]	[]	[]	[]
75	'AQG'	0.9417	'Linear (GDP to...	8	4	[5500,1750,...	[2005,2006,2...	[-1.2290e+0...	-1.8474e+05	[-0.6866,1.34...	5.6725e+03	[]	1.0252e+05	1.6977e+07	0.0051	[]	0.8749
76	'AQI'	-0.1168	'Linear (GDP on...	1	1	[69333,8017...	[2005,2006,2...	[5.4972e+04,...	2.9780e+04	[-0.2071,-0.3...	9.9300e+08	[]	[]	[]	[]	[]	[]
77	'AQP'	0.3495	'Linear (GDP to...	8	8	[685658,597...	[2005,2006,2...	[4.6225e+05,...	-3.4497e+06	[-0.3258,-0.0...	6.3115e+04	[]	[]	[]	[]	[]	[]
78	'ARH'	0.7943	'Linear (GDP to...	8	8	[244243,255...	[2005,2006,2...	[2.1621e+05,...	-4.4447e+05	[-0.1148,0.05...	2.3892e+05	[]	[]	[]	[]	[]	[]
79	'ARI'	0.6050	'Linear (GDP to...	8	8	[323422,279...	[2005,2006,2...	[2.4755e+05,...	-4.8687e+05	[-0.2346,-0.0...	6.0194e+04	[]	[]	[]	[]	[]	[]
80	'ARK'	0.6373	'nonlinear (GD...	2	2	[11040,5520...	[2005,2006,2...	[8.1574e+03,...	-62.0123	[-0.2611,1.44...	[]	[]	[]	1.3168	[]	4.2855	[]

APPENDIX B – List of MATLAB Scripts for The Global Demand Model Analysis

	Script Name	Description
1.	GPW_read.m	Reads GPW data and returns a 3432x8640 matrix for the selected year, 2.5 minute resolution cells
2.	Read_UN_population_Historical.m	Imports the UN population data for years 2000-2010 from the provided spreadsheet from the UN website
3.	Read_UN_population.m	Imports the UN population projection for years 2010-2040 from the provided spreadsheet from the UN website
4.	Geoshow_GPW.m	Plots a color map of population data
5.	Gridded_pop.m	Distribute the country level population to the 2.5 minute resolution grid saved by year
6.	identify_countries.m	Reads countries ID matrix from the GPW data and creates Countries_Info.mat file where population and other information about each country is stored.
7.	catchment_area_POP.m	Finds Population within catchment areas around airports
8.	Cell_proportion.m	finds the proportion of each cell in the grid according to the population distribution pattern in GPW
9.	catchment_area.m	Finds cell indices surrounding airport in catchment area within a specified radius
10.	catchment_area_GDP.m	Finds GDP within catchment areas around airports
11.	country_to_grid.m	Converts GDP data from country level to a global grid of 2.5x2.5 minute resolution

12.	GDP_Beyond_2030.m	Extrapolation of GDP for years beyond 2030 (USDA forecast) and calculation of GDP per capita around airports
13.	Local_GDP.m	calculate the loacl GDP per capita in catchment areas around airports
14.	Read_USDA_GDP.m	Import GDP data from USDA spreadsheet
15.	apt_coordinates.m	This function finds the airport coordinates from the struct file "airport_info" and returns both lat and lon of the current airport
16.	G_ECON_GDP.m	This script creates a grid of G-econ GDP data and plots the values on the map
17.	GDP_National_to_Grid.m	Convert GDP from National to Grid Level by country
18.	Gridded_GDP.m	This script transforms the 1-degree resolution GDP matrix to 2.5-minute resolution matrix (it also distributes the values to cells within countries borders)
19.	Read_GDP.m	Import data from spreadsheet
20.	Import_OAG_Data.m	Reads the OAG Schedule file and returns a struct array
21.	OAG_Airlines_Analysis.m	Find OAG total seats per airport per carrier by year
22.	OAG_Totals_zero_stops.m	Find OAG total seats per airport by year
23.	OAG_Summary.m	Reformatting OAG data by airport instead of by year
24.	Herfindahl_index_calculation.m	Calculation of Herfindahl Index by airport by year
25.	OAG_Summary_Herfindahl.m	Adding Herfindahl Index to OAG_summary

26.	Read_Airport_List.m	Import airport data (ID, coordinates, region…etc) from OAG spreadsheet
27.	Read_OAG.m	Read and save OAG data for each year divided into 2 parts (to reduce file size)
28.	Split_variables.m	Split the fields of large OAG data to separate struct files to reduce the file size
29.	Airport_list_all.m	Create the airport list of all airports in OAG
30.	Decreasing_Trend_AGR.m	Annual Growth Rates (AGR) for airports with decreasing trend
31.	LinearRegression_ALL.m	Model 1
32.	NonLinearRegression_ALL.m	Model 2
33.	LinearRegression_ALL_GDP_Herfindahl.m	Model 3
34.	NonLinearRegression_cap_ALL_10.m	Model 4
35.	NonLinearRegression_cap_ALL_20.m	Model 5
36.	NonLinearRegression_cap_ALL_30.m	Model 6
37.	NonLinearRegression_cap_ALL_40.m	Model 7
38.	LinearRegression_GDP_ALL.m	Model 8
39.	Regression_Analysis.m	Run the complete regression analysis
40.	Regression_model_selection.m	Model selection algorithm
41.	Regression_model_selection_part2.m	Model selection algorithm (2nd iteration)
42.	Airport_all_Trends.m	Plot the trends of airport data
43.	Airport_Trends_normal.m	Plot normalized passengers and GDP trends
44.	Airport_Trends_variables.m	Plot the Dependent vs. Independent variables
45.	Get_Growth_Rate.m	Return the fixed growth rate of airport by respective region
46.	Plot_Airport_List.m	Plot a list of airports on the world map
47.	Plot_Forecast_Model.m	Plot the forecast result of the model for individual airport