TAF-M Methodology 2023

Federal Aviation Administration

Address: 800 Independence Avenue, SW, Washington DC 205890
Website: www.faa.gov

TABLE OF CONTENTS

INTRODUCTION

Terminal Area Forecast

Terminal Area Forecast (TAF) is the official FAA forecast of aviation activity for U.S. airports. FAA office Aviation Policy and Plans (APO) is the office responsible for the forecasts. TAF contains active airports in the National Plan of Integrated Airport Systems (NPIAS) including FAA-towered airports, federal contract-towered airports, nonfederal towered airports, and non-towered airports. The TAF is comprised of TAF-Modernization (TAF-M) airports and TAF-Legacy (TAF-L) airports. The total number of airports is over 3000 in TAF. This document focuses on introducing the forecast methodology for the TAF-M airports.

TAF-M versus TAF-L airports

TAF distinguishes between TAF-M and TAF-L airports because of the difference in the forecast methodology. The forecast methodology for the TAF-M airports is based on advanced econometric models. Model estimation and forecasting using sophisticated statistical technique is made possible for the TAF-M airports due to the availability of large datasets. On the contrary, due to the limitation in data, forecast methodology is largely based on trend analysis for the TAF-L airports.[1]

TAF-M Airports

TAF-M airports are the airports in TAF with more than 100,000 annual enplanements. The number of TAF-M airports changes every year because of the annual fluctuations in enplanements by airport. The number of TAF-M airports is usually around 200 every year. For example, there are 218 TAF-M airports in the 2023 TAF, and there are 223 TAF-M airports in the 2022 TAF.

Even though TAF-M airports is a small subset of all the airports in TAF (200 or so airports out of over 3000 airports), these airports capture approximately 99 percent of enplanements in

[1] APO is at the preliminary stage of experimenting with advanced time series models for the subset of the TAF-L airports that are FAA facilities. This subset of the TAF-L airports has more data available than the TAF-L non-FAA facilities because FAA facilities report operation data to the FAA. Even though the data availability for the TAF-L FAA facilities is still quite limited compared to the TAF-M airports, it is sufficient for the application of advanced time series modeling.

the United States. Bureau of Transportation Statistics has abundant airline data for TAF-M airports, which allows the application of more advanced econometric models to facilitate the forecast production.[2]

TAF-M Domestic and International

TAF-M forecast models include domestic passenger demand forecast and the international passengers. Modeling methods are different between domestic and international passengers. Domestic passenger modeling uses panel data models. The data has a panel data structure with the groups defined as origin-destination airport-pairs and the time series dimension defined as quarterly data from calendar year 2000 until the year with the latest data.

Due to the difference in data limitation, international passenger forecast modeling is performed with time series exponential smoothing model. The data for the international passengers are only available at the airport level, not at the airport-pair level. As a result, international passenger forecasts are executed at the airport level. This is in contract to the domestic passenger modeling where the forecasts are built at the airport-pair level and the numbers are aggregated to the airport level.

TAF Forecast Parameters

The annual production of TAF results in publication of annual enplanements and airport operations for all the airports in NPIAS. Enplanements include domestic and international enplanements, and it has four user groups: air carrier enplanements, commuter enplanements. Airport operations include commercial operations, general aviation operation, and military operations. The definitions for these user groups are published in the “glossary” document on the TAF website: https://taf.faa.gov/

TAF-M Forecast Parameters

This report focuses on the generation of domestic enplanements and commercial operations for the TAF-M airports. It lays out the method to conduct the forecasts for the origin and

[2] https://www.transtats.bts.gov/

destination (O&D) passengers and the process of converting passengers to the number of commercial operations. Once the passenger forecasts are complete, the number of forecasted passengers during the entire forecast horizon until 2050 are fixed, and the forecasted operations are generated by converting the number of passengers into the number of operations given the assumptions on the mix of aircraft types and load factors for a given O&D airport-pairs. The airport-pair forecasts are aggregated to the airport level. At the end of the production, APO publishes airport level forecasts on https://taf.faa.gov/.

This report does not cover the process of converting O&D passengers into segment passengers, and from segment passengers to enplaned passengers. It also does not discuss the process of separating enplaned passengers into air carrier enplanements, commuter enplanements, us flag enplanements, and foreign flag enplanements. There is also no discussion on grouping the commercial operations into air carrier and air taxi operations. General aviation forecasts and military operation forecasts are also not covered in this report.

DOMESTIC TAF-M MODEL

Modeling overview

Domestic TAF-M modeling process has two main parts: in-sample estimation and the out-of-sample forecasts. The purpose of the in-sample estimation is to obtain the estimated values of important parameters. The estimated parameters can then be used in the next step to generate domestic passenger forecasts.

Two types of econometric models are used in the sample estimations, both models belong to the family of panel data models. The first type is the Correlated Random Effects (CRE) model. The application of the CRE model was well documented in Wooldridge's publications and workshops.[3] It is an attractive choice given the characteristics of our dataset. The second type is the Arellano-Bond (AB) model, which is a dynamic panel data model.

CRE model uses estimated income elasticity of air travel demand to forecast passenger demand. It is a good choice for airports with stable growth history and therefore the forecasts can be developed with economic data such as personal income growth projections.

AB model is a time series model for a panel dataset. It uses historical trend to help with the forecasts. As a result, it is a good choice for airports with high growth history or negative growth history. It works well for airports whose forecasts do not follow the economic data projections.

Correlated random effects (CRE) model

In empirical panel data literature, Fixed Effects (FE) model is often chosen over Random Effects (RE) model as RE imposes a strict assumption of zero correlation between the individual heterogeneity and all the covariates, which is rarely true in the empirical panel data. On the contrary, such correlation is allowed to be non-zero under the FE framework.

---

[3] See "Correlated Random Effects Models with Unbalanced Panels" by Jeffrey M. Wooldridge in Journal of Econometrics (2019, Volume 211, Issue 1, Pages 137-150) and "Estimate" workshop slides in 2014 for the "Linear Panel Data Models" section. "Estimate" workshop was hosted by the Department of Economics, Michigan State University.

However, FE model has an unfortunate property that the time-invariant variables are removed during the estimation, which is problematic if the variable of interest does not vary with time. Furthermore, among the time varying variables, those with very little time series variation will not be accurately estimated with the FE approach.

CRE is an econometric model designed for panel datasets that combine the features in the FE and RE models. CRE unifies RE and FE in that the RE estimates of the time-varying covariates are the FE estimates given that the average across time for a given airport-pair is included.

A standard panel data model using Wooldridge's (2013) notation is:

$$y_{it} = \eta_t + \boldsymbol{x_{it}\beta} + \boldsymbol{z_i\delta} + c_i + u_{it},\ t = 1, \dots, T \qquad (1)$$

- $\{u_{it}: t = 1, \dots, T\}$ are the idiosyncratic errors.
- $v_{it} = c_i + u_{it}$ is the composite error at time *t*.
- c_i is the unobserved airport-pair heterogeneity for airport-pair *i*.
- y_{it} is the O&D passengers at the airport-pair *i* at quarter *t*.
- $\boldsymbol{x_{it}}$ is the vector for the time-varying independent variables (covariates) such as airfares and metro level income.
- $\boldsymbol{z_i}$ is the vector for the time invariant covariates like airport-pair distance.

The central issue with the panel data model is the variable c_i which is the airport-pair heterogeneity. It is an unobservable varible that does not vary much over time but is corelated with the other covariates in the model. FE model will take care of the airport-pair heterogeneity, but it will produce inaccurate coefficient estimates for the covariates that do not have much time series variation. Furthermore, the time invariant covariates will drop out of the FE model. This poses an issue in APO's model because, the airport-pair distance variable would be eliminated in the FE model. In addition, most of the variation in the most important covariate, metro level personal income, comes from the cross-sectional dimension. The time series variation in the personal income variable is relatively small.

CRE model offers a more flexible alternative than the FE model so that the analyst can choose not to use FE estimators for some of the covariates with very small time series variation. Doing so can minimize the loss in the accuracy of the estimated coefficients. By decomposing c_i as $c_i = \psi + \bar{\boldsymbol{x}}_i\boldsymbol{\xi} + a_i$, equation (1) can be expressed as:

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + \boldsymbol{z}_i\boldsymbol{\delta} + \psi + \bar{\boldsymbol{x}}_i\boldsymbol{\xi} + a_i + u_{it} \qquad (2)$$

In equation (2), $\bar{\boldsymbol{x}}_i$ is the vector of the time average covariates. The RE estimate of $\boldsymbol{\beta}$ when $\bar{\boldsymbol{x}}_i$ is included is the FE estimate. Airport-pair distance would be preserved in the term $\boldsymbol{z}_i\boldsymbol{\delta}$ using equation (2). Furthermore, metro level personal income variables would be more accurately estimated in the CRE framework.

CRE model in-sample estimation

The in-sample estimation for the CRE model uses DB1B historical data from calendar year 2000 until the latest calendar year data, combined with the real personal income provided by the S&P Global.

The empirical passenger demand model is specified as follows:

$$\log\left(Passenger_{ijt}\right) = \beta_0 + \beta_1 \log\left(Fare_{ijt}\right) + \beta_2 \log\left(Route_{ijt}\right) + \beta_3 \log\left(Distance_{ij}\right) + \beta_4 \log\left(Income\ Origin_{ijt}\right) + \beta_5 \log\left(Income\ Dest_{ijt}\right) + a_{ij} + u_{ijt} \qquad (3)$$

where *ij* indexes airport-pair between origin *i* and destination *j* and *t* indexes quarter. a_{ij} is the airport-pair heterogeneity for a given airport-pair *ij*, which can be understood as the unobserved airport-pair effect. The airport- pair unobserved effect contains things such as the underlying business model, managerial ability, or the cost structure, things that are roughly constant over time during the sample periods. The error u_{ijt} is the idiosyncratic error. It represents unobserved factors that change over time.

$\log\left(Passenger_{ijt}\right)$ is the log of sum of the O&D passengers flying on airport-pair *ij* at quarter *t*. O&D passenger refers to the passengers in airline O&D survey ticket data (or, 10% sample) flying from an origination airport *i* to a destination airport *j*. There can be intermediate stops between airport *i* and airport *j*. Passengers on either non-stop flights or multiple-stop flights are included.

$\log(Fare_{ijt})$ is the log of the average market fares paid by the O&D passengers flying on airport-pair *ij* at quarter *t*. Average fare is reported in the ticket portion of the O&D or 10% sample data.

$\log(Route_{ijt})$ is the log of the total number of routes provided by the airlines flying on airport-pair *ij* at quarter *t*. For example, there are various ways to fly from DCA to RDU, such as, in addition to direct, from DCA to CLT to RDU or from DCA to ATL to RDU. APO counts the number of unique routes for each directional market and add up all the unique combinations serving a given airport-pair *ij* at quarter *t*.
$\log(Distance_{ij})$ is the log of non-stop market miles for airport-pair *ij*. This variable does not vary with time.

$\log(Income\ Origin_{ijt})$ is the log of total real personal income (millions 2012$) at the MSA level for the origination airport on airport-pair *ij* at quarter *t*. Similarly, $\log(Income\ Dest_{ijt})$ is the log of total real personal income (millions 2012$) of the MSA associated with the destination airport *j*.

CRE model out-of-sample forecast

The forecasting procedure after CRE model in-sample estimation is a straightforward process where a dynamic equation, developed from log-log approximation, is used to forecast domestic O&D passengers for a given airport-pair *ij* in quarter *q* and forecast year *y*. For each forecast year *y*, equation (4) is executed for each quarter, from quarter 1 through quarter 4, then *y* is set to *y+1*, so on and so forth.

$$Domestic\ O\&D\ Passengers_{ij,q,y+1} = Domestic\ O\&D\ Passenger_{ij,q,y} * \Bigg(1 + \hat{\beta}_1\left(\frac{Fare_{ij,q,y+1}}{Fare_{ij,q,y}} - 1\right) + \hat{\beta}_2\left(\frac{Route_{ij,q,y+1}}{Route_{ij,q,y}} - 1\right) + \hat{\beta}_3\left(\frac{Distance_{ij,q,y+1}}{Distance_{ij,q,y}} - 1\right) + \hat{\beta}_4\left(\frac{Income\ Origin_{ij,q,y+1}}{Income\ Origin_{ij,q,y}} - 1\right) + \hat{\beta}_5\left(\frac{Income\ Dest_{ij,q,y+1}}{Income\ Dest_{ij,q,y}} - 1\right)\Bigg) \quad (4)$$

Notice that the forecast equation (4) has a time subscript *y* that is different from the previous time subscript *t* that is used for the in-sample estimation. This is because the forecast is generated with annual growth rate of the same quarter for each airport-pair *ij*. For example, the income growth at the origin airport $\left(\frac{Income\ Origin_{ij,q,y+1}}{Income\ Origin_{ij,q,y}} - 1\right)$ measures the annual personal income growth rate given the same quarter.

All the variables in equation (4) are in their original level rather than the logarithm level.[4] This is important because the final forecasts need to be expressed in the levels, not the logarithm values for practical purposes.

To implement equation (4), one will need forecast values for covariates on the right-hand side such as *Fare* and *Route*, etc. Currently, APO has yet to develop the forecast values on the time-varying covariates in equation (4) other than *Income Origin* and *Income Dest*. Consequently, $\left(\frac{Fare_{ij,q,y+1}}{Fare_{ij,q,y}} - 1\right) = 0$ and $\left(\frac{Route_{ij,q,y+1}}{Route_{ij,q,y}} - 1\right) = 0$ for all *y* in the forecast years. $\left(\frac{Distance_{ij,q,y+1}}{Distance_{ij,q,y}} - 1\right) = 0$ because the nonstop distance at any given airport-pair *ij* is not changing in all future y.

Given these conditions, the effective dynamic forecast equation becomes:

$$Domestic\ O\&D\ Passengers_{ij,q,y+1} = Domestic\ O\&D\ Passenger_{ij,q,y} * \left(1 + \right.$$

$$\left. + \hat{\beta}_4 \left(\frac{Income\ Origin_{ij,q,y+1}}{Income\ Origin_{ij,q,y}} - 1\right) + \hat{\beta}_5 \left(\frac{Income\ Dest_{ij,q,y+1}}{Income\ Dest_{ij,q,y}} - 1\right)\right) \quad (5)$$

Equation (5) is a much simpler version of equation (4). $\hat{\beta}_4$ and $\hat{\beta}_5$ are the income elasticity estimates for the domestic air travel demand. With the income elasticity estimates combined with the forecast values of the real personal income, provided by S&P Global, APO is able to forecast domestic O&D passengers for all airport-pair *ij* from the first forecast year *y* until the last forecast year which is 2050 for the 2023 TAF.

In the future, APO plans to expand equation (5) to include domestic airfare growth projections. Equation (5) can be expanded to equation (6):

[4] Equation (4) can be derived from equation (3) through a Cobb-Douglas type transformation.

$$Domestic\ O\&D\ Passengers_{ij,y+1} = Domestic\ O\&D\ Passenger_{ij,y} * \left(1 + \hat{\beta}_1\left(\frac{Fare_{ij,y+1}}{Fare_{ij,y}} - 1\right) + \hat{\beta}_4\left(\frac{Income\ Origin_{ij,y+1}}{Income\ Origin_{ij,y}} - 1\right) + \hat{\beta}_5\left(\frac{Income\ Dest_{ij,y+1}}{Income\ Dest_{ij,y}} - 1\right)\right) \qquad (6)$$

This can be accomplished by introducing an equilibrium model of demand and supply. In addition to the demand side model specified in equation (3), APO is planning to add a supply-side equation with the airfare as the dependent variable. With a supply and demand equilibrium model, APO can predict future values in the domestic airfare which will make the estimation for equation (6) possible in the future.

Arellano-Bond (AB) model

In addition to the CRE model, APO also uses the Arellano-Bond model, a dynamic panel data model, to forecast the domestic O&D passengers for some airports. The primary difference between AB model and CRE model is that AB model uses historical trend to help with generating the forecasts. Technically, this means that there is lagged dependent variable in the covariate list for the AB model. The in-sample estimation is specified generally as below:

$$y_{ijt} = \rho y_{ij,t-1} + \boldsymbol{x}_{ijt}\boldsymbol{\beta} + \boldsymbol{z}_{\boldsymbol{ijt}}\boldsymbol{\gamma} + c_{ij} + u_{ijt} \qquad (7)$$

- $E\left(u_{it} \middle| y_{i,t-1}, \dots, y_{i0}, \boldsymbol{x}_i, z_{it}, \dots, z_{i1}\right) = 0$
- $\boldsymbol{x_{ij}} = \left(\boldsymbol{x_{ij1}}, \boldsymbol{x_{ij2}}, \dots, \boldsymbol{x_{ijT}}\right)$
- $\boldsymbol{z}_{ij} = (\boldsymbol{z}_{ij1}, \boldsymbol{z}_{ij2}, \dots, \boldsymbol{z}_{ijT})$
- $The\ \{\boldsymbol{x}_{ijt}\}\ are\ strictly\ exogenous$
- $The\ \{\boldsymbol{z}_{ijt}\}\ are\ sequentially\ exogenous$
- $c_{ij}\ is\ the\ airport - pair\ heterogeneity$
- $u_{ijt}\ is\ the\ idiosyncratic\ error$

In equation (7), y_{ijt} is the domestic O&D passengers flying on airport-pair *ij* at quarter *t*. $y_{ij,t-1}$ is the one-year lagged domestic O&D passengers on airport-pair *ij*. Vector $\boldsymbol{x}_{ijt}$ includes the real personal income variables at the MSA level for the origination airport and the destination airport on airport-pair *ij*. Vector $\boldsymbol{z_{ijt}}$ are the instruments. Because the sample has significant number of airport-pairs and each airport-pair has long time series data, the estimators in equation (7) are statistically consistent.

The estimator for the airport-pair heterogeneity c_{ij} is specified in equation (8). $\bar{y}_{ij}$ and $\bar{y}_{ij,t-1}$ are the time averages of the O&D passengers on airport-pair *ij*. $\bar{\boldsymbol{x}}_{ij}$ and $\bar{\boldsymbol{z}}_{ij}$ are the time averages for vectors $\boldsymbol{x}_{ijt}$ and $\boldsymbol{z}_{ijt}$.

$$\hat{c}_{ij} = \bar{y}_{ij} - \hat{\rho}\bar{y}_{ij,t-1} - \bar{\boldsymbol{x}}_{ij}\widehat{\boldsymbol{\beta}} - \bar{\boldsymbol{z}}_{ij}\widehat{\boldsymbol{\gamma}} \quad (8)$$

The one-step ahead forecast is specified as:

$$\hat{y}_{ij,T+1} = \bar{y}_{ij} + \hat{\rho}\left(y_{ijT} - \bar{y}_{ij,-1}\right) + \left(\boldsymbol{x}_{ij,T+1} - \bar{\boldsymbol{x}}_{ij}\right)\widehat{\boldsymbol{\beta}} + (\boldsymbol{z}_{ij,T+1} - \bar{\boldsymbol{z}}_{ij})\widehat{\boldsymbol{\gamma}} \quad (9)$$

The empirical specification for equation (7) for airport-pair *ij* at quarter *t* is:

$$\log\left(Passenger_{ijt}\right) = \gamma_0 + \gamma_1 \log\left(Passenger_{ij,t-4}\right) + \gamma_2 \log\left(Income\ Origin_{ijt}\right) + \gamma_3 \log\left(Income\ Dest_{ijt}\right) + a_{ij} + u_{ijt} \quad (10)$$

Unlike the in-sample estimation equation for the CRE model, the covariate list for the AB model is much shorter with only one lagged dependent variable and the real personal income at the origination and destination airport metro areas. The rest of the covariates in equation (3) are excluded from equation (10) because including them would lead to highly correlated covariates due to the lagged dependent variable. The lagged dependent variable $Passenger_{ij,t-4}$ captures plenty variations from the dependent variable.

The estimator for the airport-pair heterogeneity c_{ij} can then be empirically calculated in equation (11):

$$\hat{c}_{ij} = \boldsymbol{mean}\left(\log\left(Passenger_{ijt}\right)\right) - \boldsymbol{mean}\left(\log\left(Passenger_{ij,t-4}\right)\right)\widehat{\gamma_1} - \boldsymbol{mean}\left(\log\left(Income\ Origin_{ijt}\right)\right)\widehat{\gamma_2} - \boldsymbol{mean}\left(\log\left(Income\ Dest_{ijt}\right)\right)\widehat{\gamma_3} \quad (11)$$

The empirical forecast equation for airport *ij* at quarter t is:

$$\log\left(Passenger_{ijt}\right) = \widehat{\gamma_1} \log\left(Passenger_{ij,t-4}\right) + \widehat{\gamma_2} \log\left(Income\ Origin_{ijt}\right) + \widehat{\gamma_3} \log\left(Income\ Dest_{ijt}\right) + \hat{c}_{ij} \quad (12)$$

Lastly, the log values need to be transformed into the levels by taking the exponential and multiply the exponential value with a correction factor as described in Wooldridge's book "Introductory Econometrics: A Modern Approach" (2016).[5] Due to the nature of the AB model, it captures the historical trend of the dependent variable and incorporates the trend into the forecast equation. Consequently, it is well suited for the airports with histories of either very high growth or very low growth, i.e., growth patterns that do not fall into the average of most airports. The forests for these airports do not typically follow the personal income growth projections, which means an alternative model other than CRE is needed.

[5] See Chapter 6 "Multiple Regression Analysis: Further Issues". Section 6-4c "Predicting y When log(y) Is the Dependent Variable". Page 190-194.

ASSIGNING O&D PASSENGERS TO ROUTES

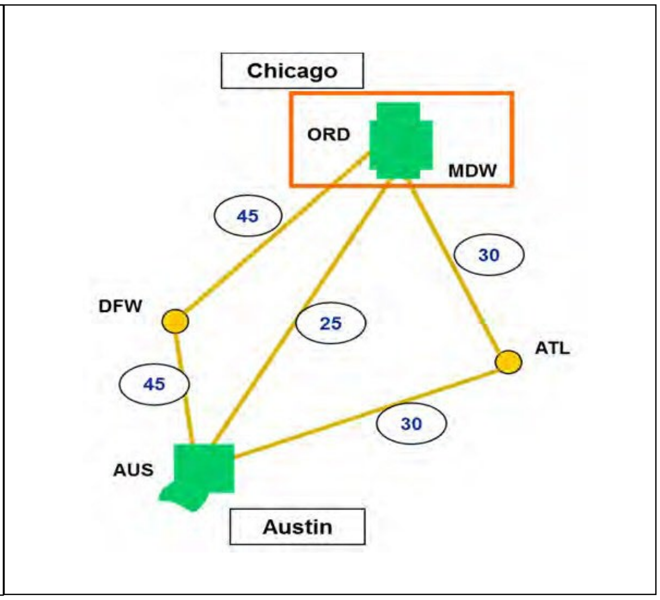
Process overview

Assignment of passenger flow between i-j to various routes (i.e., i-j; i-k-j; i-k-l-j, etc.) within the NAS is necessary to obtain segment or route-level forecasts. This is accomplished using an assignment algorithm where the number of passengers (say, between i and j) is distributed across various routes (direct or i-j; or indirect possibilities such as i- k-j (2-coupon or segments); or i-k-l-j (3-coupon or segments), etc.) are based on the historical information available for the same quarter last year. This ensures that the entire network is taken into consideration during route assignment.

Routing passengers in a metropolitan pair

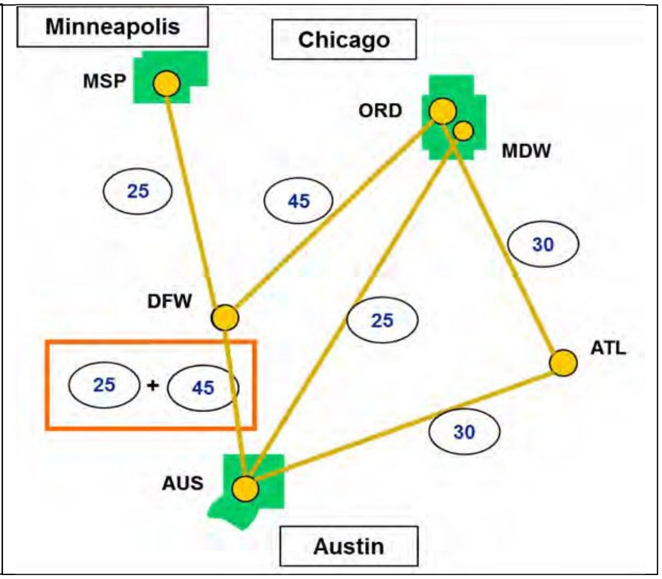
The assignment process begins by selecting an origin and a destination metropolitan area (i.e., MSAs). Once the areas are finalized for the assignment, the next step is to determine the number of routes that have been flown between the two metropolitan areas in the same quarter of the last observed year. The number of routes will include both non-stop (i-j) and multi-stop routes (or indirect routes). After the number of routes are identified, historical data is overlaid to determine the number of passengers flying each specific route.

The historical coupon data is pulled from the same quarter of the prior year to to minimize the impact of seasonality on route evolutions. Percent distribution of passengers by route is then applied to the passenger origin- destination demand forecast to arrive at the number of passengers expected to travel each specific route segment.

Take, for example, the case of 100 passengers that are projected to fly between Austin (in this case, origin *i*) and Chicago (in this case, destination *j*) MSAs. Our first task is to identify valid routes that were flown between the pairs at present (i.e., same quarter last year or the reference point). Notice that the option of multiple airports within a single MSA, Chicago with two airports ORD and MDW in this example, is treated as part of the observed "route" choice. Choice of airport in a multi- airport MSA is a research area that APO is presently exploring. We allocate passengers among existing route segments according to the current percentage distribution for that O&D market. Thus, 25% is observed to fly direct between AUS and MDW; while 45% takes AUS-DFW to ORD; and the remaining 30% takes AUS-ATL- MDW.	

We currently do not attempt to predict new routes and keep these observed routes constant going forward. Evolution of new routes in the NAS is a research area that APO is exploring together with researchers in the community [for an application of route choice in limited context, see Bhadra and Hogan (2008): "Choice of Route Networks: A Qualitative Choice Model for Over-Land and Over-Water Routes", *Journal of Aircraft*, Vol. 45(1), January-February 2008, page. 56-63.

Expansion to NAS

The concept described above is then applied to an additional pair, say, for example, Austin to Minneapolis, of metropolitan markets to determine the passenger demand by route segment. The assignment process continues until the entire NAS is mapped and passengers are distributed across all routes within the NAS. The assignment process is accomplished using SQL, as manual calculation is not feasible for the over 35,000 O&D market pairs, and numerous routes often exceeding hundreds of thousands, serving the primary O&D markets in the NAS.	

Upon completing this process for all MSAs and associated routes, we compare and adjust routed passengers with T100 segment passengers (Form 41), a database where commercial airlines report all passengers that flew routed segments. The database also provides valuable information regarding types of aircraft that were used to fly these passengers in the segments. This information is mined and used to allocate and project aircraft departures in the routed segment which we describe in the following section.

DETERMING AIRCRAFT DEPARTURE AND OPERATIONS

Process overview

After determining the distribution of passengers across various routes within the NAS, it is important to determine the type of aircraft that is flown on each specific route, as it will help determine the number of departures (and operations) on the specific route.

Determining number of aircraft departures on a specific route segment requires the information of the following variables. They include, but are not limited to:

- Number of passengers in segments
- Performance limits of A/C
- Distance between segments
- Operating costs per mile
- Type of airport in both ends of the segments

Notice that from the section described above, we have information about number of routed passengers, both history and forecast, distance, and types of airports (i.e., large, medium and small hub in the NAS). Performance limits of aircraft are determined by the types of aircraft while operating costs vary by types of aircraft and airlines and available from Form 41. At present, there are over 90 distinct aircraft types in the system. Because modeling over 90 distinct types of aircraft is complex and computationally infeasible, the types are aggregated to reduce these complexities. This narrows down the number of distinct aircraft types to be used for modeling to a manageable level. Examples of the aircraft type groupings that can be used include seat type, number of engines, aircraft range, and other definitions of missions. At present, we use classification according to seats as specified by AEE and is defined in Table 1.

Table 1: Seat Class Definition

Aircraft ID Class	Aircraft Seat Class	Min Seats	Max Seats
1	0	0	0
2	19	1	19
3	1	20	50
4	2	51	100
5	3	101	150
6	4	151	210
7	5	211	300
8	6	301	400
9	7	401	500
10	8	501	600
11	9	601	650
12	999	651	999999999

Groupings other than seats have been tried and tested in the past as well [see, for example, Bhadra (2005); "Choice of Aircraft Fleets in the US Domestic Scheduled Air Transportation System: Findings from a Multinomial Logit Analysis", *Journal of the Transportation Research Forum*, Vol. 44(3), Fall, 2005, page.143-162.

Short-run assignment of aircraft to routes

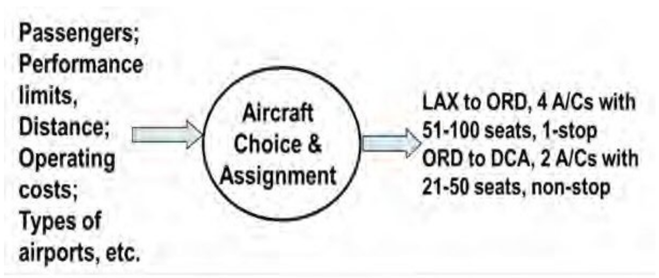
Oftentimes, airlines' choice of aircraft fleet, particularly in the short run of 4-5 years, is somewhat rigid. This is so because once the airlines have aircraft in their fleet inventory and/or orders are firm, the choices to fly them are fixed due to both network and financial commitment. Of course, there are some limited flexibilities that may still be available to airlines through arrangement with other airlines and/or leasing companies, but generally speaking, inventory and firm orders in the book generally guide airlines' choice of routes and markets. Given that, we assume that aircraft serving the segments of the NAS within the first 5 years (or 20 quarters) of the forecast is same as they have been observed in the past. In other words, aircraft that have flown a particular segment will continue to do so in the first 5 years (or 20 quarters) of the forecast and there will be no change. We assign the same aircraft as observed in the last year of the same quarter on a particular segment and continue for 20 quarters. Beyond that, aircraft choice can be modeled as described next.

Long-run assignment: applying the econometric model

The multinomial choice model [described in Bhadra (2005) in detail] is applied to the aircraft choices to arrive at departure forecasts for the years following the first five years of the forecast. Without going into technical detail, figure below illustrates the process of forecasting aircraft operations. Using routed passengers, performance limits, distance and types of airports, number of departures between two segment pairs are determined. The result of applying multinomial choice model is the number of aircraft that can be flown on a particular route or segment based on the passenger demand is described in below.

Since performance limits, distance and types of airports do not change over time, therefore, numbers of departure are driven primarily by routed passengers and seasons (i.e., different quarters). These counts, multiplied by 2, provide us the number of operations accounting for both landings (i.e., arrivals) and take-offs (i.e., departures).

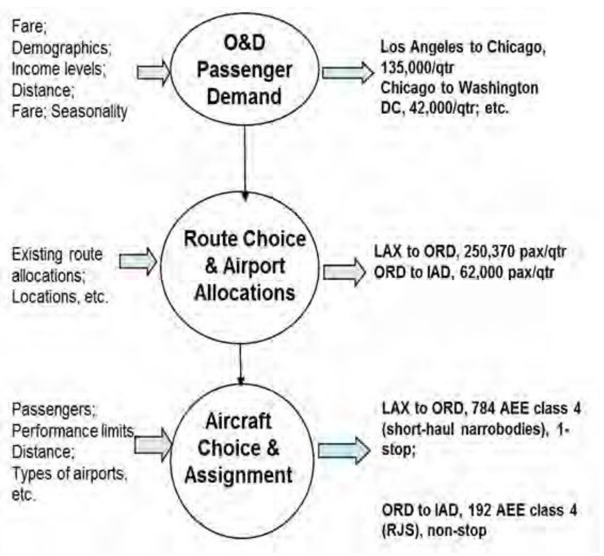
Types of aircraft departures associated with segments result from the multinomial choice models where aircraft types are aggregated according to number of seats. Once that assignment is complete, we now can determine number of departures between two segments, for example, LAX to ORD where four aircraft (with 51-100 seats) are required. Similarly, two aircraft (with 21-50 seats) are needed to fly non- stop between ORD to DCA. This process continues until all segments have been exhausted.



SUMMARY OF TAF-M PROCESS IN THE DOMESTIC NAS

The TAF-M process, as described above, can be summarized in the diagram on the right side of the panel: first, we estimate and forecast O&D passenger demand between airport-pairs, *i-j* and seasonality (i.e., quarters 1 through 4) across samples (at present, 2000- 2014); second, using those forecasts and combining them with observed routes, we determine segment or routed passenger flows by airport pairs.

Finally, using those routed passengers and combining them with performance limits of types of aircraft, distance, and types of airports, we estimate and forecast departures and operations by types of aircraft for the years beyond year five into the forecast. For short-run, we assign the observed aircraft into the routes.



This network and aircraft-integrated process is carried through all segments of the commercial service airports within the United States in order to forecast passengers and departures. When these segments are aggregated at the level of origin airports, they produce passenger and operation activities at the airport, equivalent to TAF-L from before. Aggregation of segments at the destination level airports may provide landing activities but that is not presently done. This is what is known as TAF-M.

ADDITIONAL DETAILS

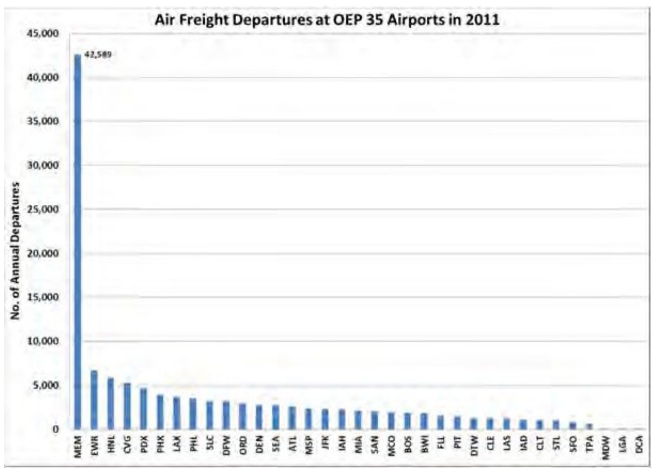
Once we have output from the TAF-M process described above, additional analysis is performed to create the official terminal area forecast. In this section, we briefly describe these stand along processes.

Treatment of domestic cargo

Within the domestic NAS, cargo departures account for approximately a little less than 5%. In some airports, for example, MEM (Memphis, TN) and SDF (Louisville, KY), their numbers are higher; but in many of the commercial airports, cargo operations are large.

As the panel on the right shows that although MEM far surpassed, by a magnitude of 8 on average, annual cargo departures in other large airports are significant as well. Ranging somewhere between 1000-6500 annual departures, these airports facilitate air cargo transportation that has been growing quite substantially over the last two decades.

From TAF-M process, we can easily identify these departures. For departures that are associated with zero passengers are identified, by the DOT, as cargo departures. Based on the estimate and forecast of departures demand described above in TAF-M process above, we account for cargo departures for which associated passengers is zero.



We report these departures, by both segment routes and aggregated at the airport, as cargo departures. Multiplying these departures by two yields cargo operations at a particular airport.

Forecast international passenger and cargo

Originally developed by the MITRE Corporation, estimation and forecast of international passenger, departures and cargo departures are presently done outside the integrated TAF-M. In order to undertake this step, we first process T100International segment data by all carriers. At present, data covering the period of 2000-2014 are used. T100 international segment data provide the number of passengers and other details by flown segments.

Passengers in many international segments are relatively low. For this reason, we define thick segments or markets in international module as sum of passengers in the last 4 quarters >= 10,000. Thin segments/markets are defined as those that have passengers less than 10,000 as sum in the last 4 quarters. Similar to the case of domestic markets, there are many thin international segments. On average, thin international markets/segments number around 4,500 while thick markets/segments are around 2,000. While numbering over double the number of thick markets, thin markets account for less than 10% of total international passengers. Thick markets, on the other hand, carry over 90% of all international passengers. To put it differently, thick segments accounting for only 1/3rd of total number of segments account for over 90% of passengers while 2/3rd of total segments are relatively thin and account for only 10% of passengers.

Segment passengers, both thick and thin, are modeled and forecasted using ESM procedure researched, developed and tested by the MITRE Corporation. Time series specification using Winters' method in SAS is the procedure by which international passengers are estimated and forecasted by segment pairs. Assuming segment load factors and average seat factors (i.e., fixed aircraft type) being fixed at the observed levels, departures are then calculated using passenger forecasts. Both passenger and departures, and operations, in turn, then added onto the main TAF-M databases.

Reallocation of aircraft by types

It is evident from the aircraft orders, both delivered and those in order book, that the US, in particular and the world in general, is at a point of another evolution in aircraft fleet. Aircraft that were introduced in 1990s to facilitate the evolution and expansion of hub-and-spoke network are beginning to go out of circulation. Generally speaking, these are smaller jets and are of types of ERJ135, ERJ145 and CRJ100, CRJ200, etc. These are being gradually replaced by larger variants of ERJ170/190s, and CRJ700/900s.

One shortcoming of TAF-M is that it is not capable of predicting the evolution of new aircraft in the system. Both during the short-run (i.e., less than 5 years) and in the long-run, aircraft we assigned as per last observed in the segment as in the case of former or aircraft is assigned via estimation based on history data as in the case of the latter. In either case, new aircraft that had never been observed before cannot be assigned during the forecast periods.

In order to bridge the gap between what aircraft has been assigned in the TAF-M and what is anticipated, we formulate algorithms that reflect these reallocations of aircraft. Aircraft that are assigned or modeled to be of smaller sizes (i.e., earlier generation) are to be replaced, irrespective of the types of segments in the NAS, by newer and larger aircraft (i.e., next generation). Thus, for example, an older version of ERJ145, typically with 50 seats, will be replaced by 0.71 of ERJ170 with a typical seat configuration of 70. This is explained in detail in Section 7.

Combined with the assumption of fixed (and observed) load factors in segments, this would result in recalculation of seats available in that segment. Recalculated seats combined with fixed load factor would result in reduced number of departures and operations. It will also lead to recalculation of passengers in the segment. We impose these reallocations on a macro basis. When macro algorithms are carried through accounting for types of aircraft and their substitutions, seats, departures, and passengers from TAF-M are recalculated and aggregated at the segment levels. These numbers are then put forward as the final output of the TAF-M. These recalculated numbers are what reported in the final TAF-M.

Calibration to short-term passenger and operations using airline schedules

By the time TAF-M is put into use, typically in the beginning of the calendar year (i.e., 2016 for the present cycle), a year is already past (2015) from the time actual data were available (i.e., 2000-2014). Furthermore, a lot is already known, from airlines' schedule data, for about half of the year of 2016 in addition to the past year, 2015. In other words, by the time forecast data are finalized and ready for public use, a further need arises to calibrate it with what is known already (i.e., 2015-2016 first half).

Notice, however, with a lag of 2-3 quarters, primary data corresponding to TAF-M (i.e., O&D) are not readily available from the DOT and only partial information for the period known from T100 data which have a slightly faster publication (with one quarter lag). Therefore, we need to combine available T100 data with airline schedule data in order to update the first part of the forecast (i.e., 6 quarters). We account for this and calibrate the forecasts, for core 30 airports at present, by the available T100 and airline schedule data. Oftentimes, adjustments are minor but we make sure that trends in the forecasts correspond to what we observe from the newer data from T100 and available airline schedule.

Calibration to OPSNET data

Although the entire process is exercised using DOT data and the airline data that are available from schedule, we need to calibrate the final data by the FAA's official count. Known as OPSNET (CountOps), this is the repository of the Agency's accounting of departures and take-offs as counted by the individual towers..

In order to carry out these calibrations for departures, we take the latest observed numbers from CountOps and calculate the ratios that translate them onto TAF-M generated actual data for the corresponding periods. At that point, we also compare the rates of growth for departures between these two data sources. Once we have established these correspondences for departures at the airport levels for which CountOps are available, we allow the TAF-M generated growth rates (for the forecast periods) to drive the actual CountOps departure data into forecasts. In other words, CountOps generated actuals are driven by TAF-M generated departures growth rates during the forecast periods. Data related to passengers and other details remain unaltered.

Treatment of exception airports

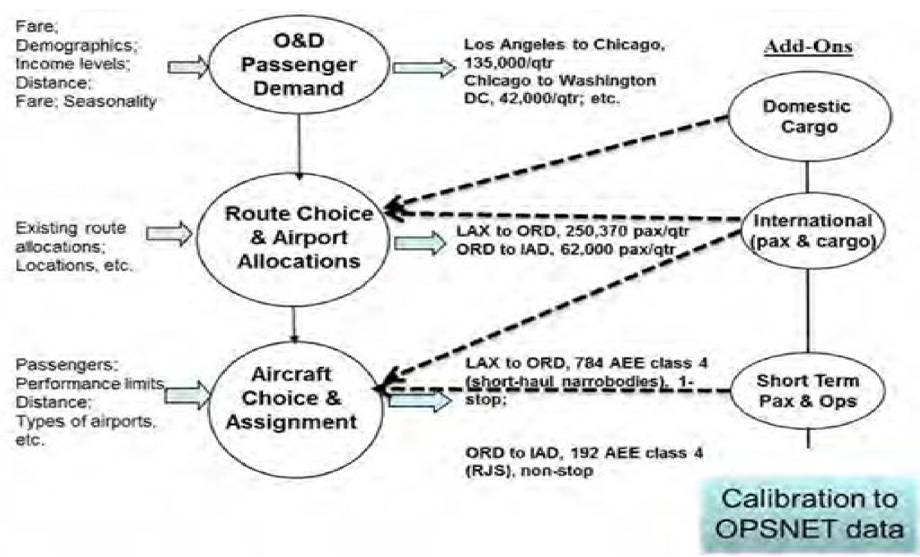
Once we generated these forecasts and compared them against the actuals and published forecasts from last year's TAF, we output these forecasts at the FAA's final forecasts for most airports. Some airports require additional treatments and they can be categorized, at present, into two: (a) constrained airports; and (b) demand scenarios affected by additional forces or supply-induced demand.

It is evident from the discussion above that TAF-M and the additional steps that we outlined above represent demand for aviation activities that are unconstrained. In other words, no constraints in airports are taken into account. However, there are a couple of airports in the NAS, La Guardia (LGA) and Reagan National (DCA), which are restricted by slots. These may change in the future but at present, these are the only two airports for which slots or operating limitations are taken into account i.e., slot restricts the unconstrained forecasts sometimes in the future. In order to take these slot restrictions and their impact on departures and, in turn, on the passengers, we further calibrate forecast data. Depending upon the time at which these slots restrict departures, we calibrate both departure mixes and passengers to arrive at the final forecasts for these two airports. LGA and DCA are the only two airports with

(a) in place requiring some further calibration. In TAF-M, demand is estimated and forecasted using the history data. History of activities at airports for which restrictions have been in place will necessarily undermine the actual potential. The case of Dallas Love (DAL) airport is a case in point. As the Wright Amendment expired in Oct, 2014, DAL experienced a surge of airlines scheduling new departures that have not been seen and/or supported by the past passenger flows. These are supply-induced demand that are, generally speaking, results of the removal of restrictions. Airlines scheduling new flights from DAL represent a mix of many factors: passenger traffic at DAL; competitive factors between DAL and DFW and airlines' anticipation of spill-overs from DFW and so on. Supply side factors, i.e., airport competition, tend to influence many of these decisions and resultant passenger traffic often far surpasses the historical trends. In order to account for these surges, particularly in the short run, we calibrate forecast output of DAL to reflect this new reality of (b).

Summary of TAF-M process in the domestic NAS and beyond

The process map in the panel below thus concludes the entire forecasting process. In addition to the processes outlined and discussed in Sections 2, 3, and 4, now we add domestic cargo, international passengers, departures and cargo into TAF-M as add-ons onto main TAF-M database.



Two further calibrations are performed in the forecast data: first, by short-run airlines schedule and T100 data to account for the changes we observed (from T100) or sure to observe (from airline schedules) during the first 6 quarters; and second, by Agency's OPSNET or CountOps data. Finally, two algorithms, one taking into account effect of slot restrictions and the other taking into account supply-induced demand, are applied to LGA, DCA; and DAL respectively.

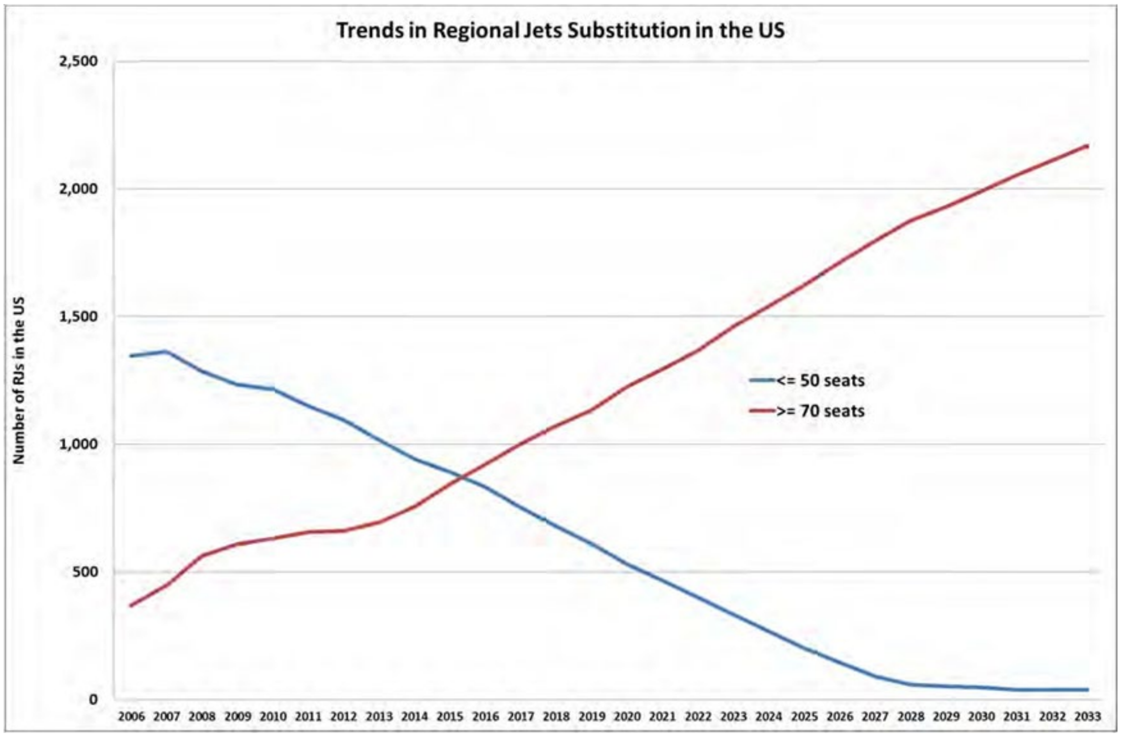
AIRCRAFT REALLOCATION IN TAF-M

This section describes the aircraft reallocation process, outlined in the earlier section "Reallocation of Aircraft by Type", in greater details. The section starts out by explaining the main concept that guides the reallocation process, followed by introducing the substitution rate calculation. The section after that lists the aircrafts that are being replaced and the aircrafts that are replacing with. Since the aircraft reallocation process is done at the airport level, we show an example of the reallocation algorithms at the Lafayette Regional Airport.

FAA's projections of regional jets (RJs)

In general, there is a substitution of larger regional jets from smaller RJs (< 50 seats). In Section 7, we will describe the steps that integrate FAA's analysis of RJs (i.e., stock of RJ aircraft and projections) into TAF-M (i.e., flow of departures by airports).

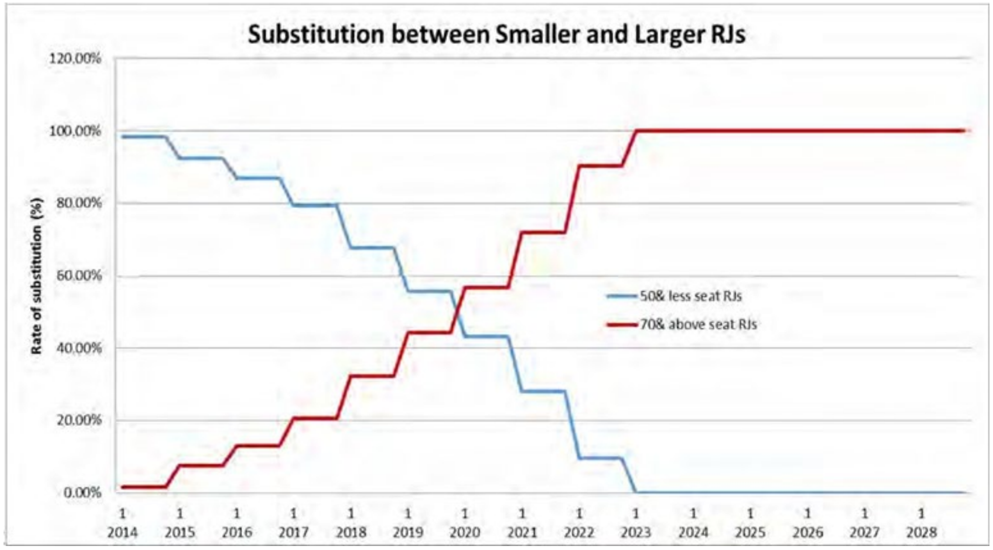
The FAA's analysis of operators changing 50-seat RJ's to 70+ (including 90-seat RJ's) can be summarized in the following figure:



Notice that the stock of smaller RJs has been going down and accelerates this trend around 2014-2016. The substitution into larger RJs is almost completed by 2028-2029. Beyond that point, it is safe to assume that RJs with 70-seats and above are driven primarily by passengers.

From stocks of RJs to rates of substitution in departures

Assuming fixed value/s of rates of utilization (i.e., in terms of hours and number of departures within an hour), the above stocks of aircraft can be translated into flow of departures. This substitution is captured by the following parameterization (i.e., essentially assumption) within the TAF-M:



The above parameterization is an excel file that we control at the macro level in order to slow down the substitution or accelerate it based upon FAA's annual national updates. In the present scenario of TAF-M, the above parameterization is in place where substitution begins slowly in 2015 and completed by 2023. The steps signify quarterly rates which do not change within a particular year (horizontal axis); it changes only annually. By construction, this parameterization was to follow the overall trend of the first graph. Table 2 summarizes the parameterization.

Table 2: Parameterization Schedule

Forecast ID	Year	Percent of 70-seat and above RJs	Year	Percent of 70-seat and above RJs
111	2015	7.5%	2028	100%
111	2016	13.0%	2029	100%
111	2017	20.7%	2030	100%
111	2018	32.2%	2031	100%
111	2019	44.3%	2032	100%
111	2020	56.7%	2033	100%
111	2021	71.9%	2034	100%
111	2022	90.4%	2035	100%
111	2023	100%	2036	100%
111	2024	100%	2037	100%
111	2025	100%	2038	100%
111	2026	100%	2039	100%
111	2027	100%	2040	100%
111	2023	100%	2036	100%

Replacement of smaller RJs by newer, larger RJs

The replacement routine is applied to a specific group of RJs. At present, there are 7 types of RJs that are replaced by 4 larger counterparts (identified by DOT-specified 3-letter numeric). Table 3 lays out the guide for the RJ replacement.

Table 3: Aircraft Replacement List

	Replacement of A/C						Replaced with A/C					
						seats						seats
461,464 with 673	461	EMB-120 E	EMBRAER	EMBRAER	EMB-120	30	673	EMBRAER	EMBRAER	EMBRAER	ERJ-175	86
	461	EMB-120 E	EMBRAER	EMBRAER	EMB-120	30	673	EMBRAER	EMBRAER	EMBRAER	ERJ-175	86
	464	EMBRAER	EMBRAER	EMBRAER	EMB-110	30						
	464	EMBRAER	EMBRAER	EMBRAER	EMB-110	30						
674,675, 676 with 677	674	EMBRAER	EMBRAER	EMBRAER	EMB-135	37	677	EMBRAER	EMBRAER	EMBRAER	EMB-17	88
	674	EMBRAER	EMBRAER	EMBRAER	EMB-135	37	677	EMBRAER	EMBRAER	EMBRAER	EMB-17	88
	675	EMBRAER	EMBRAER	EMBRAER	EMB-145	50						
	675	EMBRAER	EMBRAER	EMBRAER	EMB-145	50						
	676	EMBRAER	EMBRAER	EMBRAER	EMB-140	44						
	676	EMBRAER	EMBRAER	EMBRAER	EMB-140	44						
629 with 631	629	RJ-200ER/	BOMBARD	CANADAI	CRJ-2/4	50	631	CANADAI	CANADAI	CANADAI	RJ-700	70
	629	RJ-200ER/	BOMBARD	CANADAI	CRJ-2/4	50	631	CANADAI	CANADAI	CANADAI	RJ-700	70
628 with 638	628	CANADAI	CANADAI	CANADAI	RJ100/ER	56	638	CANADAI	BOMBARD	CANADAI	CRJ-900	86
	628	CANADAI	CANADAI	CANADAI	RJ100/ER	56	638	CANADAI	BOMBARD	CANADAI	CRJ-900	86

First important thing to notice in Table 3 is that all smaller RJs, i.e., “replacement of A/C” are replaced larger RJs by using Ed’s assumptions (i.e., “Replaced with A/C”). Notice also that although aircraft type 628 is above the 50-seat cut-off range, we decided to put it together with the 50-seat and lower as they too are expected to go out of service. The above (rate of) substitution along with the types of A/Cs are imposed at the level of airport (and not at segment level) and form the macro statement defining aircraft reallocation routine in TAF-M.

An important assumption for the reallocation process is that the passengers originating from any airport remain unchanged before and after the reallocation. If we further assume that load factor stays the same, the assumption implies that the capacity (available seats) remains the same. This is reasonable because the passenger growth driven by the market demand is already accounted for during the O&D passenger forecast stage (Section 1). Reallocation only changes the fleet mix so that the newly acquired 70 seat RJs, combined with the reduced fleet in the 50 seat RJs, can continue serving the same number of passengers.

APPENDIX: THE CASE OF LAFAYETTE REGIONAL AIRPORT (LFT)

The example below, focusing on airport LFT in 2015:Q1, demonstrates in details the reallocation algorithm. Year 2015 is the first year during the forecast periods for Forecast ID 111, the most current forecast.

forecastid	year	quarter	origin	aircrafttype	Dep	Pax	Seat
111	2015	1	LFT	674	413.9	12,298.1	15,313.6

The table shows that departures performed is projected to be 413.9 operations for aircraft type 674 (Embraer 135), carrying 12,298.1 passengers with a total seat count of 15,313.6. In reality, the numbers should be integers. We keep the decimal points here to maintain a consistent level of precision.

Incorporating reallocation

The reallocation routine proceeds in the following order: identification of the substitution rate, calculation of the remaining operations that will be performed by the smaller RJs, calculation of the additional operations that will be performed by the newer, larger RJs, and finally consolidating all records.

Identify the substitution rate

Based on Table 3, we learn that aircraft 674 at LFT will be replaced by 677. The rate of substitution is 7.5% in 2015 based on Table 2. We then conclude that only 92.5% of the 674 departures will be performed and the rest 7.5% will be replaced by 677. ting all records.

Calculate the remaining operations that will be performed by the 50-seat RJs

With the substitution rate of 7.5%, we can calculate how many departures, passengers, and seats will still be carried out by 674:

Remaining 674 departures

= Original 674 departures x (1 - substitution rate)

= 413.9 x (1 - 7.5%)

= 382.8

Remaining 674 passengers

= Original 674 passengers x (1 - substitution rate)

= 12,298.1 x (1 - 7.5%)

= 11,375.7

Remaining 674 seats

= Original 674 seats x (1 - substitution rate)

= 15,313.6 x (1 - 7.5%)

= 14,165.1

forecastid	year	quarter	origin	aircrafttype	Remainder_Dep	Remainder_Pax	Remainder_Seat
111	2015	1	LFT	674	382.8	11,375.7	14,165.1

Calculate the additional operations that will be performed by the 70-seat RJs

Based on the assumption that the total passengers and seats remain constant before and after the aircraft reallocation, the addition of 677 passengers and seats should be equivalent to the reduction of 674 passengers and seats.

Addition of the 677 passengers

= Reduction of the 674 passengers

= 12,298.1 x 7.5%

= 922.4

Addition of the 677 seats

= Reduction of the 674 seats

= 15,313.6 x 7.5%

= 1,148.5

Addition of the 677 departures

= Addition of the 677 seats/maximum seat capacity of the 677 aircraft (88 seats)

= 1,148.5⁄88

= 13.1

forecastid	year	quarter	origin	aircrafttype	Added_Dep	Added_Pax	Added_Seat
111	2015	1	LFT	677	13.1	922.4	1,148.5

Consolidate the records

After consolidating the records between aircrafts 647 and 677, there are now two records for airport LFT at 2015:Q1

forecastid	year	quarter	origin	aircrafttype	Dep	Pax	Seat
111	2015	1	LFT	674	382.8	11,375.7	14,165.1
111	2015	1	LFT	677	13.1	922.4	1,148.5

Chances are there might been existing 677 operations for airport LFT in 2015:Q1. It is not the case in this particular example. However, if there were existing 677 operations, we would need to re-calculate 7.5.2 and 7.5.3 so that the final output table is grouped by year, quarter, origin, and aircraft type.

To compare the pre-reallocation with the post-reallocation numbers, the table below shows that the number of 674 operations in 2015:Q1 at LFT is 413.9 before reallocation and 382.8 afterwards. On the other hand, the number of 677 operations at LFT is 0 versus 13.1 after reallocation.

forecastid	year	quarter	origin	aircrafttype	Dep	Pax	Seat	Dep_recalc	Pax_recalc	Seat_recalc
111	2015	1	LFT	674	413.9	12,298.1	15,313.6	382.8	11,375.7	14,165.1
111	2015	1	LFT	677	0.0	0.0	0.0	13.1	922.4	1,148.5