Building the Timetable from Bottom-Up Demand: A Micro-Econometric Approach

Dipasis Bhadra, Jennifer Gentry, Brendan Hogan, and Michael Wells
The MITRE Corporation's Center for Advanced Aviation System Development
7515 Colshire Drive, McLean, Virginia 22102

Keywords: Aviation, Decision Support Tools, Timetable, Econometrics, Forecasting

ABSTRACT

The aviation community has a rich collection of tools that simulate the operational flows of the National Airspace System (NAS). In nearly all cases, modeled operational flows of aircraft in the NAS begin with a schedule generated outside of the model. In the past, the schedule has been derived by translating the Federal Aviation Administration's (FAA's) Terminal Area Forecast (TAF) into flights. The downside to this, however, is that NAS operations are made up of specific airport-to-airport flows, which may be different from terminal area growth attributable to those airports. The challenge is to move from a generic traffic count at a specific terminal to a schedule of flights that includes a "when" and a "where" dimension.

Modeled NAS operational performance is highly dependent on the characteristics of the forecasted operations; hence it is critical that the traffic schedule be created correctly. The top-down approach based on TAF projections achieves its goal of replicating the intended volume of flights at each airport, but it does not necessarily achieve the desired operational-level integrity. In other words, the existing method is not capable of forecasting route-specific growth in operational flows.

At the MITRE Corporation's Center for Advanced Aviation System Development (CAASD), we are building a framework which attempts to fill in the gaps mentioned above using a bottom-up, demand-driven micro-econometric approach. Our ultimate goal is to produce a schedule of flights that is linked with origin and destination (O&D) operations via passenger route choice. It should thus be in sync with the Official Airline Guide (OAG), but not driven by it. Our method is comprised of six basic steps, beginning with estimation and forecasts of traveler demand between O&D city pairs, and culminating with the creation of a forecasted schedule that incorporates all major aspects of passenger demand.

INTRODUCTION

The aviation community has developed numerous tools for simulating the operational flows of the NAS[1] [1–3]. Some of these modeling capabilities are quite detailed in approximating the metrics they set out to depict. For example, CAASD's Detailed Policy Assessment Tool (DPAT) measures queuing delays occurring in the NAS throughout the various phases of flight. Taken together, these delays can reach significant levels on a bad weather day. Alternatively, other models have been developed that simulate airline schedule evolution to mitigate the effects of congestion. For instance the National Aeronautics and Space Administration Logistic Management Institute's (NASA/LMI) model provides airlines with a series of actions they can take in response to congestion, including depeaking, off-hours operations, use of secondary airports, and using larger aircraft [3].

In all cases, these tools must be fed a schedule of flights, or timetable, which is generated outside of the model. A typical timetable contains columns for the origin airport, destination airport, departure time, arrival time, equipment type, and possibly the carrier. (Note we use the term "timetable" because it also

[1]The NAS is a large network of airports and air traffic control facilities (ATC). ATC facilities are classified into three categories: airport towers, terminal radar approach control facilities (or, TRACONs), and air route traffic control centers (ARTCCs or, en route centers). Towers are located at airports and direct airport traffic on the ground and within approximately 5 nautical miles of the airport to altitudes of about 3000 feet. There are 496 towers, of which 266 are under FAA direct control and 230 are managed under contract. TRACON facilities sequence and separate aircraft as they approach and leave airports beginning approximately 5 nautical miles and ending approximately 50 nautical miles from the airport and at altitudes up to about 10,000 feet. En route centers control aircraft in transit and during approaches to TRACONs. The airspace that most en route centers control extends above 18,000 feet for commercial aircraft. At present, there are 22 en route centers.

estimates itineraries for unscheduled traffic.) Table 1 provides an example.

Timetable *input* strongly influences the resulting modeled *output*. For example, modeling an airport with only ten scheduled operations a day will produce drastically different results than when modeling the same airport with 1000 scheduled operations a day. The timetable determines the level and general directional flow of these operations. Using a realistic timetable of aircraft operations is therefore a critical component to the operational modeling effort.

In the past, timetables have been derived by extrapolating counts and forecasts of airport terminal operations into individual flights. The FAA measures overall NAS traffic in terms of annual operational counts at each terminal. It publishes a forecast of this traffic each year in the Terminal Area Forecast. Since the FAA forecasts these counts into the future, they make a logical data set to use for growing terminal area traffic, and thus growing any hypothetical schedules as well. The downside to this method, however, is that in reality NAS operations are made up of flows which are associated with a particular location and time of day (similar to a flight plan). This is not necessarily equivalent to extrapolating terminal area growth into individual operations. Thus, the challenge is to move from a generic traffic count at a specific terminal to a timetable of flights that includes a "when" and a "where" dimension.

The top-down approach described above achieves its goal of replicating the intended volume of flights at each airport (i.e., predicted TAF levels), but it does not necessarily achieve the desired operational level of integrity. In other words, the existing method is not capable of forecasting route-specific growth in operational flows. By simply matching terminal traffic forecasts with the OAG schedule (published schedule of flights submitted by commercial airlines), the existing method clearly misses out on rich route-specific information.

The information that is missing in this process is the origin and destination of the passengers on the flights, and information about the routes over which they fly. For example, if some city pairs are expected to experience above- or below-average growth, then some routes (and thus some specific airport pairs) will experience above- or below-average growth. A similar story can be told for hub airports, some of which may add capacity in the near future, and others of which may remain capacity constrained.

PROCESS DESCRIPTION

At CAASD, we are building a framework which attempts to fill in the gaps mentioned above using a bottom-up, demand-driven microeconomic approach

Table 1. Sample Timetable

DEP APT (step 2)	ARR APT (step 2)	EQUIP (step 3)	DEP TIME (step 4)	ARR TIME (step 4)
JFK	BOS	AEST	10:30	11:30
JFK	IAD	CRJ1	10:15	11:31
JFK	KIN	A343	10:15	14:05
JFK	LAX	B762	10:10	13:15
JFK	LAX	B763	10:30	13:23
JFK	MCO	B752	10:25	13:11
JFK	PAP	A306	10:00	13:53
JFK	PHX	A320	10:00	13:24

Frequency (step 4) **Block Time (step 4)**

[4–5]. Our ultimate goal is to produce a timetable of flights that is linked with O&D operations via passenger route choice and carrier equipment choice. Our output should thus be consistent with the OAG, but not driven by it. Our method is comprised of six basic steps.

The first step of the process lays the foundation upon which all the other steps will be built. This step determines where people ultimately want to travel. Once we know where people want to go, we use a logit model to determine how to get them there. This second step produces the actual segments that will be listed in our timetable. For instance, a person planning to travel from Seattle to New York may have a stopover in Chicago (see Figure 1). This process translates a single trip into two separate flights, one from Seattle to Chicago, and the other from Chicago to New York.

The third step determines what type of aircraft will be flown on each flight segment. For instance, the flight from Seattle to Chicago may require a different type of plane, because the distance from Seattle to Chicago is twice as great as the distance from Chicago to New York.

The remaining steps assign arrival and departure times to the flights and also take into account flight activity that is not driven by domestic passenger demand (i.e., cargo, international, and general aviation). While the methodology does not encompass all of the complexities airlines must account for when creating their schedules, it does attempt to capture the same passenger demand element that is the primary driver of their schedules.

METHODOLOGY

1. Estimating and Forecasting Domestic O&D Passenger Demand

People fly because they want to go to places for business and leisure reasons. These decisions are primarily driven by local economic and demographic characteristics. In addition, industry characteristics such as fare and market share of major carriers, seasonality, and the structure of airport hubs all play important roles in eventually determining the O&D demand. Differentiating the NAS by distances, we estimate a set of econometric relationships that define these relationships on O&D data [4].

To estimate those relationships, data from the Department of Transportation's (DOT's) Origin Destination Survey ("10% ticket sample") are matched with local economic and demographic information for each origin and destination airport.[2] Using this information and drawing on well-established econometric methodologies in demand estimation, we estimate O&D passenger demand between city pairs. In this framework, O&D demand is estimated based on local metropolitan variables as opposed to national economic and demographic conditions, and hence called bottom-up demand. Note this is an ongoing process; and as new data become available, we plan to re-estimate these relationships.

Finally, by combining these estimated relationships with commercially available forecasts of local economic and demographic variables, we end up with yield forecasts of passenger flows by O&D metropolitan areas.

2. Assigning O&D Passengers to Routes

We now have forecasts of passenger demand for travel between metropolitan areas. But choosing to travel somewhere is not the only decision a consumer must make. They must also choose how and when they will fly. Unfortunately, good data on passenger flows by *day* or *time* does not exist. However, the 10% ticket sample does have data on passenger routes. This is important, because over one-third of itineraries involve at least one connecting flight. Knowing that a certain number of people want to go from Seattle to New York is only part of the story, and obscures the fact that many of these passengers will change planes in a hub such as Chicago O'Hare or Dallas-Fort Worth. Furthermore, flights through these hubs are filled with passengers going to and from a variety of O&D Metropolitan Statistical Areas (MSAs).

As suggested by [6], we use the following process to convert O&D passenger flows into airport-to-airport (or "segment") passenger flows. First, using data from the 10% ticket sample, we estimate how route characteristics such as travel time, fare, and connections affect the likelihood of passengers choosing a given route from among the set of routes available. We do this using a multinomial logit model. Once the model is calibrated using historical data, the resulting equation is applied to each O&D pair of MSAs. The result is a distribution of O&D passengers among available routes.

After we have passengers assigned to routes, the model goes through all routes and sums passenger counts by segment. Figure 1 illustrates this using a stylized example. As you can see, when completed we have estimated quarterly passenger flows by airport pair.

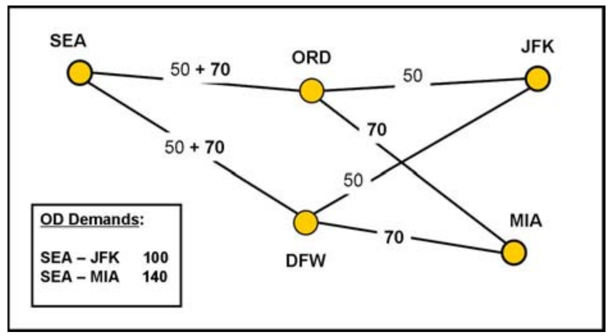

Figure 1. Example of Adding Segment Traffic

As with the calibration of the model, the routes available for passengers to choose from are those which are observed in the actual data. While this does preclude the model from "thinking outside the box" to determine other potentially feasible routes, the converse is also true—we do not end up with too many connecting flights going through small, non-hub airports.

3. Determining Aircraft Equipment Mix

Our third step is critical—taking the airport-to-airport passenger flows and translating total seat demand into the likely set of aircraft types that will fly each route. This step is necessary due to the variation in actual aircraft sizes, which implies that a given number of passengers do not uniquely determine the number of aircraft operations. Estimated passenger counts must be combined with estimated aircraft size to determine a likely equipment mix for each given route.

Choice of aircraft thus emerges as a function of passengers, frequency, trip distance, and other route characteristics. We can therefore estimate a multinomial logit model that enables us to determine the most likely choice of aircraft type. To do this, we turn to DOT's T-100 "Segment" data, which combines historical passenger counts with equipment type, along with flight characteristics such as distance.

An investigation into aircraft utilization over the last 5 years indicates that the most utilized aircraft in the NAS has been narrow-bodies, which transport

[2]Primary data for this analysis is based on the 10 % O&D sample obtained from the Bureau of Transportation Statistics (BTS) [see http://ostpxweb.dot.gov/aviation for details]. In addition, we use T-100 schedule data collected by the BTS. We combine the O&D travel data with local economic, demographic and spatial variables collected by the Bureau of Economic Analysis (BEA) (see [4] for more details).

approximately 60% of scheduled passengers. Narrow-body is a broad classification usually representing single aisle aircraft (e.g., 737 100 through 500 series; and A320). This class of aircraft has a seat range of 90–162 passengers; best cruise speeds at 550–625 miles per hour, and is observed to have been in maximum use for flights ranging from 500 to 750 miles. We have examined actual data and classified the vast majority of aircraft based on these characteristics. In all, there are five natural groupings of aircraft, or categories. Each category has particular performance and capacity characteristics. Category 1 primarily consists of turboprops, that on average, typically fly segments that are less than 250 miles (e.g., SF-340, ATR-42/72, etc.). Category 2 consists of regional jets (e.g., ERJs and CRJs) that fly an average distance of 250-500 miles between MSAs. Category 3 is made up of narrow-bodies (737-100 to 500, A320s, and 727-200, etc.) that fly an average distance of 750-1500 miles between city pairs. Category 4 is the narrow-bodies that tend to fly longer distances, on average 750-1500 miles (i.e., 737 700/LR, A330, etc.). The 5th category is the wide-body category that fly the long haul flights (e.g., 747, 757, 767, 777, L-1011, etc.). As noted, these distances are averages, and many aircraft within one category also travel distances defined under other categories. Each category is also associated with an average number of seats (passenger capacity) and a best cruise speed. Together, the five classifications account for more than 93% of all scheduled passenger activities (see Figure 2).

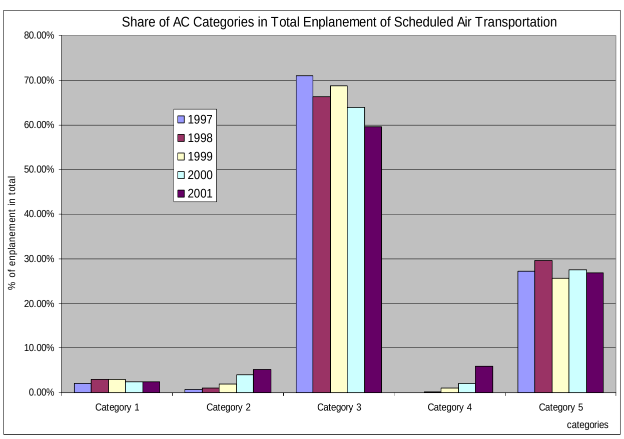

Figure 2. Aircraft Categorization

Modeling aircraft choice, based on historical passenger activities, is a tricky task, especially during a transitional time, like the one resulting from the terrorist attacks on September 11, 2001. The U.S. aviation industry is currently undergoing serious structural changes. At the end of 2002, more than 1400 aircraft were temporarily parked in the Mojave Desert. This is a relatively large percentage considering that the current aircraft inventory consists of only 3623 wide- and narrow-bodies, and around 1020 regional jets (RJs). Routes are being rationalized based on individual profitability in an attempt to improve aircraft utilization. Despite all these changes, the central element of scheduling flights remains intact: carry passengers between two points in the most efficient way.

4. Assigning Times for Scheduled Domestic Flights

Our fourth step determines when the flights will occur. The first part of this task is to determine exactly how many flights will occur. To do this we combine passenger movements between airport pairs (determined in step 2), with the aircraft that are predicted to fly between those airports (determined in step 3), and then apply a load factor. The load factor applied will be specific to each city pair and will be derived from BTS data. For example, if 1000 passengers are predicted to fly between LaGuardia and O'Hare on a given day, in a category 3 aircraft (that holds approximately 133 passengers) with a load factor close to 75%, the resulting frequency is 10 flights a day. This task is critical to the calibration of our model since the frequency calculation determines the total level of operations at an airport. By fine tuning the frequency calculation during the model validation stage, the overall number of operations in the timetable can be adjusted up or down to make it more accurate.

The next task is to take the actual flights and assign arrival and departure times. The timetable will contain commercial operations for 292 airports. The number of airports in the timetable with unscheduled or general aviation (GA) activity will be considerably larger (see step 6). Both commercial and unscheduled departure and arrival times will be assigned using historical data, when available.

Historical data for commercial traffic is obtained from the OAG. This data is then transformed into an arrival and departure distribution of operations over a given day. Several different "days" were obtained; one weekday, and one weekend, from each of the four quarters, for a total of eight representative days. These different "days" represent specific patterns in seasonal and weekday passenger travel.

Current baseline OAG operations will be used as the timetable starting point for scheduled times between city pairs. This data will be processed and altered to accommodate changes in forecasted equipment and international traffic. The historical airport operational distributions will be used to determine time assignment for additional flights. For example, if eight flights currently operate between LaGuardia and O'Hare each day, and ten frequencies are predicted, then the two additional flights will need to be assigned departure and arrival times based on historical data. The initial eight flights will receive the departure and arrival times that already exist in the current OAG schedule, with some minor adjustments.

To assist in the process of assigning departure and arrival times, the airports have been grouped into four tiers. The airports with the most dominant schedules were assigned to tier 1 (the FAA's capacity critical airports), and the airports with the least dominant schedules were assigned to tier 4 (typically airports only served a couple times a day by just one carrier), with the other airports falling somewhere in between. This categorization will help to determine a flight's scheduled arrival and departure times .

Once an arrival or departure time is established at an airport for a given flight, there is very little "slack" left in the schedule on the other end because the equation—departure time, plus block time[3], equals arrival time—is fairly tight. This causes a problem when a departure is created at an airport that then dictates an arrival at the destination airport during an unlikely time or vice versa.[4] To mitigate this problem, an airport's tier will be used to determine how strictly additional arrivals and departures should conform to the airport's historical distribution of flights. For instance, a hub airport traditionally has strong arrival and departure banks, thus additional flights should conform to those times when banks occur. On the other hand, adding flights to small spoke airports that have only a handful of operations should not necessarily comply with a historic distribution (i.e., If there are only five flights a day at an airport and a sixth is added, that sixth flight should not necessarily be added at a time when flights have historically occurred). Hence when scheduling between tiers, the more dominant airport's schedule will prevail. Intra-tier flights (i.e. flights to and from tier one airports) will undergo additional logic in order to find a departure and arrival time that fits with the airport's historical distributions. This preference will also be enforced by the order in which these flights are assigned in the timetable. Flights between tier 1 airports will be scheduled first, then flights between tier 1 and the remaining tiers. Next flights between tier 2 airports will be scheduled and so on, until all commercial flights have been scheduled.

5. Adding in Scheduled International Flights

Step five adjusts our tentative timetable by taking into account aggregate flows of international passenger and cargo traffic to and from the continental United States (CONUS). Although our modeling focus is the CONUS, we must still account for the additional terminal area traffic, especially important for the international gateway airports, that is generated by flights with only one of two cities in the CONUS. This will be accomplished using a modified top-down approach. All non-CONUS destinations will be associated with growth rates. The rates will then be applied to the number of seats currently being flown to or from those destinations. Applying the growth rate to the number of seats is in line with a passenger demand focus, and also allows for smaller increments of growth.

In 2000, around 26 million passengers traveled to the U.S. from around the world. While a majority (43%) of these passengers originated in Western Europe, the Far East had a respectable 29% share, followed by South America's share of 11%. Almost all of this traffic takes place through 11 gateway airports in the U.S., and therefore, greatly influences the schedule at those airports.

Unlike our O&D model for domestic air travel, here we plan to use regional growth rates from external agencies (e.g., U.S. Department of Commerce International Trade Agency (ITA), FAA) to drive our forecasts of international passenger travel. Using these forecasts and assuming the types of aircraft that are currently flown between these destinations, we can derive the forecast demand for scheduled departures and arrivals. This data will then be added to our domestic schedule.

6. Adding in Non-Scheduled Flights

The last step is to account for unscheduled, or GA traffic. Both the terminal and TRACON handle a large amount of GA traffic. It is estimated that for every scheduled flight, there is another one and half unscheduled operations [7]. There were an estimated 218,000 active GA aircraft in the NAS, which flew almost 40 million operations in 2000. Almost four-fifths of this traffic was in the domain of VFR[5] and thus less likely to crowd the en route air space. However VFR traffic impacts airport towers and TRACONs the same as IFR traffic. Given its significant utilization of NAS infrastructure, we must include a model of GA traffic in our timetable.[6]

[3] Block time, the time it takes to leave the gate at one airport and arrival at the gate of the destination airport, is largely a function of distance, winds and aircraft speed, but can also be highly influenced by taxi-in and taxi-out times at the arrival and departure airports respectively.

[4] Some time frames will also be unavailable due to constraints on airport operating hours.

[5] VFR stands for Visual Flight Rules. All commercial aircraft are required to fly instrument flight rules (IFR). GA or unscheduled traffic can fly VFR or IFR,

[6] Notice that a GA timetable is somewhat *fictitious*. GA traffic, for all practical operational purposes, is unscheduled traffic and hence does not produce timetable on its intent to fly between cities at a given time. It is worth noting here that published schedules are different than the flight plans that IFR GA flights are required to submit. Creation of a schedule, based on their behavior modeled using economic and/or other logic, runs the risk of being truly unreal. Nonetheless, we proceed with this method because of our need to model this entity in simulations of the NAS where they compete with

GA traffic is comprised of many different types of operators. Unscheduled business operators tend to file and fly IFR flight plans. Like in the case of O&D travel, particularly the upper end or premium fare travel, IFR flights can be sensitive to economic or financial factors. Specific location and time data for these flights can be derived using the FAA's Enhanced Traffic Management System data.

Data on GA traffic that file and fly VFR flight plans as well as military traffic are not time and location specific, and thus individual operations must be derived using the top-down approach described in the introduction. Based on the historical trends and economic factors, composite growth rates will be applied to both VFR and IFR operations to produce forecasts of activity.

CONCLUSION

This framework is being used to develop a timetable of aircraft operations that will support modeling efforts in evaluating NAS performance. In addition, by changing the inputs, this framework can be used to perform various types of "what if" policy analysis, and thus can stand on its own as a useful analytical tool.

REFERENCES

1. Wieland, F. 1999. "The Detailed Policy Assessment Tool (DPAT)," MITRE Technical Report MTR99W00000012, McLean, VA: The MITRE Corporation.
2. Wieland, F., C. Wanke, B. Niedringhaus, and L. Wojcik. 2002. "Modeling the NAS: A Grand Challenge for the Simulation Community." Proceedings of the Grand Challenges in Modeling and Simulation Conference, San Antonio, TX: 2002 SCS Western Multi-conference.
3. Long, D., D. Lee, E. Gaier, J. Johnson, and P. Kostiuk. 1999. "A Method for Forecasting Commercial Air Traffic Schedule in the Future," NASA/CR-1999-208987, Hampton, VA: NASA/LMI, Langley Research Center.
4. Bhadra, D. 2003. "Demand for Air Travel in the United States: Bottom-Up Econometric Estimation and Implications for Forecasts by O&D Pairs," *Journal of Air Transportation* (forthcoming); paper also presented at the 6th Air Transport Research Society Conference (ATRS 2002) hosted by the Boeing Corporation, July 14 – 16, 2002, Seattle, Wa; and AIAA/ATIO Technical Forum, October 1–3, 2002, Los Angeles, CA.
5. General Accounting Office. 2002. "Commercial Aviation: Air Service Trends of Small Communities Since October 2000, Report to Congressional Reporters," GAO Report No. GAO-02-432, Washington, DC.
6. Oppenheim, N. 1995. *Urban Travel Demand Modeling: From Individual Choices to General Equilibrium*. New York: John Wiley and Sons
7. Federal Aviation Administration. 2003. "Aerospace Forecasts: Fiscal Years 2002-2013," Washington, DC: U.S. Department of Transportation.

---

scheduled commercial and non-commercial traffic for scarce air space resources.

Paper forthcoming (July, 2003) in the *Annual proceedings of the Summer Computer Simulation Conference (SCSC'03).* A version of this paper was also presented at the NEXTOR-FAA *Conference on Air Traffic Management and Control*, Virginia Tech Graduate Center, June 2-3, 2003.