https://ntrs.nasa.gov/search.jsp?R=20180002161 2018-07-23T13:00:29+00:00Z

AIAA Aviation
13-17 June 2016, Washington, D.C.
16th AIAA Aviation Technology, Integration, and Operations Conference

Analyzing Double Delays at Newark Liberty International Airport

Antony D. Evans*
University of California Santa Cruz, Moffett Field, CA, 94035

and

Paul Lee†
NASA Ames Research Center, Moffett Field, CA, 94035

Downloaded by NASA AMES RESEARCH CENTER on August 17, 2016 | http://arc.aiaa.org | DOI: 10.2514/6.2016-3456

When weather or congestion impacts the National Airspace System, multiple different Traffic Management Initiatives can be implemented, sometimes with unintended consequences. One particular perceived inequity that is commonly identified is in the interaction between Ground Delay Programs (GDPs) and time based scheduling of internal departures by the Traffic Management Advisor (TMA) (now operationally superseded by the FAA's the Time-Based Flow Management system). Internal departures under TMA scheduling can take large GDP delays, followed by large TMA scheduling delays, because they cannot easily fit into the arrival flow at the runway. In this paper we examine the causes of these 'double delays' through an analysis of arrival operations at Newark Liberty International Airport (EWR) from June to August 2010. TMA scheduling delays are found to be generally higher than TMA airborne metering delays, regardless of prior GDP delays. Depending on how the double delay is defined, between 42% and 62% of all internal departures in GDP and TMA scheduling experienced double delays in this period. A deep dive into the data reveals that contributors to double delays include upstream flights departing before their Expect Departure Clearance Times (EDCTs); differences in the rates used for setting EDCTs and TMA Scheduled Times of Arrival; differences in the arrival demand expected based on EDCTs and the arrival demand entering TMA; and shorter en route times between takeoff and entry into TMA than assumed in the calculation of flight EDCTs, all of which undermine the sequencing and spacing underlying flight EDCTs. Double delays are also found to coincide with periods in which the virtual runway arrival queue being served by a TMA is large, there are periods of high demand relative to capacity, and there are high airborne metering delays. Data mining techniques are used to confirm that each of these factors contribute to the occurrence of double delay and/or high internal departure scheduling delay across three months of data from June to August 2010. Predictors of the occurrence of double delay and high TMA scheduling delay are built using logistic regression, providing prediction accuracies of 69% and 73%, respectively.

I. Introduction

ON a day with significant convective weather or widespread congestion within the National Airspace System (NAS), traffic managers in different facilities will typically issue multiple Traffic Management Initiatives (TMIs) to deal with the various problems they face. This means that many flights will be impacted by multiple different TMIs. For example, as described in Ref. 1, a flight may be held on the ground through a Ground Delay Program (GDP), before being rerouted around a thunderstorm, and then subject to Miles in Trail (MIT) as it passes through a congested sector. Three different traffic managers may implement these three initiatives in order to deal with specific problems in the airspace for which they are responsible. However, the joint impact of all three initiatives together may not be well coordinated, and there may be perceived inequities in their implementation, at

* Associate Research Scientist, University Affiliated Research Center, M/S 210-6, Senior Member.
† Senior Research Associate, Human-Systems Integration Division, M/S 262-4.

This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

least in part because tools currently available to traffic managers make good coordination difficult[1]. It may therefore be possible for the efficiency and equity of the Traffic Flow Management (TFM) procedures to be improved.

One concept that is under development by NASA that has the aim of improving the efficiency of TFM operations is Region Trajectory Based Operations (Region TBO). Region TBO is an integrated concept for operators and service providers to collaboratively organize aircraft trajectories into well-managed flows that match traffic demand to available capacity. The concept is to leverage FAA and NASA pre-departure, en route and arrival technologies, improving system predictability and equity of impact in the face of various airspace, airport, and weather constraints without compromising safety. The project will focus on improving arrival operations into the New York Metroplex through the integration of new and existing local-, regional- and national traffic flow initiatives. The reason for the focus on New York is that air traffic operations in and around the New York Metroplex are particularly complex, with high traffic demand, a highly constrained airspace and uncertainty from factors such as weather. These combine to create a challenging operational environment resulting in chronic congestion and delays that have impact across the NAS.[2]

One particular perceived inequity in the application of multiple TMIs, commonly identified by traffic managers and airline personnel, is in the interaction between GDPs and time based scheduling of internal departures[1] by the Time-Based Flow Management (TBFM) system, formerly the Traffic Management Advisor (TMA)[‡,3]. A GDP is a TMI implemented by the FAA to control the flow of aircraft into an airport by delaying flights destined for that airport at their respective origin airports based on distance (scope) to the constrained airport. A GDP is typically implemented for a period of time when increased aircraft spacing is considered necessary between landing aircraft, to ensure safety, and is often associated with adverse weather and/or high demand situations in which demand exceeds capacity. A GDP does this by specifying Expect Departure Clearance Times (EDCTs) for all flights destined for the constrained airport, within the specified scope and duration of the GDP, based on the predicted capacity of the destination airport. Because aircraft are held at their origin airports, a GDP is a strategic tool, controlling aircraft on the order of hours before they actually land at the constrained airport.

TBFM, and formerly TMA, serves a related function to a GDP, controlling arriving flights into an airport, but does so tactically. TBFM allows control of arriving aircraft that enter the destination Air Route Traffic Control Center (ARTCC or commonly 'Center') from an adjacent Center or departing from feeder airports close to the destination airport. TBFM sets up arrival schedules to the arrival airport runway and meter fixes on a first-come, first-served basis, independent of the TMIs that were initiated upstream. TBFM assigns scheduled times of arrival (STAs) at meter fixes for each arrival and "freezes" the STAs at the freeze horizon, from which the controllers then manage each aircraft's estimated times of arrival (ETAs) to conform to their STAs. This mechanism is called *TBFM metering*.

For aircraft whose departure airports are located inside the TBFM freeze horizon (referred to as internal departures), TBFM imposes additional constraints. TBFM is used to fine tune the arrival demand by scheduling departure times for the internal departures only when there are open slots at the runway and in the overhead stream. This mechanism is called *TBFM scheduling*. Note that TBFM scheduling and TBFM metering to an arrival airport can be run separately or concurrently. Airborne flights metered by TBFM typically absorb less than about 7 minutes of delay[4]. More delay can be absorbed through airborne holding, but this is undesirable because of the high cost of absorbing delay in the air, so is avoided. In contrast, because TBFM scheduled internal departures incur their delay on the ground, at lower cost, they have lower priority and can be delayed much longer.

The observation that is commonly made by traffic managers and airline personnel is that a GDP and TBFM can interact to give unfairly long delays for internal departures. This happens when a flight is given a significant GDP delay, followed by a significant TBFM scheduling delay. This is sometimes referred to as a 'double delay' or 'double penalty'. Because these flights are not airborne and the delay is incurred on the ground, TBFM scheduling delays are not restricted, as are airborne metering delays. The result is that while the delayed internal departures receive similar GDP delays to non-internal flights, their TBFM delays may be significantly larger.

Based on discussions with Subject Matter Experts (SMEs) double delays are thought to arise when the slot at the runway or in the overhead stream into which an internal departure is scheduled to enter, based on its GDP generated EDCT, no longer exists. This is despite the fact that EDCTs are calculated in order to meet the expected runway arrival rate, applying a first-come, first-served order based on scheduled arrival times[5] (although airlines can make substitutions and revisions, subject to constraints). Instead of delaying all the flights in the arrival stream in order to

[‡] TBFM is a direct development of TMA, with improvements in TMA's core time based metering capability and its trajectory modeler, an expansion of TMA and its departure capabilities to additional locations, and enhancements to TMA's departure capabilities[3].

Downloaded by NASA AMES RESEARCH CENTER on August 17, 2016 | http://arc.aiaa.org | DOI: 10.2514/6.2016-3456

open up a slot for the internal departure, the internal departure is typically held on the ground until a slot becomes available in the arrival stream into which it would fit[§]. For this reason, the Region TBO concept includes mechanisms to control the flow entering TBFM to ensure that gaps still exist in the arrival stream for internal departures. There could be a number of reasons for a slot not being available in the arrival stream, including poor EDCT compliance from the flights entering TBFM airborne metering, a GDP arrival rate that was set too high (e.g., the capacity at the destination airport deteriorated more than expected relative to when the GDP was put in place), shorter en route times between takeoff and entry into TBFM than assumed in the calculation of flight EDCTs, other ATC actions upstream of TBFM that have impacted the airborne flights, and because the GDP plan is simply not achievable by TBFM because, e.g., flights are concentrated over one meter fix.

A number of studies exist on this topic. A comprehensive initial analysis into the interaction between different TMIs, including specifically GDPs and TMA, was carried out by the Volpe National Transportation Systems Center in 2011[1]. This study, at the national level for the period from June to August 2010, compared distributions of EDCT compliance for the examined flights in GDP and scheduled by TMA, and in GDP but not scheduled by TMA. While a majority of flights not in TMA were EDCT compliant (departing within ±5 minutes of the EDCT) and were fairly symmetrically distributed around zero, flights in TMA scheduling showed significantly poorer compliance, with a skewing towards departing after the EDCT, suggesting an additional delay for the TMA scheduled aircraft. This positive delay beyond the EDCT in TMA scheduling allowed the amount of double delay to be estimated. Ref. 6 presents a detailed study of the use of the Traffic Management Advisor Flow Program, a type of GDP intended to reduce demand for a capacity constrained airport that is running TMA airborne metering and internal departure scheduling. That study uses fast-time simulations to explore the impact that Traffic Management Advisor Flow Programs have on arrival operations, and to identify situations in which flights receive double delays. The experiments clearly illustrated an imbalance in the distribution of delays assigned to internal departures and non-internal departures. Ref. 7 addresses impacts of multiple TMIs by computing and presenting candidate strategic TFM plans in the framework of a decision support tool. Fast-time simulation is used to explore combinations of one GDP, one Airspace Flow Program (AFP), and two re-route advisories in Atlanta Center (ZTL). Ref. 8 presents an analytical approach that quantifies the interactions between multiple TMIs based on a Brownian Motion formulation. The model developed takes demand and capacity estimates as inputs and estimates the delay distribution associated with the TMI controlling the demand. This is then used to estimate the probability of over-controlling or under-controlling the flow for the selected TMI parameters. Interactions between multiple TMIs are characterized by evaluating the probability that one TMI over- or under-controls the flow to a downstream TMI. Ref. 9 analyzes the nature and magnitude of additional delays experienced by flights subject to ground holding, finding that at four of the five airports studied, arrivals subject to ground holding incurred higher gate-arrival delays than arrivals not subject to ground holding.

While Ref. 1 and 6 provide an indication that double delays are a problem in the NAS, it is still not clear what the underlying causes of such double delays are, and under what circumstances they occur in real operations. Without this understanding, it is difficult to develop a concept that would effectively manage them, making for more equitable TFM operations. The objective of this paper is to identify the underlying causes of double delay and circumstances in which they occur in real operations, so as to be able to develop a concept that is able to better manage them. This is done through an analysis of arrival operations into Newark Liberty International Airport (EWR), one of the primary airports within the New York Metroplex which is significantly impacted by GDPs and TMA metering and scheduling, using data from June to August 2010. This is the same data used in Ref. 1, and was provided for our use by the Volpe National Transportation Systems Center. It represents the busy summer period when most convective weather impacts air traffic in the continental United States. Only 2010 data were available at the time of writing. Section II quantifies the double delay problem at EWR, followed in Section III by the approach to identify the causes and circumstances surrounding double delays. This is followed by results and a discussion in Section IV, and conclusions and recommendations in Section V.

[§] A functionality exists in TMA (a check box) that, if activated, does force internal departures to be slotted into the arrival stream ahead of airborne flights entering TMA after it. However, this is not typically used because of the negative impact of airborne metering delays.

Downloaded by NASA AMES RESEARCH CENTER on August 17, 2016 | http://arc.aiaa.org | DOI: 10.2514/6.2016-3456

Downloaded by NASA AMES RESEARCH CENTER on August 17, 2016 | http://arc.aiaa.org | DOI: 10.2514/6.2016-3456

II. Quantifying the Double Delay Problem

A. Are Internal Departures Inequitably Delayed?

As described by Ref. 1, TMA can improve arrival operations when a GDP is in place for three reasons. Firstly, a GDP is only able to apply its controls at departure. The rate at which aircraft then arrive at the destination airport is impacted by uncertainties in departure times and en route times. This can result in groupings of flights that exceed the airport arrival rate. TMA is specifically designed to smooth out these groupings in an efficient way. Secondly, TMA has a more detailed model of the arrival airspace than the Traffic Flow Management System (TFMS), which calculates EDCTs, and is able to meet constraints at both meter fixes and the runway, accounting for different separation requirements of different aircraft weight classes. TMA is thus able to fine tune the flow to better take advantage of the available capacity (both at the runway and in the terminal area). And thirdly, TMA has a more accurate estimate of the actual capacity at the runway, because the control takes place much closer to the runway. In contrast a GDP must use forecast runway arrival rates hours in advance. So the fine tuning that is possible in TMA results in a flow that more accurately meets the actual available capacity than is possible with a GDP alone.

In Ref. 1, average delays at different stages of the departure process were extracted for internal departures (flights departing from airports eligible for TMA scheduling) and non-internal departures (flights departing from airports not eligible for TMA scheduling) in TMA and GDP. These include:

- Total Departure Delay, calculated as: $D_{Total} = t_{Act} - t_{Sched}$
  t_{Act} refers to the actual flight departure time, and t_{Sched} to the original scheduled departure time[**]. Total Departure Delay can be decomposed into the following delays components:
- Pre-GDP Delay, calculated as: $D_{Pre-GDP} = t_{Ready} - t_{Sched}$
  t_{Ready} refers to the time when the flight is ready to depart, but before it is included in a GDP and issued an EDCT. Pre-GDP delay therefore includes delays due to e.g., mechanical issues with the aircraft. This delay is only valid for flights in GDP.
- GDP Delay, calculated as: $D_{GDP} = t_{EDCT} - t_{Ready}$
  t_{EDCT} refers to Expect Departure Clearance Time. GDP delay therefore refers to the delay assigned by the GDP, including substitutions, revisions and adaptive compression. This delay is only valid for flights in GDP.
- Post-GDP/Pre-TMA Delay, calculated as: $D_{PostGDP/PreTMA} = t_{TMA\ Ready} - t_{EDCT}$
  $t_{TMA\ Ready}$ refers to the TMA ready time, which is the departure time requested by the tower controller of the Center Traffic Management Coordinator, representing the earliest time that the flight would be able to take off. Post-GDP/Pre-TMA Delay therefore refers to additional internal departure delay after the EDCT, but prior to being scheduled in TMA. This delay is only valid for flights in GDP and TMA scheduling.
- TMA Scheduling Delay, calculated as: $D_{TMA\ Sched} = t_{TMA\ STD} - t_{TBFM\ Ready}$
  $t_{TMA\ STD}$ refers to the TMA scheduled departure time. TMA scheduling delay is therefore the scheduled delay assigned by TMA to an internal departure. This delay is only valid for flights in GDP and TMA scheduling.
- Take-Off Delay, calculated as: $D_{Take-off} = t_{Act} - t_{STD}$
  t_{STD} refers the assigned departure time. For flights in TMA scheduling, this refers to the flight departure time scheduled by TMA ($t_{TMA\ STD}$). For flights not in TMA scheduling, it refers to the flight EDCT (t_{EDCT}).

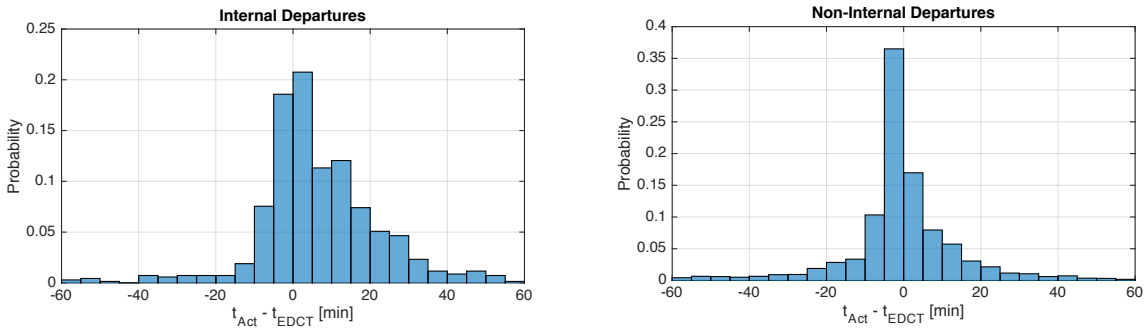
Using the same data set used by Ref. 1, which includes flight data and multiple TMI data (including GDP and TMA data) for the period June 1, 2010 to August 31, 2010 (TMA was then referred to as the Traffic Management Advisor), we calculate averages of each of these delays, as in Ref. 1, for departures destined for EWR. The data is also filtered for only those flights with post-GDP/pre-TMA delays and takeoff delays between -60 minutes and +60 minutes. This filtering is required because a small number of flights were identified with unrealistically high (or highly negative) post-GDP/pre-TMA delays or takeoff delays, which skew the results. In most of these cases the TMA times recorded are inconsistent with the GDP times, and are likely because the flight was departing near the start of the GDP, and may therefore have already been in the departure queue when it was given its EDCT, or because of issues with the data. The results are shown in Table 1 below.

[**] If a flight is not scheduled (e.g., general aviation), the original departure time filed with the flight plan is used instead.

Table 1. Comparing average departure delays per flight for internal departures in GDP and TMA scheduling, and non-internal departures in GDP, for flights destined for EWR from June and August 2010.

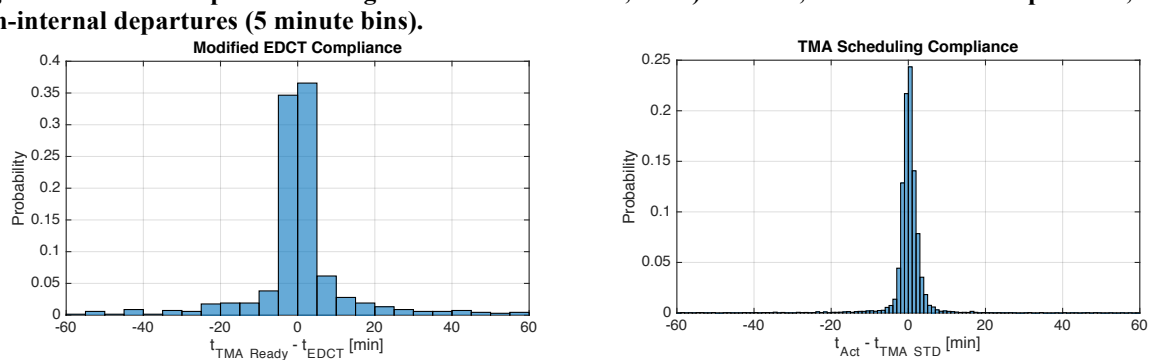
Category	Flight Count	D_{Total} [min/flt]	$D_{Pre\text{-}GDP}$ [min/flt]	D_{GDP} [min/flt]	$D_{PostGDP/PreTMA}$ [min/flt]	$D_{TMA\ Sched}$ [min/flt]	$D_{Take\text{-}off}$ [min/flt]
Internal Departure	602	64.6	4.4	51.2	-0.3	10.6	-1.4
Non-Internal Departure	1,299	53.1	8.0	47.1	-	-	-2.0

Table 1 shows that internal departures destined for EWR have departure delays, during periods under GDP and TMA scheduling, that are on average higher than non-internal departures – by more than 10 minutes. This appears to be predominantly the result of delay assigned by TMA scheduling ($D_{TMA\ Sched}$). The results also suggest directly that TMA scheduling adds departure delay compared to that of flights not subject to TMA scheduling. While the increase in delay is not large, on average, the standard deviation in $D_{TMA\ Sched}$ is 13.3 minutes, indicating that some TMA scheduling delays are significant. This is confirmed in Figure 1, which shows distributions of EDCT compliance (calculated as the difference between t_{Act} and t_{EDCT}) for internal, TMA scheduled departures (Figure 1a), and non-internal departures (Figure 1b). In both cases, zero represents EDCT compliance. The distribution of EDCT compliance for internal, TMA scheduled departures (Figure 1a) shows a notable skewing towards positive EDCT compliance (a delay beyond the EDCT), in some cases approaching 1 hour. This skewing cannot be seen for non-internal departures (Figure 1b), which shows a more symmetrical distributions around zero. However, the EDCT compliance shown in Figure 1a does not represent 'true' EDCT compliance, because it also includes departure delay associated with TMA scheduling. By modifying the EDCT compliance calculation for internal departures based on the difference between the TMA ready time and EDCT (Figure 2a), we see a tight symmetrical distribution around zero, indicating that the skewing in Figure 1a is directly associated with TMA scheduling delay. Compliance to the TMA scheduled departure time, calculated as the difference between actual departure time and TMA scheduled departure time, is shown in Figure 2b, showing an ever tighter distribution.



(a) (b)

Figure 1. EDCT Compliance for flights destined for EWR, for a) internal, TMA scheduled departures, and b) non-internal departures (5 minute bins).



(a) (b)

Figure 2. a) Modified EDCT compliance for internal, TMA scheduled departures, calculated as the difference between TMA ready time and EDCT (5 minute bins); and b) TMA scheduling compliance, calculated as the difference between actual departure time and TMA scheduled departure time (1 minute bins).

Downloaded by NASA AMES RESEARCH CENTER on August 17, 2016 | http://arc.aiaa.org | DOI: 10.2514/6.2016-3456

As Ref. 1 suggests, were there no double delays, internal departures would be expected to have EDCT compliance distributions comparable to non-internal departures (Figure 1b). However, the perceived inequitability of double delays is also a function of the airborne TMA metering delays that non-internal departures incur. If airborne metering delays were comparable to the internal departure delays shown here, then the double delays incurred by the internal departures would not be considered inequitable – because both groups of flights would incur similar TMA delays on top of their GDP delays. However, this is not the case. As shown in Table 2, the average TMA scheduling delay for internal departures to EWR that are included in a GDP, during periods in which TMA scheduling and TMA airborne metering are active, is 10 minutes. In contrast, average airborne metering delay of non-internal flights during periods in which GDP, TMA scheduling and TMA airborne metering are all active is only 3.2 minutes. These delays are on top of average GDP delays across the same periods of 46.7 minutes. It is expected that TMA scheduled internal departures that are excluded from a GDP would incur higher departure delay, because a slot is not allocated to them. These results indicate that during GDPs, TMA scheduled internal departures included in the GDP are also penalized more severely than non-internal, airborne metered flights, despite the fact that they are allocated slots in the arrival flow. Even if the differences in cost of airborne and ground delay are considered, with airborne delay assumed to be twice as costly as ground delay[10], these internal departures can be perceived to be inequitably penalized. For periods in which only TMA scheduling and airborne metering are active, but not GDP, average TMA scheduling and metering delays are slightly lower at 9.9 min and 2.5 minutes respectively, as expected given that capacity is likely to be higher when a GDP is not active. The difference between the average TMA scheduling delay and airborne metering delay suggests that it is not only during GDPs that internal departures could be perceived to be inequitably penalized. In summary, internal departures do appear to be inequitably delayed, on average, whether subject to a GDP delay or not.

Table 2. Average Delays in TMA metering and scheduling for flights destined for EWR

	Avg. TMA Scheduling Delay of Internal Departures [min]	Avg. TMA Metering Delay of Non-Internal Departures [min]	Avg. GDP Delay [min]
GDP, TMA metering and TMA scheduling active	10.0	3.2	46.7
TMA metering and scheduling active, GDP inactive	9.9	2.5	-

Note that in the above results, the airborne metering results are calculated based on an estimation of when TMA metering was being used to actively meter aircraft. The data analyzed, described in detail by Ref. 1, specifies when TMA metering is turned on. However, based on feedback from SMEs, this does not always coincide with flights being actively metered. Sometimes TMA metering is turned on but is only used to enhance situational awareness, and not to actively meter aircraft. We hence use a hidden Markov model (HMM) to estimate when TMA metering is being actively used to meter aircraft, and when it is not. This is described in Appendix A. This model estimates that in 19% of the periods in which TMA is turned on, it is not used for active metering.

B. What Percentage of Flights are Double Delayed?

In the data analyzed, 6% of flights are under GDP, while 20% of these flights are internal departures under TMA scheduling. Therefore, only 1.25% of flights are eligible to be double delayed, being internal departures under TMA scheduling, and subject to a GDP. Quantifying the degree to which these flights experience double delays at EWR depends on the definition of double delay used. If double delay is defined as any TMA scheduling delay, for an internal departure included in a GDP, that is higher than the average airborne metering delay at the time, then 62% of internal departures under TMA scheduling and GDP are classed as double delayed in the data analyzed. This represents 420 flights over the three month period analyzed. If the definition of double delay is expressed based on thresholds, the results differ slightly. Based on feedback from SMEs, a flight can be considered to be double delayed if it receives more than 15 minutes of GDP delay, followed by more than 5 minutes of TMA scheduling delay. Using these thresholds, 42% of internal departures under TMA scheduling and GDP are classed as double delayed, which represents 284 flights over the three months analyzed.

If we consider all internal departures, not just those included in GDPs, we can quantify the occurrence of high TMA scheduling delay more generally. Internal departures under TMA scheduling represent 13% of all flights in the data analyzed. Defining high TMA scheduling delay as anything higher than the average airborne metering delay at the time, 56% of internal departures under TMA scheduling are classed as highly delayed, which represents 4,010

Downloaded by NASA AMES RESEARCH CENTER on August 17, 2016 | http://arc.aiaa.org | DOI: 10.2514/6.2016-3456

flights in the data analyzed. Using the 5 minute threshold, 39% of internal departures under TMA scheduling are classed as highly delayed, representing 2,780 flights in the data analyzed.

Regardless of definition, while the percentage of flights eligible for double delays is small, the percentage of internal departures under TMA scheduling and GDP experiencing double delay is high. However, the problem is more widespread than just during GDPs, as shown by the results for all internal departures under TMA scheduling. A higher percentage of flights are eligible for high internal departure scheduling delay, while the percentage of TMA scheduled flights that are highly delayed is again high. This illustrates the importance of developing a concept that prevents high TMA scheduling delays for internal departures, both in GDP and not in GDP. This requires a better understanding of how and why double delay, and more generally high TMA scheduling delay, occurs.

III. Approach

In order to understand how and why double delay, and more generally high TMA scheduling delay, occurs, two types of analysis are performed. The first is a 'deep dive', where days in which a large number of double delays occur are identified and examined in detail. This provides insight into what is happening when flights are double delayed, and allows us to identify features that may impact the occurrence of double delay and high TMA scheduling delay, more generally. The second analysis applies data mining techniques to the identified features over a larger data set in order to confirm which are the underlying drivers of double delay and high TMA scheduling delay more generally. This is done through the development of predictors of double delay and high TMA scheduling delay (classification type problems).

A. Deep Dive

For the deep dive, we extract days with a high number of double delays, based on a combination of the metrics described in Section II-B, i.e., flights for which the TMA scheduling delay is greater than the average TMA airborne metering delay at the time, and for which GDP delay > 15 minutes and TMA scheduling delay > 5 minutes. Timelines are then plotted for all flights in each of these days. These timelines show the following airport ETAs, for each flight, as shown in Figure 3:

- ETA_{FS}: The initial estimated time of arrival at the runway, derived from the first flight plan information message. This should correspond most closely to the expected arrival time based on the flight schedule.
- ETA_{EDCT}: For flights in GDP, the runway ETA based on the flight's EDCT (the final GDP generated runway controller time of arrival before departure). These times should therefore meet the forecast airport arrival rate used by the GDP.
- ETA_{D}: For non-internal flights in GDP, an approximation of the runway ETA based on the actual flight departure time. The difference between this time and the previous time is the flight's EDCT compliance. Not relevant to internal departures..
- $ETA_{TMA\ entry}$: The runway ETA when the flight enters TMA. For scheduled internal departures this is approximated as the runway ETA at the TMA ready time. For non-internal metered flights, it is the runway ETA when the flight becomes active in TMA.
- STA_{Freeze} / ETA_{SchedD}: The runway STA accounting for delays scheduled by TMA. For non-internal metered flights, this is the runway STA at the freeze horizon (STA_{Freeze}). For scheduled internal departures it is the runway ETA at the TMA scheduled departure time (ETA_{SchedD}). The difference between this time and the previous time is the TMA scheduled delay (either on the ground for internal departures, or in the air for non-internal flights).
- ETA_{MF} / ETA_{ActD}: The runway ETA accounting for actual delays resulting from TMA control. For non-internal metered flights, this is the runway ETA at the meter fix (ETA_{MF}). For scheduled internal departures it is the runway ETA at the actual departure time (ETA_{ActD}). The difference between this time and the previous time gives an indication of the extent to which the flight met its TMA scheduled delay.
- ATA: The actual runway time of arrival. The difference between this time and the previous time gives an indication as to how much further delay was incurred in the Terminal Radar Approach Control (TRACON).

An aggregate example of this timeline is presented in Figure 4 in Section IV-A. Based on the runway arrival times plotted in the timelines, runway arrival rates are also plotted for each of the days identified to have a large number of double delays. This gives an indication of the extent to which the rates differ, particularly around when double delays occur. These arrival rates are also compared directly to plots of average GDP, TMA scheduling and TMA metering delay, and an estimate of the TMA virtual runway queue size. The latter represents the total number of flights that are being processed by TMA at any given time, including all TMA metered and scheduled flights that

Downloaded by NASA AMES RESEARCH CENTER on August 17, 2016 | http://arc.aiaa.org | DOI: 10.2514/6.2016-3456

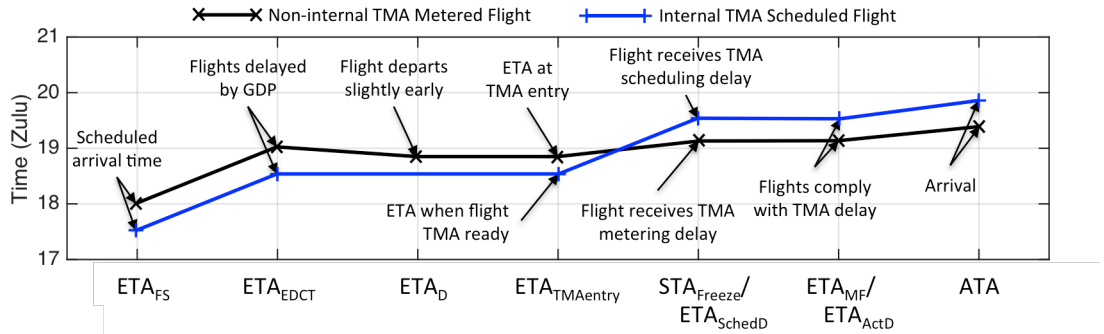


Figure 3. Sample timeline generated for two flights, one in GDP and TMA metering (a non-internal flight, in black), and one in GDP and TMA scheduling (an internal departure, in blue).

must be landed at the runway. TMA virtual runway queue size is calculated based on $ETA_{TMA\ entry}$ times, indicating when a flight enters the virtual queue, and ETA_{MF} / ETA_{ActD} times, indicating when a flight is processed by the queue. An example of these plots is shown in Figure 5 in Section IV-A.

B. Data Mining

Based on the results of the deep dive, discussed in Section IV-A, a number of features are identified as potentially having an impact on double delay. These include the following:

1. The ratio of destination arrival demand and destination arrival capacity in terms of AAR;
2. The destination arrival demand across the arrival fix to which an internal departure is routed;
3. Origin center Weather Impacted Traffic Index (WITI), as described by Ref. 11 and 12, and based on traffic data from representative low-weather-impact days in 2009 for each day of the week[††];
4. Destination center WITI;
5. Average airborne metering delay for flights arriving at the destination airport;
6. Average airborne metering delay for flights arriving at the destination airport that are routed across the arrival fix to which an internal departure is routed;
7. Average and standard deviation in EDCT compliance for arriving flights, measured as the difference between actual departure time and EDCT for non-internal flights (traditional EDCT compliance), or TMA ready time and EDCT for internal departures (modified EDCT compliance as defined in Section II-A);
8. Miles In Trail (MIT) in place for arriving traffic;
9. Whether an internal departure is in a reroute advisory (origin and destination airports are in an advisory);
10. Whether any reroute advisory is in place to the destination airport;
11. Virtual runway queue size experienced by TMA, as described in Section III-A;
12. Virtual runway scheduling queue size experienced by TMA, indicating the other internal departures that are TMA scheduled, but have not yet departed;
13. The change in estimated runway demand based on ETA_{ETD} and $ETA_{TMA\ entry}$ times. ETA_{ETD} is identical to ETA_{EDCT} in Section III-A. This feature indicates the degree to which estimated runway demand based on TMA ETAs has changed relative to expected demand before departure;
14. The change in estimated runway demand based on ETA_{ETD} (ETA_{EDCT} in this case) and STA_{Freeze} times. This feature represents a proxy for the difference between the airport capacity used to define EDCTs, and the airport capacity used by TMA;
15. The difference between the estimated runway demand based on $ETA_{TMA\ entry}$ times and the actual AAR operated at the airport;
16. The average and standard deviation of the change in airborne flight ETAs between ETD and TMA entry ($ETA_{TMA\ entry} - ETA_{ETD}$), indicating the degree to which there are differences between airline scheduled flight times used to generate EDCTs, and estimated times en route used for generating TMA ETAs.
17. The time since the start of the GDP; and
18. The time to the end of the GDP.

[††] Nov. 8 (Sun), Nov. 9 (Mon), Nov. 17 (Tue), Sep. 30 (Wed), Sep. 3 (Thu), Nov. 27 (Fri) and Nov. 21 (Sat).

Downloaded by NASA AMES RESEARCH CENTER on August 17, 2016 | http://arc.aiaa.org | DOI: 10.2514/6.2016-3456

The same feature set is also used to predict the occurrence of high TMA scheduling delay, with the exception of features 17 and 18, which are dropped because they are specific to GDPs. Some other features also differ slightly because many flights included in the training set for TNA scheduling delay are not under GDP:

- For feature 7, flights not in GDP are considered to have EDCT compliance of 0;
- For feature 13, the ETA_{ETD} for flights not in GDP is the ETA based on the last estimated time of departure (ETD) before take off;
- Feature 14 represents a proxy for the difference between the expected demand based on expected departure times and the airport capacity used by TMA; and
- Feature 16 indicates the degree to which there are differences between estimated times en route before departure and used to generate TMA ETAs.

One further feature is also included in predicting the occurrence of high TMA scheduling delay:

19. Whether the flight is subject to GDP or not.

Note that no features included describe the direct cause of high internal departure scheduling delay, which, based on SME feedback, is a lack of gaps in the arrival stream. This is because this effort aims to identify the underlying drivers of high TMA scheduling delay, as opposed to the direct cause. Future work will attempt to confirm that a lack of gaps in the arrival stream is indeed the direct cause of high TMA scheduling delay. This will require calculating expected internal departure delays were the only constraint to fit into the runway arrival stream, based on TMA runway STAs for all other flights.

In order to identify which features are the strongest drivers of double delay and high TMA scheduling delay more generally, a number of data mining algorithms, including logistic regression[13], decision tree[14], support vector machine[15], random forest[16], and k-nearest neighbors[17] were used to build predictors. The results of this analysis, including the performance of each model, are described in Section IV-B.

Because the objective of this exercise is the identification of drivers of double delay and high TMA scheduling delay, we remove features that are collinear or statistically insignificant. Colinearity is identified by calculating Variance Inflation Factors (VIF) for each feature, and removing features iteratively when VIF values exceed 10. Statistical significance is identified by calculating t-statistics, and removing features iteratively with t-statistics below 1.96 (representing a 95th percentile confidence interval). The results of this analysis, presented in Section IV-B, can then be compared to that of the deep dive and SME feedback.

IV. Results

A. Deep Dive into Days with Double Delays

Runway arrival timelines described in Section III-A were generated for EWR arrivals on June 28, 2010. Because a plot of all the flight timelines for this day is difficult to interpret, aggregate timelines are presented instead. Flight initial estimated times of arrival (ETA_{FS}) are subtracted from all times in each flight's timeline, allowing box plots to be generated for each time in the timeline, showing a single aggregate timeline for all flights considered. In Figure 4 we plot aggregate timelines for all non-internal flights on June 28, 2010 that are in GDP and in TMA airborne metering (Figure 4a), and all internal departures on June 28, 2010 that are in GDP and in TMA scheduling (Figure 4b). Associated arrival rates, virtual TMA queue size and average delays are shown in Figure 5. On June 28, 2010 a GDP was put in place from 16h30 to 03h00 Zulu (12:30am to 11pm EDT), while TMA metering was active from approximately 16h00 to 21h00 Zulu (12pm to 5pm EDT), and TMA scheduling was active from approximately 11h00 to 21h00 Zulu (7am to 5pm EDT).

As can be seen in Figure 4a by comparing ETA_{FS} and ETA_{EDCT} the median GDP delay for non-internal flights on June 28, 2010 is approximately 20 minutes. Similarly, for internal departures in Figure 4b, we see a median GDP delay of approximately 40 minutes. Median EDCT compliance can be calculated for non-internal flights by comparing ETA_{EDCT} and ETA_D in Figure 4a, which shows that most flights depart slightly early. Flights are considered EDCT compliant if they depart within ±5 minutes of their EDCT, so this is expected. For internal departures, modified EDCT compliance, as defined in Section II-A, can be calculated by comparing ETA_{EDCT} and $ETA_{TMA\ entry}$ (ETA_D is not relevant to internal departures), and shows similar results. By comparing ETA_D and $ETA_{TMA\ entry}$ in Figure 4a we see a median shift forward in time by as much as 10 minutes, suggesting that non-internal flights are making up time en route. This is likely to be because EDCTs are calculated by TFMS based on the ration-by-schedule algorithm[5] as a function of scheduled times of arrival filed by airlines, which may include schedule padding. Once the flight takes off, however, ETAs are calculated by TFMS based on current conditions, so the ETA may jump forward in time, as shown.

Downloaded by NASA AMES RESEARCH CENTER on August 17, 2016 | http://arc.aiaa.org | DOI: 10.2514/6.2016-3456

Downloaded by NASA AMES RESEARCH CENTER on August 17, 2016 | http://arc.aiaa.org | DOI: 10.2514/6.2016-3456

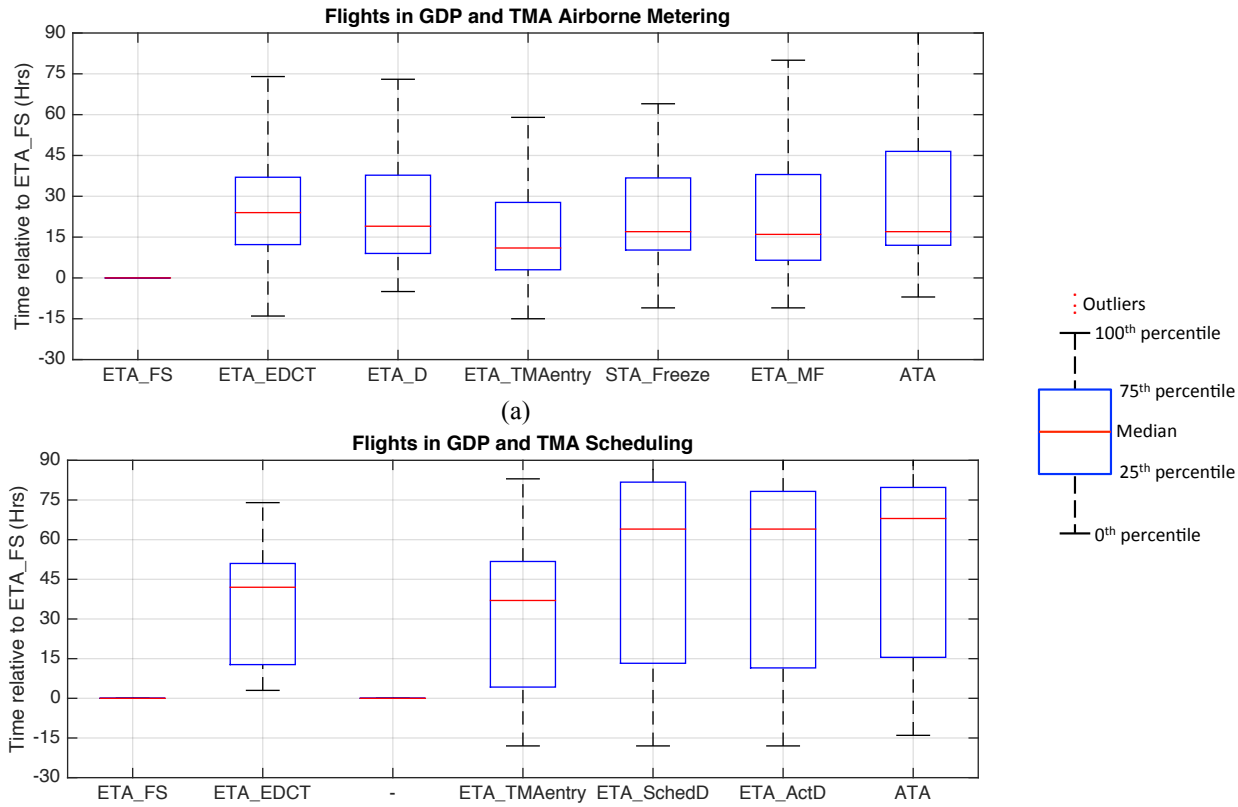


Figure 4. Aggregate timelines on June 28, 2010, for (a) non-internal flights in GDP and TMA metering, and (b) internal departures in GDP and TMA scheduling. Aggregate times relative to the initial runway ETA (ETA_{FS}) are shown using box plots for: runway ETA based on the flight EDCT (ETA_{EDCT}); runway ETA based on the actual flight departure time (ETA_D); runway ETA when the flight enters TMA ($ETA_{TMA\ entry}$); runway STA accounting for delays planned by TMA (STA_{Freeze} / ETA_{SchedD}); runway ETA accounting for actual delays resulting from TMA control (ETA_{MF} / ETA_{ActD}); and actual runway time of arrival (ATA).

Comparing $ETA_{TMA\ entry}$ and STA_{Freeze} in Figure 4a, we see a small median shift back in time by under 10 minutes, as expected due to TMA airborne metering, which typically absorbs only a limited amount of delay. Comparing $ETA_{TMA\ entry}$ and ETA_{SchedD} in Figure 4b, however, we see a significant median shift back in time for internal departures by nearly 30 minutes, indicating large scheduling delays. The median shift from STA_{Freeze} to ETA_{MF} for non-internal flights (Figure 4a), and from ETA_{SchedD} to ETA_{ActD} for internal departures (Figure 4b), is small, indicating that most flights incur the delay allocated to them by TMA. Similarly, the median shift from ETA_{MF} and ETA_{SchedD} to ATA is also small, indicating that relatively little extra delay is incurred in the TRACON, as expected.

These results, which are similar although less extreme if calculated across the whole three months of data analyzed, suggest that two contributors to double delays may be flights departing before their EDCT and differences between airline scheduled flight times and estimated times en route used to generate TMA STAs, both of which undermine the sequencing and spacing underlying the flight EDCTs.

Runway arrival rates based on $ETA_{TMA\ entry}$ and STA_{Freeze} / ETA_{SchedD} times are plotted in Figure 5a, approximating TMA demand and throughput. The AAR actually in place at the runway is also plotted, indicating the airport capacity. Note that this AAR may have not been the same as that used to calculate flight EDCTs, or TMA scheduling and metering delays, although in the latter case is likely to be reasonably close. These rates can be directly compared to the virtual TMA queue at the runway plotted in Figure 5b, and the average GDP, TMA scheduling and TMA metering delays in Figure 5c. Note that in Figure 5c, average delays are calculated based on the actual flight landing times (ATAs). This means that for flights delayed by TMA, there is an offset between the TMA arrival rate in Figure 5a and the corresponding delay in Figure 5c. This is especially true for internal departures under TMA scheduling, some of which are delayed by up to an hour.

Downloaded by NASA AMES RESEARCH CENTER on August 17, 2016 | http://arc.aiaa.org | DOI: 10.2514/6.2016-3456

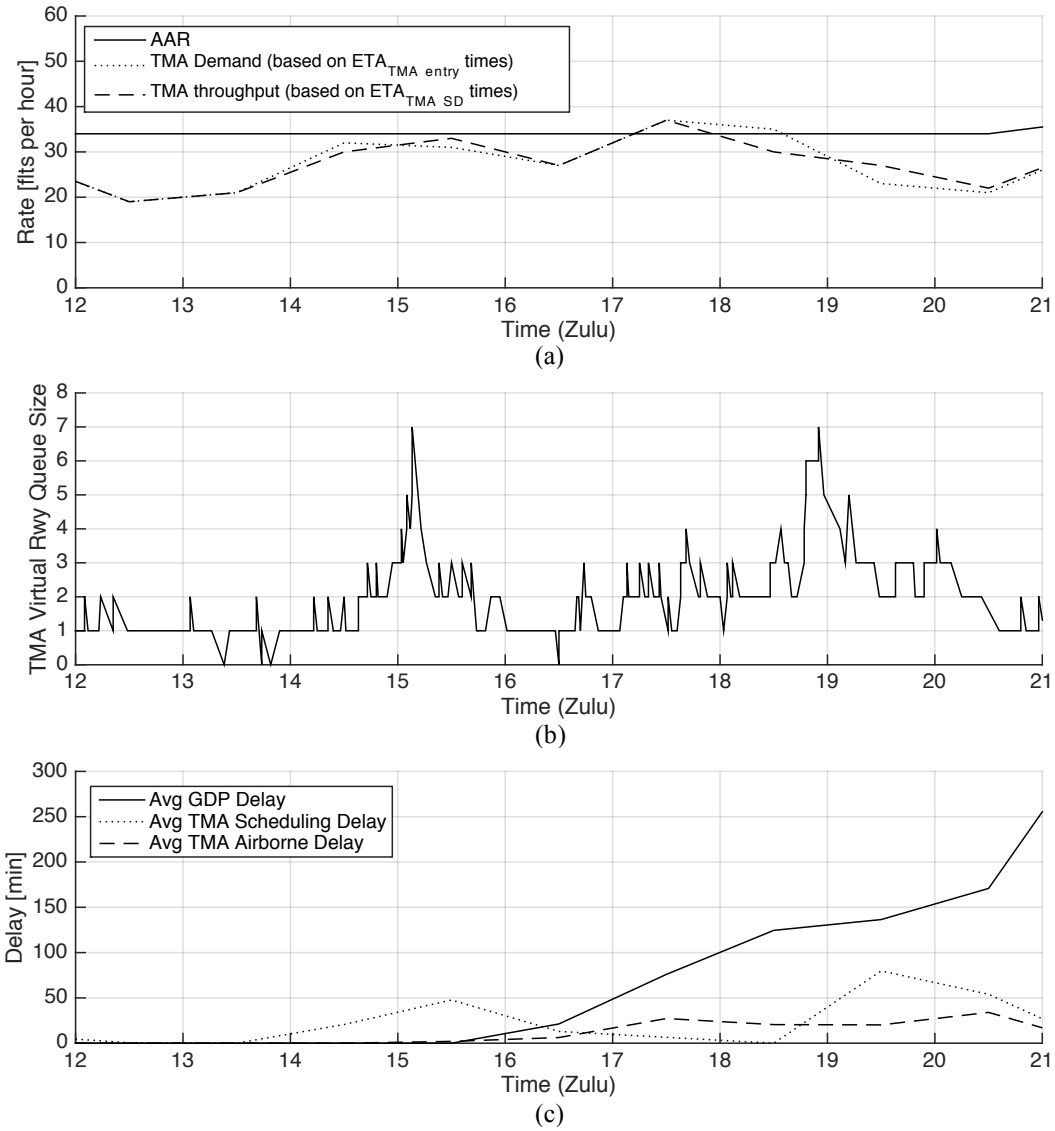


Figure 5. For EWR on June 28, 2010; (a) average runway arrival rates, (b) virtual TMA arrival queue at runway, and (c) average arrival delays.

In Figure 5a, TMA demand only exceeds the actual runway capacity (AAR) between 17h00 and 18h00 Zulu, but it approaches capacity between 14h00 and 16h00 Zulu as well. Between 14h00 and 15h00 Zulu TMA reduces the runway arrival rate slightly (shown by the lower TMA throughput), by delaying flights. This leads to an increase in the virtual TMA queue in Figure 5b just past 15h00 Zulu. The high average TMA scheduling delay for flights landing between 15h00 and 16h00 Zulu in Figure 5c suggests that it is primarily internal departures that are delayed to reduce the arrival rate. This is because TMA metering was not active until 16h00 Zulu. To deal with the delayed flights, between 15h00 and 16h00 Zulu TMA throughput is higher than TMA demand in Figure 5a, almost clearing the queue in Figure 5b by 15h30 Zulu. Between 17h00 and 18h00 Zulu, TMA throughput matches TMA demand, despite the fact that it is above the AAR. In TMA, runway arrival rates are set using a buffer, which defines a higher separation between arriving aircraft than required by the arrival separation matrix. A higher TMA throughput than AAR may be because the TMA buffer was set above the AAR, or because the particular mix of aircraft types arriving at the runway was such that more aircraft were able to land than normal. Between 18h00 and 19h00 Zulu TMA throughput drops below the AAR, while TMA demand remains just above it. The virtual TMA queue in Figure 5b consequently increases in size dramatically just before 19h00 Zulu. Accordingly TMA scheduling delay in Figure 5c for flights landing between 19h00 and 20h00 Zulu increases. TMA metering delays, however, do not change significantly, as there is a limit on the amount of delay that is typically absorbed in the air.

The above results suggest that double delays are associated with periods of high demand, and with periods in which TMA demand exceeds the TMA throughput or airport AAR. It also suggests that high internal departure scheduling delays are associated with periods when airborne metered flights are already absorbing as much delay as is possible in the air. Other days with a high number of double delays show similar results. However, in order to confirm that these results apply more generally, we apply data mining techniques to three months of data from June to August 2010, including features that capture these observations. Results of this analysis are described in the next section.

B. Data Mining

As described in Section III-B, data mining techniques have been applied to a limited set of features in order to identify the dominant drivers of double delay, and more generally, high TMA scheduling delay. Double delay is defined as a GDP delay greater than 15 minutes, followed by a TMA scheduling delay greater than 5 minutes, while high TMA scheduling delay is defined as greater than 5 minutes. Accordingly, predictors are built for the occurrence of double delay and high TMA scheduling delay. For the double delay predictor, the number of observations from the dataset (EWR arrivals, June to August 2010) is low (310 observations), because of the requirement to be in GDP, which has consequences for the performance of the model. For the high TMA scheduling delay predictor, however, the number of observations from the dataset is much larger (3,803 observations). Both data sets are reasonably well balanced with 43% of the TMA scheduled flights in GDP classed as double delayed and 38% of the TMA scheduled flights classed as having high TMA scheduling delay.

Because the objective of this study is the identification of drivers, we remove features that are collinear or statistically insignificant. Based on the VIF analysis, the difference between the estimated runway demand based on $ETA_{TMA\ entry}$ times and actual AAR operated at the airport (feature 15 in Section III-B) was found to be collinear with the ratio of destination arrival demand and capacity (feature 1 in Section III-B), so was dropped from both models. A number of features were also found to be statistically insignificant at the 95th percentile level, and were dropped. 6 features were retained for the double delay predictor:

- The ratio of destination arrival demand and destination arrival capacity in terms of AAR;
- Average EDCT compliance;
- Virtual runway queue size experienced by TMA;
- The change in estimated runway demand based on ETA_{ETD} (ETA_{EDCT} in this case) and STA_{Freeze} times;
- The average change in airborne flight ETAs between ETD and TMA entry ($ETA_{TMA\ entry} - ETA_{ETD}$); and
- The time to the end of the GDP.

10 features were retained for the high TMA scheduling delay predictor:

- The ratio of destination arrival demand and destination arrival capacity in terms of AAR;
- The destination arrival demand across the arrival fix to which an internal departure is routed;
- Average airborne metering delay for flights arriving at the destination airport;
- Whether an internal departure is in a reroute advisory;
- Whether any reroute advisory is in place to the destination airport;
- Virtual runway queue size experienced by TMA;
- Virtual runway scheduling queue size experienced by TMA;
- The change in estimated runway demand based on ETA_{ETD} and $ETA_{TMA\ entry}$ times;
- The change in estimated runway demand based on ETA_{ETD} and STA_{Freeze} times; and
- The average change in airborne flight ETAs between ETD and TMA entry ($ETA_{TMA\ entry} - ETA_{ETD}$).

Table B-1 in Appendix B lists all features that were excluded from each predictor, with the associated t-statistics/VIF values that led them to be excluded.

As described in Section III-B, a number of data mining algorithms are used to build predictors of the occurrence of double delay and high TMA scheduling delay. These include logistic regression, a decision tree, a random forest with 25 trees, a Support Vector Machine (SVM) with linear kernel, and K-nearest neighbors with K=10. All quantitative (non-categorical) features are scaled, and 10-fold cross validation is applied. The performance of each model is shown in Table 3 and Table 4, predicting the occurrence of double delay and high TMA scheduling delay respectively. In these tables, *Misclassification Error* refers to the fraction of incorrect predictions from all predictions made. *Accuracy* refers to the fraction of correct predictions from all predictions made (misclassification error and accuracy add to 1). *True Positive Rate* or *Recall* refers to the number of elements correctly labeled as belonging to the positive class divided by the total number of elements that actually belong to the positive class. *True Negative Rate* refers to the number of elements correctly labeled as belonging to the negative class divided by the total number of elements that actually belong to the negative class. *False Positive Rate* refers to the number of

Downloaded by NASA AMES RESEARCH CENTER on August 17, 2016 | http://arc.aiaa.org | DOI: 10.2514/6.2016-3456

elements incorrectly labeled as belonging to the positive class divided by the total number of elements that belong to the negative class. *False Negative Rate* refers to the number of elements incorrectly labeled as belonging to the negative class divided by the total number of elements that belong to the positive class. *Precision* refers to the number of elements correctly labeled as belonging to the positive class divided by the total number of elements labeled as belonging to the positive class.

Table 3. Data mining results for different algorithms applied to feature set for EWR arrivals from June 1, 2010 to August 31, 2010, predicting occurrence of Double Delays (GDP Delay > 15 minutes and TMA Scheduling Delay > 5 minutes).

	Logistic Regression	Decision Tree	Random Forest	SVM	K-Nearest Neighbors
Misclassification Error	0.31	0.41	0.34	0.32	0.35
Accuracy	0.69	0.59	0.66	0.68	0.65
True Positive Rate/Recall	0.64	0.53	0.64	0.64	0.60
True Negative Rate	0.73	0.63	0.68	0.70	0.70
False Positive Rate	0.27	0.37	0.32	0.30	0.30
False Negative Rate	0.36	0.47	0.36	0.36	0.40
Precision	0.64	0.52	0.60	0.62	0.60

Table 4. Data mining results for different algorithms applied to feature set for EWR arrivals from June 1, 2010 to August 31, 2010, predicting occurrence of high Internal Departure TMA scheduling delay (TMA Scheduling Delay > 5 minutes)

	Logistic Regression	Decision Tree	Random Forest	SVM	K-Nearest Neighbors
Misclassification Error	0.27	0.34	0.28	0.27	0.28
Accuracy	0.73	0.66	0.72	0.73	0.72
True Positive Rate/Recall	0.72	0.56	0.68	0.71	0.71
True Negative Rate	0.73	0.72	0.75	0.75	0.72
False Positive Rate	0.27	0.28	0.25	0.25	0.28
False Negative Rate	0.28	0.44	0.32	0.29	0.29
Precision	0.62	0.55	0.63	0.63	0.61

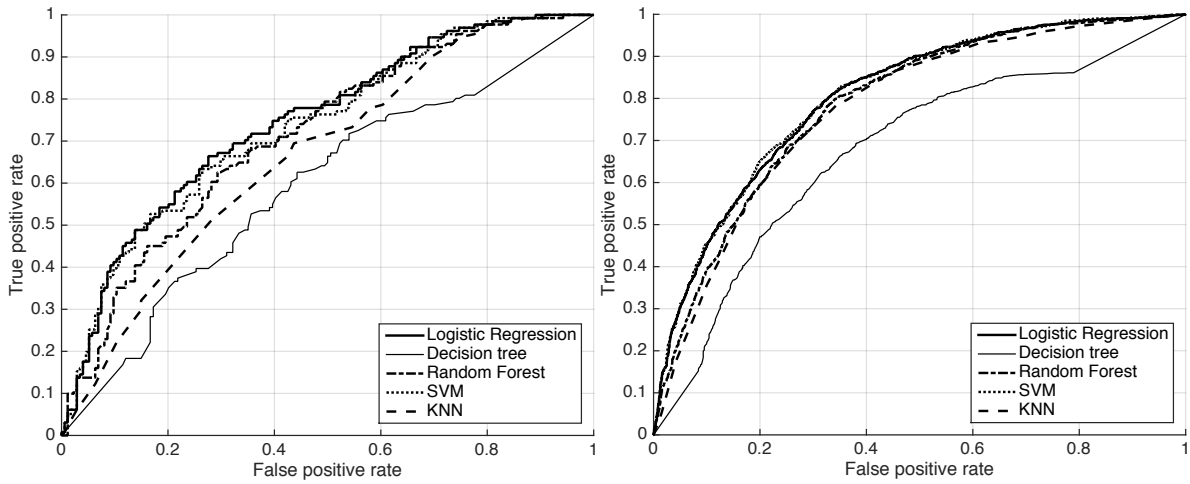
Receiver Operating Characteristic (ROC) curves are also generated for each model, varying the threshold for what probability constitutes a positive response, as shown in Figure 6 (in Table 3 and Table 4 a threshold of 0.5 is used). The closer the ROC curve is to false positive rate = 0 and true positive rate = 1, the better the model performance.

As can be seen in Table 3 and Figure 6a, for modeling the occurrence of double delay the logistic regression performs best across all metrics, with an accuracy of 0.69. Considering Table 4 and Figure 6a, for modeling the occurrence of high TMA scheduling delay the logistic regression and SVM perform best across most metrics, with an accuracy of 0.73. This is higher than the accuracy of the double delay model, but given the larger number of observations this is expected. The logistic regression is used to generate results in this paper because of the ease by which its results can be interpreted.

The estimated coefficients for the two logistic regression models, with t-statistics, are presented in Table 5 and Table 6. We first present the results of the high TMA scheduling delay model (Table 5), because it applies more generally, and is based on a better performing model. We then present the results of the double delay model (Table 6), focusing particularly on the differences relative to the high TMA scheduling delay model, highlighting the unique characteristics of double delay versus high TMA scheduling delay.

Downloaded by NASA AMES RESEARCH CENTER on August 17, 2016 | http://arc.aiaa.org | DOI: 10.2514/6.2016-3456

Downloaded by NASA AMES RESEARCH CENTER on August 17, 2016 | http://arc.aiaa.org | DOI: 10.2514/6.2016-3456



(a) (b)

Figure 6. Receiver operating characteristic (ROC) curve for different data mining algorithms applied to feature set for EWR arrivals from June 1, 2010 to August 31, 2010, characterizing (a) double delay (GDP delay > 15 minutes and TMA scheduling delay > 5 minutes), and (b) high internal departure TMA scheduling delay (TMA scheduling delay > 5 minutes).

Table 5. Logistic Regression Model parameters, predicting occurrence of high internal departure TMA scheduling delay (TMA scheduling delay > 5 minutes). Percentage increase in odds (% Inc. Odds) represents the increase in odds that an internal departure would experience high TMA scheduling delay as a result of an increase in the associated variable by one standard deviation. Standard deviations in each variable are included in the column 'Std. Dev.'

Feature	Estimate	tStat	% Inc. Odds	Std. Dev.
(Intercept)	-2.91	-17.65	-	-
Arrival Demand/AAR	1.87	9.84	547.8%	0.27
Average Airborne Metering Delay	0.22	9.62	24.2%	2.2 min
$\text{Demand}_{\text{ETA_ETD}} - \text{Demand}_{\text{STA_Freeze}}$	0.04	5.97	4.4%	11 ac/hr
Virtual Runway Queue Size	0.18	4.65	19.3%	1.9 ac
Arrival Demand across Meter Fix	0.02	3.37	1.9%	6.9 ac/hr
Virtual Runway Scheduling Queue Size	0.15	2.97	15.8%	1.4 ac
Destination in Reroute Advisory	-0.20	-2.77	-18.5%	-
Average ($ETA_{TMA\ entry} - ETA_{ETD}$)	0.02	2.44	1.8%	7.5 min
$\text{Demand}_{\text{ETA_ETD}} - \text{Demand}_{\text{ETA TMAentry}}$	0.02	2.15	1.5%	12 ac/hr
In Reroute Advisory	-2.21	-2.13	-89.1%	-

The t-statistics in Table 5 indicate that the most consistent contributor to the occurrence of high TMA scheduling delay is a high ratio of arrival demand to capacity (AAR), which has a t-statistic of 9.84. This feature also makes the largest contribution to the probability of high TMA scheduling delay, with an increase in odds of over 500% associated with an increase in the ratio of arrival demand to capacity by 0.27 (one standard deviation). This result is consistent with the results of the deep dive described in Section IV-A, and confirm that these observations apply more generally across the larger dataset (June to August 2010), and not only on the days analyzed in the deep dive. The result is also expected given that a high ratio of demand to capacity is required for internal departures to incur TMA scheduling delays.

The second most consistent contributor to high TMA scheduling delay, based on the t-statistics in Table 5 is high average airborne metering delay. This is consistent with the observations from the deep dive, where high TMA scheduling delays are observed to coincide with periods when airborne metering delay is absorbing the maximum delay possible. This again confirms that these observations apply more generally across the larger dataset. The result

also makes sense, because when airborne metered flights are receiving high metering delays, it is probable that there would be few gaps for internal departures to fit into the arrival stream. Hence TMA scheduling delays would be high. The feature also makes a reasonably large contribution to the probability of high TMA scheduling delay, with an increase in odds of over 20% associated with an increase in airborne metering delay by 2.2 min (one standard deviation).

The third most consistent contributor to high TMA scheduling delay, based on the t-statistics in Table 5, is the change in estimated runway demand based on ETA_{ETD} and STA_{Freeze} times. This feature represents a proxy for the difference between the expected demand based on expected departure times and the airport capacity used by TMA. The significance of this feature is consistent with the observations made in the deep dive that high TMA scheduling delays are associated with periods in which TMA demand exceeds TMA throughput. The feature makes a relatively small contribution to the probability of high TMA scheduling delay, however, with an increase in odds of high TMA scheduling delay of only 4.4% when demand based on ETA_{ETD} times exceeds that based on STA_{Freeze} times by 11 ac/hr (one standard deviation).

Virtual runway queue size processed by TMA, as well as the virtual runway scheduling queue size, also consistently contribute to high TMA scheduling delay. TMA does not strictly operate as a queue at the runway, because internal departures are not served based on first-come, first-served, but must wait for a gap in the arrival stream. However, the virtual runway scheduling queue, for internal departures, is served based on first-come, first-served, with any TMA scheduled flight served before other TMA scheduled flights that are TMA ready after it. These results, however, indicate that even the full TMA queue, including both scheduled and metered flights, contributes consistently to high TMA scheduling delay. This is expected partly because the virtual runway queue size is a direct indicator of how many flights must still be delayed by TMA when an internal departure is TMA ready. The increase in odds of high TMA scheduling delay associated with an increase in both these features by one standard deviation is reasonably high (between 16% and 19%).

Table 5 also indicates that arrival demand across the arrival fix to which an internal departure is routed is a consistent contributor to high TMA scheduling delay, as it is not only the runway, but also the meter fixes that are constraints in TMA. This is consistent with SME feedback. However, the increase in odds of high TMA scheduling delay associated with this variable is low, at only 1.9% for an increase in arrival demand of 6.9 ac/hr (one standard deviation), indicating that the runway is the primary constraint.

The t-statistics in Table 5 also indicate that the average change in airborne flight ETA between estimated departure time and TMA entry ($ETA_{TMA\ entry} - ETA_{ETD}$) is also a consistent indicator of high TMA scheduling delay. This feature indicates the degree to which there are differences in estimated times en route before departure and as used to generate TMA ETAs. The impact, however, is small, with a 7.5 min average shift forward in ETA between departure and TMA entry associated with only a 1.8% increase in odds. A metric that captures some of the consequences of this effect is the change in estimated runway demand between departure and TMA entry, which can also be seen in Table 5 to be a consistent indicator of high TMA scheduling delay. It also has a relatively low impact on odds of 1.5%.

Finally, the inclusion of the destination airport in a reroute advisory, as well as the inclusion of the whole route from origin to destination in a reroute advisory, also correlate consistently (but negatively) with the occurrence of high TMA scheduling delay. The latter in fact has quite a large impact on odds, with the inclusion of the route in a reroute advisory reducing the odds of high TMA scheduling by 89%. This may be because a reroute advisory is indicative that action is being taken to reduce the delay otherwise required on the route, so, with a reroute advisory in place, the required TMA scheduling delay is lower. A reroute may also open up space in the overhead stream for internal departures. This suggests that, where possible, a reroute advisory may have a significant impact on reducing TMA scheduling delay.

Given the results of the more general case identifying drivers of high TMA scheduling delay, we now focus on the more specific case of identifying drivers of double delay in Table 6, highlighting the differences relative to the high TMA scheduling delay model, allowing us to identify the unique characteristics of double delay versus high TMA scheduling delay.

Downloaded by NASA AMES RESEARCH CENTER on August 17, 2016 | http://arc.aiaa.org | DOI: 10.2514/6.2016-3456

Table 6. Logistic Regression Model parameters, predicting occurrence of double delay (GDP delay > 15 minutes and TMA scheduling delay > 5 minutes). Percentage increase in odds (% Inc. Odds) represents the increase in odds that an internal departure would experience double delay as a result of an increase in the associated variable by one standard deviation.

Feature	Estimate	tStat	% Inc. Odds	Std. Dev.
Virtual Runway Queue Size	0.44	5.39	55.0%	1.9 ac
(Intercept)	-2.41	-3.75	-	-
Arrival Demand/AAR	1.44	2.80	320.6%	0.27
Average EDCT Compliance	-0.03	-2.63	-3.1%	13 min
$\text{Demand}_{\text{ETA_ETD}} - \text{Demand}_{\text{STA_Freeze}}$	0.03	2.57	3.3%	11 ac/hr
Average ($ETA_{TMA\ entry} - ETA_{ETD}$)	0.04	2.28	4.1%	7.5 min
Time to GDP End	-0.002	-2.12	-0.2%	150 min

The results of the logistic regression modeling the occurrence of double delay, in Table 6, identify similar features to those identified for high TMA scheduling delay. A high ratio of arrival demand to capacity is a fairly consistent contributor to double delay, with a t-statistic of 2.80, although it is not as consistent a contributor to the occurrence of double delay as high TMA scheduling delay. However, this feature does still show the largest increase in odds (320%) associated with an increase in the feature by 0.27 (one standard deviation). The most consistent contributor to double delay is, instead, virtual runway queue size processed by TMA, which has a t-statistic of 5.39, and an increase in odds of double delay associated with an increase in queue size by 1.9 aircraft (one standard deviation) of 55%. Unlike the prediction of high TMA scheduling delay, virtual runway scheduling queue size is not found to be significant to the occurrence of double delay. Also unlike the prediction of high TMA scheduling delay, average airborne metering delay is not found to be a consistent contributor to the occurrence of double delay, suggesting the it is more indicative of high TMA scheduling delay outside GDP. A larger set of observations of flights in GDP would be needed to confirm this.

The change in estimated runway demand based on ETA_{ETD} and STA_{Freeze} times is another consistent contributor to the occurrence of double delay, as it is for high TMA scheduling delay. However, this feature differs slightly from that used in the prediction of high TMA scheduling delay because, in this case, ETA_{ETD} is entirely based on EDCTs, since all flights in the analysis are under GDP. Therefore, this metric represents a proxy for the difference between the airport capacity used to define EDCTs and the airport capacity used by TMA. Its inclusion in Table 6 is indicative that when EDCTs are generated based on arrival rates that are ultimately not realizable by TMA, double delays result, which is expected.

There is a similar difference for the feature describing the average change in airborne flight ETA between ETD (EDCT in this case) and TMA entry. Because all flights are in GDP, in Table 6 this feature indicates the degree to which there are differences between airline scheduled flight times used to generate EDCTs, and estimated times en route used to generate TMA STAs, which are likely to undermine the sequencing and spacing underlying the flight EDCTs. This is consistent with the observations from the deep dive. The change in estimated runway demand between EDCT and TMA entry is not found to be significant to the occurrence of double delay. This may, however, also be because of a lack of data, and would need to be confirmed with an increased set of observations.

Two other features that were found to be significant to the occurrence of high TMA scheduling delay in Table 5, but are not found to be significant to the occurrence of double delay in Table 6, are the inclusion of the destination airport in a reroute advisory, and the inclusion of the whole route from origin to destination in a reroute advisory. This may be because the GDP is already indicative that action is being taken at the destination airport.

Finally, two features are found to be significant in the occurrence of double delay in Table 6 that were not found to be significant in the occurrence high TMA scheduling delay in Table 5. These are average EDCT compliance – particularly early departure relative to EDCT – and time to the end of the GDP. The significance of average EDCT compliance is consistent with the observations of the deep dive, in which flights were observed to consistently depart before EDCT on days with high double delays, confirming that these observations apply more generally across the larger dataset. This result is expected because early departure is likely to undermine the sequencing and spacing underlying the flight EDCTs. The increase in odds of double delay, however, associated with an early departure relative to EDCT by 13 minutes (one standard deviation) is not large (3%). The estimate for the feature describing the time to the end of the GDP is negative, indicating that double delays are more likely near the end of a

Downloaded by NASA AMES RESEARCH CENTER on August 17, 2016 | http://arc.aiaa.org | DOI: 10.2514/6.2016-3456

GDP, which is expected. The increase in odds of double delay associated with a reduction in the time to the end of a GDP by 150 minutes (one standard deviation) is, however, very small (0.2%).

V. Conclusion and Recommendations

When weather or congestion impact the National Airspace System, multiple different Traffic Management Initiatives can be implemented, sometimes with unintended consequences. One particular perceived inequity that is investigated in this paper is the interaction between GDPs and TMA (now TBFM) scheduling of internal departures. Internal departures under TMA scheduling can take large GDP delays, followed by large TMA scheduling delays, because they cannot be easily fitted into the arrival stream at the runway. Through an analysis of arrival operations at Newark Liberty International Airport (EWR) from June to August 2010, average TMA scheduling delays are found to be significantly higher than airborne metering delays – more so than warranted based on the differences in cost associated with airborne and ground delay – regardless of prior GDP delays. Depending on how the double delay is defined, between 42% and 62% of all internal departures in GDP and TMA scheduling experienced double delays in this period. While a relatively small percentage of flights is eligible for double delays, a higher percentage of flights are eligible for high internal departure scheduling delay, and the percentage of TMA scheduled flights that are highly delayed is also high – between 39% and 56%, depending on how it is defined. This illustrates the importance of developing a concept that prevents high TMA scheduling delays for internal departures, both in GDP and not in GDP.

A deep dive into the data reveals that contributors to double delays include upstream flights departing before their Expect Departure Clearance Times (EDCTs); differences in the rates used for setting EDCTs and TMA Scheduled Times of Arrival; differences in the arrival demand expected based on EDCTs and the arrival demand entering TMA; and shorter en route times between takeoff and entry into TMA than assumed in the calculation of flight EDCTs, all of which undermine the sequencing and spacing underlying flight EDCTs. Double delays are also found to coincide with periods in which the virtual runway arrival queue being served by TMA is large, periods of high demand relative to capacity, and high airborne metering delays.

Data mining techniques are used to confirm that each of these factors contribute to the occurrence of double delay and/or high internal departure scheduling delay across three months of data from June to August 2010. Predictors of the occurrence of double delay and high TMA scheduling delay were built using logistic regression, providing prediction accuracies of 69% and 73%, respectively. The predictor inputs include features describing each of the above factors as well as other factors such as whether a flight is in a reroute advisory. Features exhibiting colinearity and statistical insignificance are removed to ensure that the results are interpretable.

In order to reduce double delay and high TMA scheduling delay, a concept must be developed that addresses the identified factors. The Region TBO concept under development at NASA specifically includes mechanisms to control the flow entering TMA to ensure that gaps still exist in the arrival stream for internal departures. The exact manner in which this is to be done is still the subject of research. However, the results in this paper suggest that if the concept were to enable the spacing and sequencing defined by EDCTs to be preserved on entry into TMA, this could contribute to reducing double delays. This would require addressing differences in en route times used for calculating EDCTs and TMA ETAs, as well as the accuracy of the arrival rates assumed for the calculation of EDCTs. The latter may be difficult to improve because of capacity prediction uncertainties.

While the analysis described in this paper goes some way to identify the causes of double delay and high TMA scheduling delay, it is also important to confirm that the direct cause of these delays is a lack of gaps in the arrival stream, and to identify whether preserving the gaps in the arrival stream as defined by EDCTs would be sufficient to reduce double delay and high TMA scheduling delay. This will require the calculation of expected internal departure delays based on fitting internal departures into the runway arrival stream, using TMA runway STAs for all other flights, and using runway ETAs at EDCT. This is left for future work.

Acknowledgments

The authors would like to acknowledge helpful comments from Nancy Smith, Heather Arneson, Deepak Kulkarni and Yao Wang of NASA Ames Research Center. The authors would also like to acknowledge helpful insights from subject matter experts Danny Vincent and Mark Evans. Finally, the authors would like to thank Sarasina Tuchen and the Volpe National Transportation Systems Center for providing the multi-TMI data on which this study is based.

Downloaded by NASA AMES RESEARCH CENTER on August 17, 2016 | http://arc.aiaa.org | DOI: 10.2514/6.2016-3456

Downloaded by NASA AMES RESEARCH CENTER on August 17, 2016 | http://arc.aiaa.org | DOI: 10.2514/6.2016-3456

References

[1] Dwyer A.M., Epstein L., Futer A., Hogan M., Howard K., Oisen R., Sharick B., "Interactions of Multiple Traffic Management Initiatives: An Initial Analysis, Version 2," *Volpe National Transportation Systems Center*, Report No. VNTSC-TFM-11-11, Cambridge, MA, 2011.

[2] Department of Transportation, "NY flight delays have three main causes, but more work is needed to understand their nation wide effect," *Inspector General Office*, Washington, DC, URL: https://www.oig.dot.gov/sites/default/files/NY%20Delays%20Final.pdf [cited April 22, 2015].

[3] NextGen Implementation Plan 2015, *Federal Aviation Administration*, Washington DC, URL: https://www.faa.gov/nextgen/media/NextGen_Implementation_Plan-2015.pdf [cited April 13, 2016].

[4] Based on feedback from Subject Matter Experts.

[5] Brinton, C., Atkins, S., Cook, L., Lent, S., and Prevost, T., "Ration by Schedule for airport arrival and departure planning and scheduling," *IEEE Integrated Communications Navigation and Surveillance Conference,* pp. I3-1, Herndon, VA, 2010.

[6] Grabbe, S., Sridhar, B., Mukherjee, A., Morando A., "Traffic Management Advisor Flow Programs: an Atlanta Case Study," *AIAA Guidance, Navigation, and Control Conference*, Portland, OR, August 2011.

[7] Wanke C., Taylor C., "Exploring Design Trade-offs for Strategic Flow Planning," *AIAA Aviation Technology, Integration, and Operations Conference*, Los Angeles, CA, August 2013.

[8] Rebollo J., Brinton C., "Brownian Motion Delay Model for the Integration of Multiple Traffic Management Initiatives," *Eleventh USA/Europe Air Traffic Management Research and Development Seminar*, Lisbon, Portugal, 2015.

[9] Bilimoria, K., "Analysis of Additional Delays Experiences by Flights Subject to Ground Holding," *AIAA Aviation Technology, Integration, and Operations Conference*, Washington, DC, June 2016.

[10] Mukherjee, A., Hansen, M., "A dynamic stochastic model for the single airport ground holding problem," *Transportation Science* 41(4), pp. 444-456, 2007

[11] Sridhar, B., Swei, S.S.M., "Classification and Computation of Aggregate Delay Using Center-based Weather Impacted Traffic Index," *7th AIAA Aviation Technology, Integration and Operations Conference*, Belfast, Northern Ireland, 2007.

[12] Eshow, M.M., Lui, M. and Ranjan, S., "Architecture and capabilities of a data warehouse for ATM research," *33rd IEEE/AIAA Digital Avionics Systems Conference,* pp. 1E3-1, Colorado Springs, CO, October 2014.

[13] Cox, D.R., "The regression analysis of binary sequences (with discussion)," *J Roy Stat Soc B,* 20: 215–242, 1958.

[14] Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, P.J., *Classification and Regression Trees*, Wadsworth, 1984.

[15] Cortes, C., Vapnik, V., "Support-vector networks," *Machine Learning*, 20 (3): 273, 1995.

[16] Ho, T.K., "Random Decision Forests," *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, Montreal, QC, 14–16 August, 1995, pp. 278–282.

[17] Altman, N.S., "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, 46 (3): 175–185, 1992.

Downloaded by NASA AMES RESEARCH CENTER on August 17, 2016 | http://arc.aiaa.org | DOI: 10.2514/6.2016-3456

Appendix A: Estimate of Active Metering

As described in Section II-A, based on SME feedback, TMA metering being turned on and displayed on controller screens does not necessarily mean that aircraft are being actively metered. TMA metering is sometimes only used to enhance situational awareness, and not to actively meter aircraft. We use a hidden Markov model (HMM) to estimate when TMA metering is being actively used to meter aircraft, and when it is not. This is appropriate because the state of TMA metering, whether it is being used for active metering or not, can be assumed to be a Markov process, i.e., memoryless, whereby predictions of the future state can be made based solely on the present state. It is hidden because the state is not revealed in the data available.

The data is filtered for only those flights in which TMA metering is turned on and being displayed on controller screens. These flights are assumed to be in one of two states: in active metering, or not in active metering (TMA being used for situational awareness only). Ideally, the HMM emissions, from which we will attempt to identify the state, should be flight TMA compliance, or the accuracy of flights meeting their STAs at the meter fix, as defined by equation (1).

$$TMA\ Compliance = ATA_{Meter\ Fix} - STA_{Freeze\ Horizon} \tag{1}$$

ATA refers to Actual Time of Arrival. As flights cross the freeze horizon 19 minutes before they are estimated to cross the meter fix, their STA at the meter fix is frozen. Controllers then use control techniques, such as speed control or lateral maneuvers, to absorb sufficient delay to meet this time at the meter fix. TMA compliance, in equation (1) can provide an indication of how accurately flights meet this time. Flights are considered to be TMA compliant if they meet the time within ±1 minute.

Unfortunately, flight STAs at the freeze horizon are not available in the data analyzed. However, runway STAs are available for different points in each flight's trajectory, including when the flight became active in TMA, when it crossed the freeze horizon, and when it crossed the meter fix. If we assume that the predicted transition time through the TRACON does not change between the freeze horizon and meter fix crossing times, then TMA compliance can be approximated by equation (2), measured instead based on runway STAs (*RwySTA*).

$$TMA\ Compliance = RwySTA_{Meter\ Fix} - RwySTA_{Freeze\ Horizon} \tag{2}$$

TMA compliance, estimated using equation (2) is used to define the HMM emissions, from which we attempt to identify the TMA metering state.

A discrete HMM is used to estimate the state. Despite the fact that TMA compliance can be considered to be of a Gaussian form, making a Gaussian HMM appropriate to estimate the state, flights are considered to be TMA compliant based on whether they meet the meter fix STA within ±1 minute. TMA compliance can therefore be categorized as either less than -1 minute (not TMA compliant), between -1 minute and 1 minute (TMA compliant), or greater than +1 minute (also not TMA compliant). Such categorization of the HMM observations makes a discrete HMM appropriate.

Fitting the discrete HMM to the TMA compliance data calculated for all TMA metered flights using equation (2), results in transition and emissions matrices as follows:

$$Transition\ Matrix = \begin{bmatrix} 0.966 & 0.034 \\ 0.093 & 0.907 \end{bmatrix}$$

$$Emissions\ Matrix = \begin{bmatrix} 0.056 & 0.909 & 0.035 \\ 0.313 & 0.316 & 0.371 \end{bmatrix}$$

The transition matrix presents the transition probabilities, which control the way the hidden state is chosen given the hidden state in the previous time period. The probabilities of staying in the current state are highest (the diagonal entries), while the probability of changing state are significantly lower (the off-diagonal entries), as expected.

The emissions matrix presents the emissions probabilities, which govern the distribution of the observed variable at a particular time given the state of the hidden variable at that time. In the first state (the top row), the probability of the observed variable, TMA compliance, falling in the first and last categories (less than -1 minute and greater than +1 minute) is very low (0.056 and 0.035, respectively), while the probably of it falling in the middle category (between -1 minute and 1 minute) is significantly higher (0.909), as would be expected during active metering. This suggests that this first state in 'actively metering'. In the second state (the bottom row) the probability of the observed variable, TMA compliance, falling in all three categories is approximately the same, as would be expected when flights are not being actively metered. This confirms that the second state in 'not actively metering'.

Figure A-1 shows an application of the fitted HMM to observed TMA compliance data for TMA metered flights at EWR on June 29, 2010. Note that the state is arbitrarily given the value of 1 and 2 for 'not actively metering' and

‘actively metering’, respectively. The model appears to do a reasonable job of capturing when active metering is taking place, with periods of poor TMA compliance corresponding to TMA State = 1, and periods of poor TMA compliance corresponding to TMA State = 2.

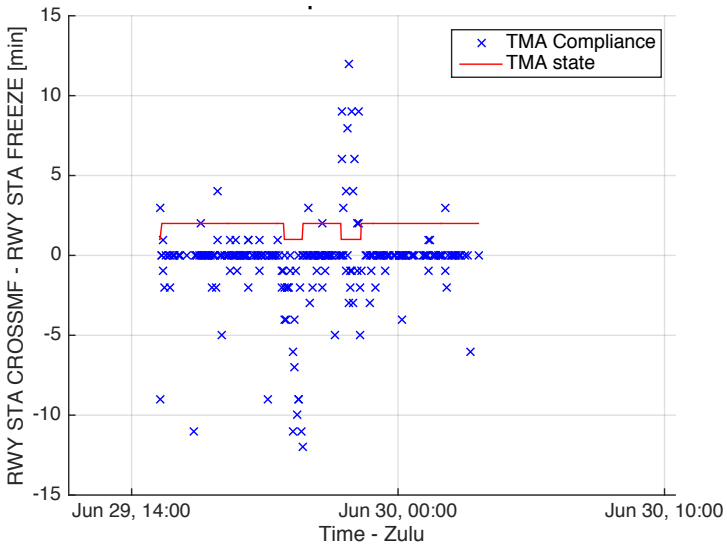


Figure A-1. Example application of HMM model – TMA metering at EWR, June 29, 2010. TMA State = 1 for ‘not actively metering’ and TMA State = 2 for ‘actively metering’.

Downloaded by NASA AMES RESEARCH CENTER on August 17, 2016 | http://arc.aiaa.org | DOI: 10.2514/6.2016-3456

Appendix B: Feature Exclusion For Data Mining

Table B-1. Features included and excluded (due to colinearity based on VIF>10, or statistical insignificance based on tStat < 1.96) in predictors of double delay (GDP delay > 15 min, and TMA scheduling delay > 5 minutes), and high internal departure TMA scheduling delay (TMA scheduling delay > 5 minutes).

Feature	Double Delay Predictor			High TMA Scheduling Delay Pred.		
	Incl.	Excluded		Incl.	Excluded	
		VIF	tStat		VIF	tStat
Arrival Demand/AAR	✔			✔		
Arrival Demand across Meter Fix	✗		-0.73	✔		
Origin WITI	✗		-0.27	✗		0.56
Destination WITI	✗		0.06	✗		1.19
Average Airborne Metering Delay	✗		0.89	✔		
Average Airborne Metering Delay across Meter Fix	✗		-0.18	✗		0.37
Average EDCT Compliance	✔			✗		-1.43
Standard Deviation EDCT Compliance	✗		-0.12	✗		-0.98
MIT	✗		0.02	✗		1.16
In Reroute Advisory	✗		-0.36	✔		
Destination in Reroute Advisory	✗		0.24	✔		
Virtual Runway Queue Size	✔			✔		
Virtual Runway Scheduling Queue Size	✗		0.02	✔		
$Demand_{ETA_TMAentry} - Demand_{ETA_ETD}$	✗		0.90	✔		
$Demand_{ETA_ETD} - Demand_{STA_Freeze}$	✔			✔		
$Demand_{ETA_TMAentry} - AAR$	✗	75.8		✗	58.2	
Average ($ETA_{TMA\ entry} - ETA_{ETD}$)	✔			✔		
Standard Deviation ($ETA_{TMA\ entry} - ETA_{ETD}$)	✗		-0.08	✗		-0.14
Time since GDP Start	✗		-1.44			
Time to GDP End	✔					
In GDP				✗		0.45

Downloaded by NASA AMES RESEARCH CENTER on August 17, 2016 | http://arc.aiaa.org | DOI: 10.2514/6.2016-3456