Variables Influencing RNAV STAR Adherence

Michael Stewart
San José State University
NASA Ames Research Center
Moffett Field, CA, USA
Michael.J.Stewart@nasa.gov

Bryan Matthews
Stinger Ghaffarian Technologies Inc.
NASA Ames Research Center
Moffett Field, CA, USA
Bryan.L.Matthews@nasa.gov

Vijay Janakiraman
USRA
NASA Ames Research Center
Moffett Field, CA, USA
vijai@umich.edu

Ilya Avrekh
Stinger Ghaffarian Technologies Inc.
NASA Ames Research Center
Moffett Field, CA, USA
ilya.avrekh-1@nasa.gov

Abstract— **In this study we investigated how variables in the aviation domain impact adherence levels of aircraft flying area navigation arrivals with optimized profile descents (RNAV OPDs). Variables categories were: *weather, aircraft, procedure,* and *traffic.* Non-adherence events analyzed were: *miss above, miss below, skip before merge,*** and ***skip after merge. Miss below* and *miss above* describe when a flight does not comply vertically with a procedure. *Skips* refer to a flight leaving a procedure, then returning. Findings of this work reveal that vertical events are most impacted by altitude restriction size, steepness of flight paths, and merging routes. Lateral events were impacted by merging routes, number of speed restrictions, and the flow rate of the arrival traffic. This study helps increase understanding of how the system is functioning and identifies where procedures are not flexible enough to handle the variability in normal operations.**

Keywords—RNAV, Data Mining, Decision Trees, RADI, Adherence

I. Introduction

The benefits of using continuous descent arrivals have been quantified in terms of fuel and noise, so why is full adherence around ten percent? Pre-planned flight paths, known as procedures, aim to alleviate some of the need for air traffic control (ATC) intervention and allow aircraft to self-regulate. Specifically, this is the case for standard terminal arrival routes with optimized profile descents (OPD STARs). These procedures provide a vertical, lateral, and typically a speed profile for the aircraft to follow. Expectations for utilizing profiles as designed are: reduced ATC communications, reduced fuel consumption, reduced noise, and increased predictability of aircraft trajectories [1]. However, for this idea to work, the procedure's design must encompass enough of the variability in the operational domain (e.g., weather, aircraft differences, and traffic volume) to afford the aircraft the opportunity to stay within the design boundaries of the procedure in an uninterrupted manner. When variables in the domain preclude using the intended procedural flight path (non-adherence), alternative ATC instructions must be implemented for the aircraft to remain clear of terrain and traffic. The reasons for this might be pilot initiated, controller initiated, human error, or a combination of these. Ultimately, the responsibility is left to ATC for assuring separation, then pilots for adapting appropriately to the new instructions.

Our goals for this research were to identify and quantify some of the variables in the domain that contribute to the non-adherence of aircraft flying OPD STARs. Understanding this topic first requires an operational definition of adherence, rationale for the types of variables we included in our analyses, events we measured to describe types of aircraft non-adherence, and how procedures have changed flight path management. First, as defined in Advisory Circular 90-100A adhering to a procedure is flying one mile either side of centerline and within 300 feet above or below an altitude restriction. The 300 feet rule is an industry standard. These criteria represent aircraft navigation equipment performance requirements and pilot precision requirements, respectively [2]. Second, we wanted to capture variables that are out of the control of the users (pilots and controllers) and quantifiable. This was intended to objectively describe the impact of the design and its interaction with the environment to the extent possible. After reviewing literature, Aviation Safety Reporting System (ASRS) reports, and subject matter expert knowledge, the measurable variables were broken into four categories: weather, procedure, aircraft, and traffic. This taxonomy can be amended in the future; we aimed to be as comprehensive as possible. Third, events used to quantify non-adherence were broken into two categories: vertical and lateral. The vertical events were defined as flying below (*miss below*) or above (*miss above*) a waypoint restriction altitude by more than 300 feet. Lateral events were defined as leaving the procedure by more than one mile and then returning, either before a merging waypoint (*skip before merge*) or after a merging waypoint (*skip after merge*) (See Figures 1 and 2.).

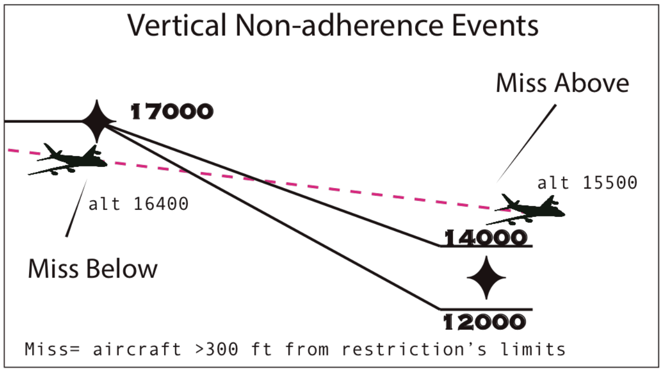

Fig. 1 Diagram of Vertical Non-adherence Events.

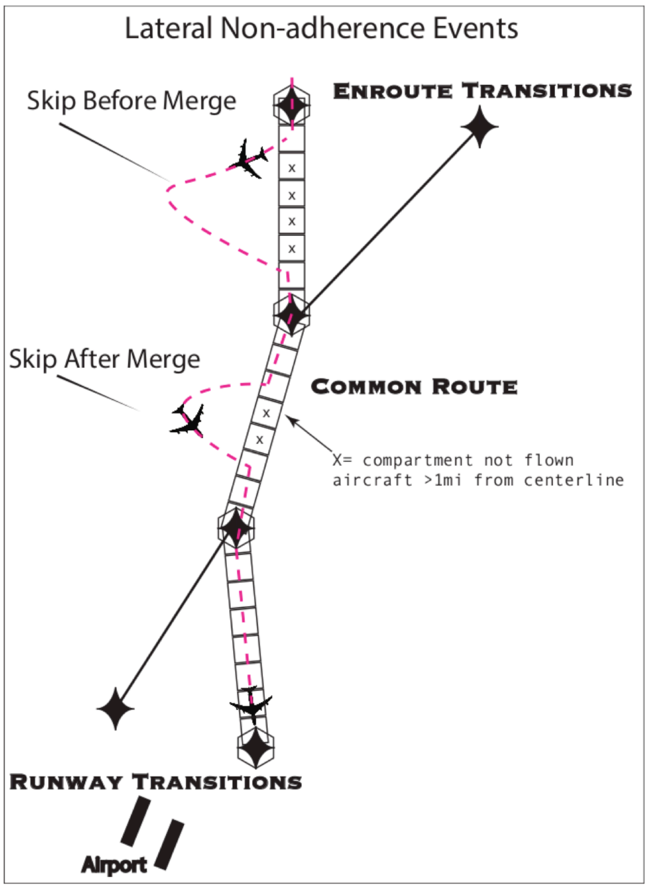

Fig. 2 Diagram of Lateral Non-adherence Events

We used the variables describing the operational domain and compared them to the non-adherence events to see if there was a relationship and a predictive capability in the variables. This information gave us a partial view of context (weather and traffic), procedure design characteristics, and aircraft type for their effect on adherence events. We accomplished this by using the RNAV Adherence Data Integration system (RADI). Historical radar track from Federal Aviaiton Administration (FAA) System Wide Information Management (SWIM) network, National Oceanagraphic Atmosphereic Administration (NOAA) wind, Corridor Integrated Weather System (CIWS), and Coded Instrument Flight Procedure (CIFP) data are integrated in this system [3]. Lastly, the fundamentals of the descent phase of flight have not changed as a result of the OPD STAR, but the tasks of speed assignment and descending, once handled dynamically by ATC, are now transferred to a fixed procedure. This rigidity creates a mismatch between the realities of the domain and how procedures are designed. To date, procedures cannot adapt to real-time conditions, which renders them unusable at times when weather, traffic, and aircraft variables exceed their design scope.

II. Background

To illustrate how often ATC and pilots are involved in amending procedures, we refer to our previous work on OPD STAR usage statistics [4]. Reviewing 24 of the United States' most trafficked airports, median full compliance with the procedures' vertical and lateral boundaries ranged from 0% to 24%. This low usage and high variability shows something is inhibiting use as designed assuming that adherence is desirable. Analyses were accomplished by reviewing historical radar tracks overlaid onto OPD STAR lateral and vertical boundaries, then comparing the waypoint crossing altitude and position with those published STAR boundaries using derived variables from the data. One limitation of this study was that it was only capable of looking at waypoint adherence and wasn't able to detect aircraft position between waypoints. Therefore, aircraft behavior between waypoints may have been missed and not included. In addition, the radar source was not capable of showing the landing runway of the aircraft, which made runway transition identification difficult in some cases. This may have artificially lowered some adherence rates. Further, some routes transitioned directly to instrument approaches before the end of the normal track, which may have inflated the number of aircraft that appeared to be leaving the procedure early, when they were actually transitioning to a different procedure for landing. However, to our knowledge, this was limited to Denver (KDEN) airport and to a subset of the aircraft capable of radial fix required navigation performance (RF) (RNP). If we added all of the ambiguous flights into the full compliance category, the results would have shown a similarly high rate of variability and still the majority of aircraft not fully complying with the procedures.

The issue that flights will sometimes be forced to leave a procedure is not new. Human-in-the-loop simulations were conducted by Johnson, 2009 to identify issues that could limit the use the OPD STARs in Atlanta [5]. Heavy traffic flow was thought to be a contributor to flights being removed from OPDs. In addition, mixed aircraft, aircraft weight variability, merging, and initial spacing were thought to negatively interact with high levels of traffic. Mixed aircraft types, wind, and variable aircraft weight was noted because of their effects on descent path angles and where an aircraft needs to start its descent. As a result, controller transparency of where the aircraft will start descending is reduced and deceleration and spacing would be difficult to predict. Merging transition routes is problematic for spacing as it also requires a predictive element from a controller to fit two streams of traffic into one. Last, initial spacing of aircraft prior to the OPD must be sufficient to handle the effects of compression as the aircraft decelerate during descent. Further, the study reported that controllers thought the OPDs limited traffic flow below that of the airport's capability. Therefore, the idea that a procedure (other than an instrument approach) may be throttling throughput was noted.

Additional research by Mercer, Callantine, and Martin, 2012 discussed several variables they believed to impact unencumbered use of OPD STARs [6]. Their study used go-around scenarios, which require spacing changes in arrival flows to allow aircraft to reintegrate as they return to the landing airport. They wanted to identify recovery methods and determine if an automated decision support tool would be useful in the recovery process. Findings of this work show that automation may be necessary to assist controllers with the complex task of predicting future aircraft spacing. This work showed two things: OPD STARs are thought to be unusable in situations of improperly spaced traffic, and systematic recovery methods were explored to augment the procedures during known disruptions. However, there may be latent issues that contribute to non-adherence that were not addressed in this work (e.g., weather and human error).

Recovering from a situation outside the procedure's design scope contrasts from a design that is intended to be more usable. Early work on the human factors implications of Constant Descent Approaches (CDAs) (predecessors of OPD STARs) looked at how controllers might be influenced by the speed restrictions in the procedure [7]. They found that predicting an airplane's future position with varied deceleration can contribute to projection error and that standardization of speed profiles could be beneficial as a compensatory method. This work highlights that predicting future aircraft location is important to controllers and design for this purpose could be beneficial. This was a motivating factor in why we chose variables related to merging time differences.

Thunderstorms or convective weather poses risks to aircraft from turbulence, hail, lighting, and windshear [8]. Pilots deviate from planned routes to avoid these risks. Quantifying aircraft deviations from planned routes requires the position and intensity of the weather cell and aircraft position in the same temporal window. In addition, the location of the route is required for a comparison to the path being flown.

In 2006, Phoenix (KPHX) airport's procedures were investigated to gauge the effects of the implementation [9]. Dependent variables investigated were: descent continuity, fuel burn, mileage, and emissions. Emissions were measured in CO2 and calculated from estimates of fuel burn. Descent continuity was measured at two points on the arrival in 100 ft increments around 8000 ft and 9000 ft. The authors concluded that the arrivals increased the descent continuity, reduced fuel burn, and had no change in mileage. However, one condition was a subset of aircraft that flew the arrivals more precisely, which is inconsistent because this is not done for the baseline condition. It is likely there is also a subset of aircraft in the baseline condition that performed better than the average, but they are not revealed. Fuel savings analyses assumed that all of the traffic was in this same subset condition, but the days sampled showed many aircraft were off the route, which makes this assumption that all traffic would fly the route precisely to be unrealistic. Further, the authors noted that increases in vectoring might have resulted from the implementation. This study did not describe the variability in the data or statistical tests to determine if the differences could be attributed to chance. Overall, the authors stated that benefits were likely if the procedures were flown fully and consistently.

Safety issues with RNAV STAR procedures have been recurring. Incidents raised by pilots and controllers led to the issuance of two safety enhancements (SEs) by the Commercial Aviation Safety Team (CAST): SE 213 and SE 214 [10,11]. SE 213 was intended to address pilot and controller procedures, and charting depiction for the purpose of error reduction. SE 214 was intended to address path compliance issues related to procedure design [7]. Efficacy of this work was measured by the occurrence rate of subjective reports (e.g., ASRS, ASAP, ATSAP). This poses a limitation and may not be an adequate indicator. ASRS reports do not represent actual rates of occurrence and may not faithfully represent categories of events [8]. Quantitative occurrence rates are needed to augment the subjective reports to understand the issues comprehensively.

A. RADI System

RADI is a continuation of our previous work to quantify RNAV STAR adherence [3]. However, it utilizes additional data sources, more precise radar, finer grained procedure descriptions, and additional derived variables. The following data sources were leveraged to capture the relevant information and address adherence variables (See Figure 5.):

1) Aircraft Flight-track Data

NASA's Sherlock archive contains radar flight track data collected over the FAA's System Wide Information Management (SWIM) data feed. The primary sources include: SWIM Flight Data Publication Service and SWIM Terminal Data Distribution System. The unique flight tracks are merged between the two sources to produce a unique trajectory capturing en route operations down through landing. The flight trajectories span 11/12/14 through 01/31/18.

2) Wind Data

The NOAA Rapid Refresh (RR) data contains estimated winds aloft at 13 KM resolution across a US map projection grid. The files were parsed over the same time frame to capture snapshots of winds and temperatures aloft above each destination airport. A subset of these measurements were gathered at 1,000 ft. intervals ranging from 4,000 - 40,000 ft. with a one hour sampling rate.

3) Convective Weather Data

The MIT Lincoln Labs Corridor Integrated Weather System (CIWS) data were parsed over the same time frame to capture weather obstructions on the STAR. Since the weather cells were not expected to change very rapidly with respect to the flights, the files were sampled every five minutes. The data are projected using a US Lambert Azimuthal Equal-Area map projection grid. The data contains Vertical Integrated Liquid (VIL) and EchoTops for each grid point. The VIL and EchoTops are quantized in a lookup table to derive the weather avoidance field [12]. The *Weather on compartment* variable characterizes the probability of a flight avoiding the weather in the terminal airspace. Weather under 40% was excluded.

4) Procedure Data

STAR characteristic data are extracted from the Airspace Data File, which is provided by ATAC Corp. and is extracted from the coded instrument flight procedures every 56 days. This corresponds to 21 chart cycles for the time frame analyzed. Data include RNAV STAR waypoints (latitude/longitude), airspeed restrictions, altitude restrictions, and descent gradient (degrees) for each waypoint.

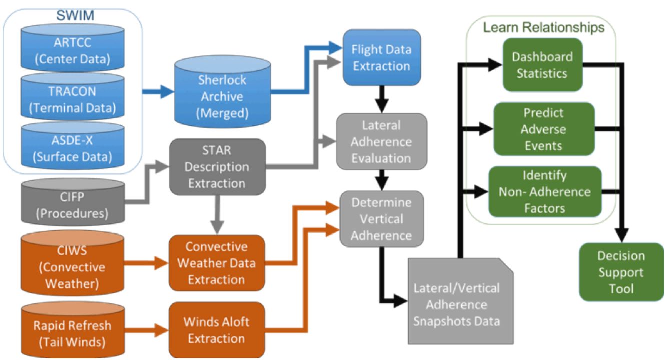

Fig. 5 System Diagram of RADI.

B. *Variables*

Variables were broken into four categories: *procedure, weather, traffic,* and *aircraft*. Due to scope restrictions, we did not include the analysis of each aircraft type. In addition, we could not detect the effects of speed restrictions on vertical profiles because we could not determine whether the airplane was adhering to the published speeds. We did not include speed variables in vertical assessments. Variables used in the analyses are listed in Table 1 and Figures 3 and 4.

TABLE I. *VARIABLE LIST*

Categories	Variables	Units	Definitions
Weather	Tailwind	Knots	Wind component for the direction of segment of STAR the aircraft is on at that altitude.
	Tailwind Differential	Knots	The difference in wind between two merging segments of a STAR.
	Weather on Compartment	40-100% <40%=0	Probability a flight will deviate convective weather at a given compartment at the time the flight passes.
	Distance to Weather	Nautical Miles	Distance from current position to weather covered segment downstream.
Procedure	Average Slope	Degrees	The high slope plus the low slope divided by two
	Distance to Previous Restriction	Nautical Miles	The distance looking back from a waypoint to the previous waypoint with a altitude restriction.
	Window Size	Feet	The size of an altitude restriction (ft) from the lower to upper limit.
	Waypoint %	% 0-100	Percent of waypoint position out of total waypoints in STAR from start.
	Merge Waypoint Type	Nominal	A waypoint where two or more transitions converge. 0=no merge 1=transitions, 2=STARs, 3=transitions and STARs.
	Number of Speed Restrictions in Sub-route	# of speed restrictions	Count of speed restrictions from enroute transition to runway transition.
Traffic	STAR Flow	# of Aircraft	The count of aircraft at a given time that are currently on the STAR assigned to the aircraft in question. Also counts *skipping* aircraft that return to the route.
	Miles in Trail	Nautical Miles	The distance the preceding aircraft is in front of the aircraft in question.
	Merge Time Differential	Seconds	Absolute difference in projected time (Sec) two conflicting flights will be at a merge.
	Regional Flight	Boolean	A flight that originated within the furthest enroute transition.
Aircraft	A319,A320, A321, A330, A340, A350, A380, B737, B747, B757, B767, B777, B787, CRJ2, CRJ7, CRJ9,DHC-8, EMB-135, EMB-145, EMB-170/5, EMB-190, MD10/11, MD80/90	Nominal	NA

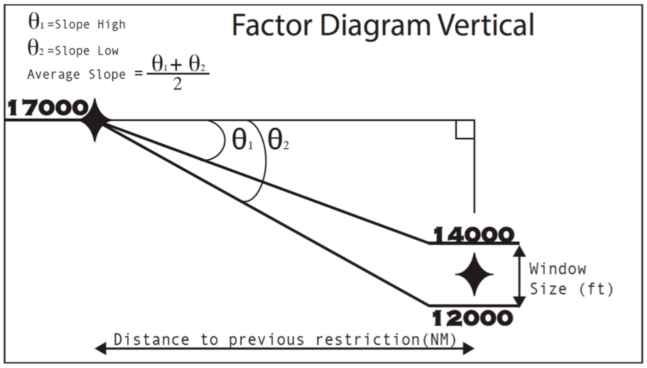

Fig. 3 Diagram of Vertical Factors.

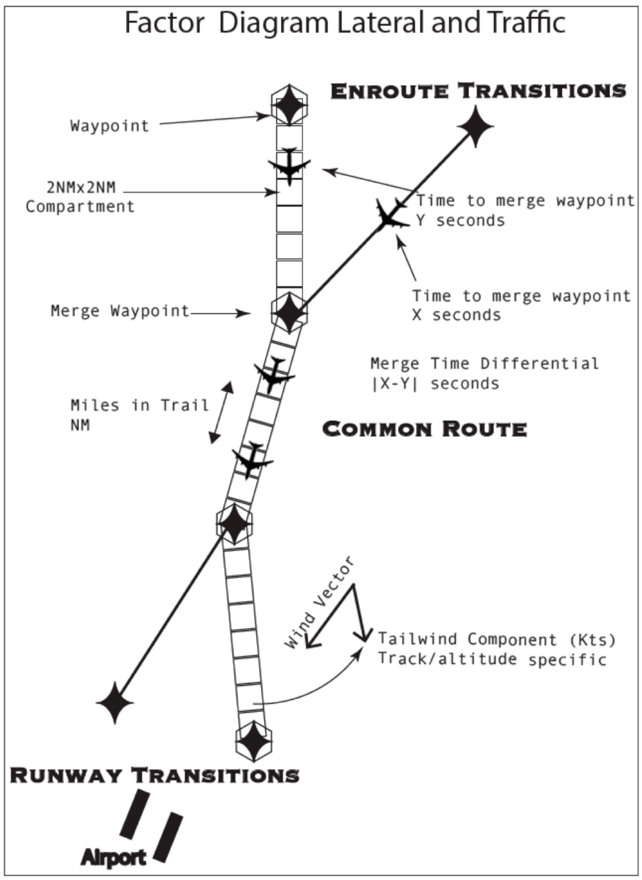

Fig. 4 Diagram of Lateral Variables.

C. *Regulatory Guidance*

Navigation system requirement guidance is provided in Advisory Circular 90-100A. Lateral navigation requirements are specified in miles from track centerline for a given amount of operational time. For RNAV 1, the system is required to be within one mile of track centerline 95% of the time. Vertical navigation requirements are not specified for RNAV 1 capable systems [2].

Guidance for procedure design exists in documents provided by the FAA. These FAA Orders are: 8260.19G, 8260.3C, and 8260.58A [13,14,15]. Lateral design standards specify waypoint spacing and type, turn angles, wind impact, and other details to ensure aircraft adherence to the procedural track. Given that there is a performance requirement (RNP) for lateral navigation, we were not focused on investigating those criteria. Vertical profiles have maximum descent gradient requirements between waypoints based on altitude and speed restrictions. Restrictions above 10,000ft are limited to 3.11°, and below 10,000ft to 3°. However, if there is a speed restriction that is less than 220 Knots (below 10,000), the maximum gradient is 2.36° both before and after the waypoint restriction. The guidance specifies a reduced descent gradient for reductions in airspeed, but does not provide the method by which it was calculated. One explanation is a reduction in drag at lower airspeeds from the reduction in dynamic pressure.

These variables illustrate that descending to a target altitude is regulated by aircraft performance limitations. Aircraft are limited by a maximum descent rate to still abide by a speed restriction. Thus, the speed restrictions and altitude restrictions are yoked. If either the descent gradient is too steep for the airplane to decelerate, the speed or altitude restriction must be prioritized over the other. Conversely, if the speed restriction were too slow, the aircraft might not be able to descend at the required rate to make the altitude. For example, if the altitude were given priority, increasing airspeed would increase the descent rate (i.e., dive to altitude); speed restriction priority would result in a reduced descent rate for a reduced speed (i.e., flatten-out to slow). In addition, controller intervention such as changing airspeed, vectoring, and changing altitude restrictions renders design standards ineffective since the interventions would not conform to the standards.

General principles for design guidance encourage using the fewest number of waypoints and waypoint restrictions to define a trajectory. Further, the guidance asks designers to consider the interaction effects of multiple restrictions in succession [13]. Human performance decrements from charting clutter resulting from navigation procedure complexity were identified by Chandra and Grayhem, 2013. Users' speed in retrieving data from charts was reduced as the amount of data increased [16]. Possible motivations for this guidance are to reduce complexity to enhance human performance, and mitigating unwanted aircraft performance from problematic restriction interactions. However, the specific negative interactions from speed and altitude restrictions are not listed.

D. *Purpose*

The purpose of this study was to identify the influence of variables outside the users' (pilots and controllers) control on specific types of non-adherence. By identifying these variables and their respective impacts, we aimed to provide empirical evidence that they contribute to differing rates of adherence. Further, we wanted to quantify the specific levels and combinations of variables that influence non-adherence to aid in design criteria. Ultimately, increasing system resilience is the motivating goal. It is a key point that we used historical data of real operations and analyzed flight performance. We did this to encompass the complexity of the domain.

The secondary purpose of this study was to introduce this analysis method as a way to measure the impact of designs and design changes (i.e., create a dependent variable set). The RADI system serves that function and acts as instrumentation for how procedures are being flown; system health cannot be measured without instrumentation.

III. Method

A. Materials

All data files were produced by the RADI system. Tableau software was used to identify trends, test variables, and test data integrity. We used Python language scikit-learn package to build decision tree models and Pandas package to load CSV data files.

B. Dataset

We analyzed 32 airports in the United States that contain 276 RNAV STARs with vertical profiles. Data collection ranged from November, 2014 to January, 2018. Atlanta (KATL) was excluded from the vertical analysis because its RNAV STARs vertical profiles were not being used due to safety issues during a section of time. "Flights" are counts of aircraft arriving, flights can contain multiple events. These are the behaviors of interest and are measurable descriptions of non-adherence (See Figs 1 and 2). The dataset contained 12,762,967 flights, of which, 10,619,068 flew some portion of a STAR. This subset contained 2,257,000 *misses below*, 952,585 *misses above*, 3,365,889 *skips before merge*, and 868,696 *skips after merge*. In addition, 1,314,306 flights complied fully vertically and laterally to the STAR.

C. Design

This study used a supervised predictive machine learning design to retrospectively analyze historical flight data. Decision tree classification was used to predict four dichotomous dependent variables (events). These were: *miss above, miss below, skip before merge,* and *skip after merge*. For each event of interest, the data were first filtered based on the specific criteria describing the event, then a separate dataset was created. Using these data, a decision tree was built. Each event was tested at one level (all STARs at all airports). This was intended to gauge the generalizability of the variables and how they influence the entire system. Hyper-parameters of a decision tree model include maximum depth of tree, minimum samples required to make a split, minimum samples to be present at a leaf node, maximum number of leaf nodes, and minimum impurity decrease required to make a split.

D. Model Specifications

Each test was made independently and was trained, tested, and validated with 60%, 20%, and 20% of the data for the specific case, respectively. To identify the best hyper-parameters of the model, a five-fold cross-validation was used. First, the training data were split into five sets, then trained on four sets and evaluated on the fifth. This was repeated five times with each set being the evaluation set. Then, average (area under curve) AUC was measured. The five-fold cross-validation was repeated on several combinations of hyper-parameter values and the average AUC was recorded. The best hyper-parameter for the model was the one that has the maximum average AUC on the five-fold cross-validation experiment. The purpose of this validation step was to test the effects of different split thresholds to maximize the true positive rates (TPR), and true negative rates (TNR). A final model was trained using the best hyper-parameters and was used as the prediction model. Class imbalance was handled by weighing the minority class to equal the majority. Last, the top three event and non-event rules with the smallest GINI impurity were reported. Event probability was calculated for each rule based on its leaf node's GINI impurity translated to rate of occurrence.

E. Test Specifications

We chose a cutoff of 60% prediction accuracy for reporting results of the decision trees.

1) Miss above and below:

In order to gauge vertical misses faithfully, waypoints without vertical restrictions were filtered. Flights that skipped the waypoint in question laterally were not assessed vertically. However, if the flight returned to the lateral path, it was vertically assessed for those waypoints flown. Thus, flights could have multiple vertical events in the same flight and also be laterally non-adherent for part of the flight.

Variables Used:

Average slope, distance to previous restriction, altitude window size, tailwind, merge waypoint type, and *waypoint percentage.*

2) Skip before and after merge:

Variables were first filtered to remove all of the skips that happened after a merge waypoint. We then identified the five compartments (ten miles) before the skip occurred and recorded the variables of interest for the flight in each compartment. These five compartments were considered the event. We assumed the variables within those compartments would be sufficiently stable and unchanged before the aircraft left the lateral path. We then compared compartments that did not have a skip and used those as the non-event comparisons.

Skips after merge waypoints used a filter to remove the events before a merge waypoint. In addition, all variables related to differentials were not applicable to that test since they only exist before routes merge.

Variables Used:

Merge time differential, Tailwind differential, STAR flow, Number of speed restrictions, Miles in trail, and *Weather on compartment, regional flight,* and *waypoint Percent.*

IV. RESULTS

A. *Miss Above*

The decision tree correctly predicted 65% of the *miss above* events and 74% of the non-event adhering flights; total accuracy was 70%. The base occurrence rate was 2.7%. Hyper parameters and training accuracy are listed in Table II. The three rules predicting the highest probabilities of the event (i.e., event class) and lowest probability of event (i.e. adherence class) are listed in Tables III and IV.

TABLE II. *MISS ABOVE* MODEL SETTING AND PEFORMANCE

Hyper-parameters	TPR Train/Test Difference	TNR Train/Test Difference
Max depth = 11 Max leaf nodes = 25 Min impurity decrease = 0.0001 Min samples leaf = 1000 Min samples split = 4000	0.00123	-0.00024

TABLE III. TOP THREE LEAF NODES FOR *MISS ABOVE* CLASS

Rank	Event Probability	Rule
1	42%	Not merge 3 & window size=1000 & Average slope >.1º
2	29%	Merge 3 & window size < 2500 & distance to previous restriction > 18 NM
3	17%	Window Size=0 & Average slope >2.9º

TABLE IV. TOP THREE LEAF NODES FOR *ADHERENT* CLASS

Rank	Event Probability	Rule
1	0.3%	Not merge 3 & window size > 1500 & Average slope < .1º
2	0.5%	Not merge 3 & window size > 4500 & Average slope > .1º
3	0.5%	merge 3 & window size > 2500 & distance to Average slope < 3.1º

Miss Below

The decision tree correctly predicted 70% of the *miss below* events and 73% of the non-event adhering flights; total accuracy was 72%. The base occurrence rate was 4.4%. Hyper parameters and training accuracy are listed in Table V. The three rules predicting the highest probabilities of the event and lowest probability of event are listed in Tables VI and VII.

TABLE V. *MISS BELOW* MODEL SETTING AND PEFORMANCE

Hyper-parameters	TPR Train/Test Difference	TNR Train/Test Difference
Max depth = 20 Max leaf nodes = 25 Min impurity decrease = 0.001 Min samples leaf = 1000 Min samples split = 2000	0.0013	-0.0002

TABLE VI. TOP THREE LEAF NODES FOR *MISS BELOW* CLASS

Rank	Event Probability	Rule
1	37%	Window Size=0 & Average slope >.8º & waypoint % < 97 & distance to previous restriction < 11.6 NM
2	31%	Window Size =0 & Average slope <.8º & waypoint % < 97 & distance to previous restriction < 10.2 NM
3	26%	Window Size =0 & waypoint % > 97 & distance to previous restriction <5.7 NM

TABLE VII. TOP THREE LEAF NODES FOR *ADHERENT* CLASS

Rank	Event Probability	Rule
1	0.7%	Window Size >3500 & Average slope > 1.9º & miles in trail > 5 NM
2	0.9%	Window Size 1000-3500 & Average slope >.44º & miles in trail > 5 NM distance to previous restriction 4.7-10.2 NM & waypoint % 52-96
3	1.2%	Window Size=0 & waypoint % > 97 & distance to previous restriction 5.7-5.9 NM

B. *Skip Before Merge*

The decision tree correctly predicted 62% of the *skip before merge* events and 57% of the non-event adhering flights; total accuracy was 60%. The base occurrence rate was 3.3%. Hyper parameters and training accuracy are listed in Table VIII. We included the top three rules for the event class listen in Table IX, but not the adhering class as it performed below 60%.

TABLE VIII. *SKIP BEFORE MERGE* MODEL SETTING AND PEFORMANCE

Hyper-parameters	TPR Train/Test Difference	TNR Train/Test Difference
Max depth = 15 Max leaf nodes = 25 Min impurity decrease = 0.0001 Min samples leaf = 7000 Min samples split = 3500	-0.0003	1.08e-5

TABLE IX. TOP THREE LEAF NODES FOR *SKIP BEFORE MERGE* CLASS

Rank	Event Probability	Rule
1	8.7%	Number of speed restrictions in sub-route <=3 & STAR Flow <=5 & merge time differential <65s & miles in trail >25NM
2	8.5%	Number of speed restriction in sub-route <=3 & STAR Flow > 5 & Regional flight = False & tailwind <17kts & miles in trail >34NM
3	7.1%	Number of speed restriction in sub-route <=3 & STAR Flow > 5 & Regional flight = False & tailwind <17kts & miles in trail >34NM & tailwind differential <3kts

C. *Skip After Merge*

This model did not reach a 60% accuracy level. Hyper parameters and training accuracy are listed in Table X.

TABLE X. *SKIP AFTER MERGE* MODEL SETTING AND PEFORMANCE

Hyper-parameters	TPR Train/Test Difference	TNR Train/Test Difference
Max depth = 17 Max leaf nodes = 25 Min impurity decrease = 0.0001 Min samples leaf = 4000 Min samples split = 1500	-0.00084	0.00027

V. DISCUSSION

The decision trees showed that *window size*, *average slope, distance to previous restriction,* and *waypoint percent* had the most impact on vertical adherence. Smaller *window sizes* were predictive of both misses above and below. *Window size* was the most impactful variable for both vertical events. *Average slope* was also predictive of both vertical events. Steeper slopes greater than 2.9º increased the *miss above* rate; shallow slopes typically lowered the rate to a point. Nearly flat slopes (<1º) were predictive of misses in both events. This could be because the ends of many procedures are flat and controllers might need to modify altitudes to increase efficiency. *Waypoint percent* was predictive of miss below events; more were near the end of the path. This is also evidence that controllers were likely intervening and the issue is logistical. It's unclear how distance to previous restriction impacts adherence and could be a poor indicator.

Lateral adherence decision trees performed less accurately than the vertical trees. The *Skip after merge* tree did not perform up to our 60% standard. *Skip before merge* was impacted by *number of speed restrictions in sub-route*, *merge time differential*, *tailwind*, and *STAR flow*. Surprisingly, increasing the number of speed restrictions consistently reduced the *skip before merge* event rate. This was not expected and will require more investigation to understand. *Merge time differentia*l of less than 69 seconds increased the *skip before merge* rate. This is logical and shows that ATC likely uses this lateral maneuver to increase spacing when two aircraft are going to meet at the same place at the same time. *STAR Flow* of six or greater increased the *skip before merge* event rate. The increase in traffic variables could be related to increases in lateral events because the separation requirements may not be met if the aircraft are left on the STAR path.

A. *Miss-classification and Decision Tree Performance*

Given that our highest accuracy was 72%, we acknowledge there is still considerable information for each event that we do not have. Possible explanations are that our variables do not totally describe the phenomena we intend. In addition, we have not identified key variables that influence the system. ATC and pilot decisions, preferences, and many traffic issues we are not measuring could all be possible reasons for misclassified events. Specifically, the *skip after merge* decision tree was not informative, indicating we do not understand many of the contributory variables.

Overall decision tree test performance trained well and generalized consistently to the test data. Although some models did not predict well, the performance of the trees in all cases was close to the training set, which indicates that the model was not over-fitting. This can be attributed to the validation step where the optimization of hyper-parameters was created. Limited predictive capability was a lack of information and not the mechanics of the method. However, portions of the model characterized operations well and yielded useful interpretable results.

B. *Interaction Effects*

The main motivation for using decision tree classification was its ability to detect non-linear trends and interactions between many variables. In addition, the ease of interpretability, based on the derived rules, makes information understandable. The derived rules for each event demonstrate the interdependencies of the domain and show that no single issue is responsible for the performance of the system. For example, we saw that average slopes greater than 2.9º and window sizes of zero (i.e., "at" restrictions) showed approximately six times more misses above than the baseline rate. However, when the window sizes increased to greater than 2,500 ft, the average slope could be less than 3.1º and adherence was better than baseline. It is likely that reducing the window size is similar to shrinking a target; therefore, increasing the precision requirement. Increasing the average slope may make speed control an issue, either artificially from speed restrictions or structurally from the maximum speed of the aircraft. This could force aircraft to shallow descent paths and miss the altitude. Further, if the slope of the procedure is steep and the altitude restriction is tight, the combination might require anticipatory actions from the pilots to ensure success. Anticipatory actions could be speed changes, faster reactions, amended flight paths, or many other adaptations. Ultimately, this could pose unnecessary complexity and increase the likelihood of errors if attention is focused elsewhere.

Combinations were not limited to two variables. We noticed other interactions with *tailwind, average slope*, and *window size*. *Tailwind* in excess of 19 knots worsened performance with *window size* between 1000ft and 4500ft and *average slopes* greater than 2.7º. This showed that two variables could be the better-than-baseline performance range and a third could impact the performance beyond the baseline. It's likely that there are many other variables involved in increasing the possibility of an event, and it will take more research to identify them.

C. *Generalizability*

All tests were performed on all of the STARs in our dataset, so variables that are found in more than one location we analyzed together for trends. This is a benefit if there is a trend in the variable and that variable is shared in different parts of the airspace (i.e., different airports). However, the decision tree models can also find trends in variables that are unique to certain procedures. Thus, a trend may not be generalizable because it was only noticed in one location. For example, the second rule for *miss above* events only pertains to one waypoint in KPHX airport called HYDRR. Whereas the third rule in the *miss above* event pertains to 27 waypoints at 14 airports.

D. Possible Rule Interpretations

Most of the non-adherence in the system is likely due to intentional adaptations. That is, pilots and controllers chose alternative flight paths for logistical reasons. Variables that might require ATC action for lateral non-adherence are *merge time differential, miles in trail*, and *STAR flow.* These variables might also require ATC intervention vertically, but we do not have the ability to detect this yet. However, *merge waypoint type* was predictive of *miss above* rates.

Other types of non-adherence fall into the unintentional category. These are altitude and track deviations that could pose a threat to the flight's separation from traffic and terrain. We believe that the physics issues associated with descent (e.g., slope, wind) have an inhibitory effect on the flight's ability to adhere to altitude restrictions. In addition, increasing precisions requirements could also influence error rates. Thus, rules showing these variables as predictive might be a source of errors in the system.

E. Practical Applications and Implications

One goal of this work was to use the RADI system to identify occurrence rates of events in the system. In addition, we can use RADI to locate the portions of procedures that the decision trees predict will have higher non-adherence rates. Locating these procedures can be done by filtering the data based on the specific rules the trees produce. This targeted search can provide possible explanations for events. Validation can then be done by searching the ASRS database for subjective safety reports to see if these events are being reported.

As an example, we applied this method to Rule 3 in Table III. We chose this rule because it had characteristics that could inhibit the flight from achieving its goal of making an altitude restriction. Figure 6 is a list of the pilot reported deviations and waypoints from procedures in the pilots' own words.

These reports show miss above events that appear to be similar to the trend of the rule: Steep profile and small altitude window target.

F. Limitations

This study does not show causal relationships between events and variables. Some of the rules generated by the decision trees are found in single instances of the airspace system. This could mean that the variables themselves are not explaining the events, but describing where a unique instance of non-adherence resides (Rules 1 and 2 in Table III). However, we do not know and will need to study this further to understand.

Although we know that some of the non-adherence events in our dataset are unintended errors, it is certain that there are many other reasons that aircraft were not flying procedures as designed. Thus, intent of pilots and controllers is not ascertainable and we do not believe that all events are unsafe.

Waypoint: LIQWD

"The Captain stated we could not make the next crossing altitude restriction for LIQWD intersection at 6,000 feet. I called out our airspeed at 224 knots to when I noticed us speeding up passed our assigned 210 knots."

Waypoint: ARRTU

"Last week, I witnessed the airplane in VNAV PATH unable to comply with published speed as it descended via the arrival between HEFLY and ARRTU. It had difficulty because the airplane inexplicably chose to cross HEFLY at the FL190 high end of the published altitude and thus had virtually no chance in crossing ARRTU at 10000 and 250 kts."

Waypoint: COPEN

"From a pilot's point of view this arrival could be better, safer, and more user friendly by simply extending the distance from BLUZZ and COPEN intersection. 9.7 miles is simply not enough to lose altitude and slow down. Could it be increased to possibly 20 or 25 miles? The chart states that from COPEN intersection to the airport is 40NM, surely there is enough room to make this possible."

Waypoint: COPEN

"I briefed that I follow the VPI (Vertical Path Indicator) as habit on ODP (Optimized Descent Profile) arrivals, and waited until the VPI carrot came alive to begin descent. While we met BLUZZ below 23000, as soon as we crossed the fix the FMS messaged "max grade unable next altitude." We advised ATC we were unable to cross the COPEN fix at 10000. ATC handed us off to approach, who gave us a vector to join the approach to 18L."

Fig. 6 ASRS Event Excerpts.

In fact, we believe most non-adherence is likely adding safety. In addition, we were limited to subject matter expertise, literature, and incident reports to identify variables of interest. More work in this area is likely needed.

Our data production method is also limited. By labeling waypoints and the compartments between waypoints differently, we lost the variables that are associated only with waypoints (e.g., *window size*). This could mean that the variables associated with the waypoints that are not carried over to the compartments near those waypoints are influencing the performance of the STAR, but remain undetected. We have not devised a logical way to account for this and cannot determine the effect of waypoint variables on the lateral adherence of a procedure.

RADI is currently limited to radar data, so airspeed is not measurable. Therefore, it is likely we are missing information that could help us find additional trends.

G. Future Work

Next steps in the work will be to integrate aircraft data into the system to replace radar as the source. This will also provide precise airspeed, groundspeed, autopilot, speed brake, fuel usage, and weight information. Additional work is needed to enhance the predictive capability of the lateral non-adherence events. Further, ethnographic observations of operations of pilots and controllers would be useful to gain a better understanding of the system.

H. Conclusion

This work demonstrates that variables in the domain that are not controlled by pilots and controllers are impacting adherence to RNAV STARs. Although we are unable to distinguish all events in terms of positive or negative, we believe that systemic non-adherence is an example of humans adapting to a domain that is too complex for a procedure to handle. Procedures that do not encompass the variability of the domain and failures are both expected and unexpected by users. Therefore, non-adherence is both positive and negative: positive when humans make the system continue to function and negative when the design puts flights in unsafe situations. We need to understand how we can compare the envisioned functionality of procedures with their actual performance to evolve. Future airspace congestion is likely to increase and having methods to capture the system's behaviors will be necessary for maintaining and increasing safety. Understanding our current strengths and limitations is vital for maturing aviation safety.

Acknowledgment

We would like to thank NASA Airspace Operations Safety Program and SMART-NAS and System Wide Safety Projects for their funding. We would also like to thank the NASA ATM Data Warehouse for access to the Sherlock, RR, and CIWS data and ATAC Corp for providing the historical Airspace Data Files. We would also like to thank Oliver Ammann for his assistance in analyzing the data.

References

[1] Faa.gov. (2018). *Fact Sheet – NextGen and Performance-Based Navigation*. [online] Available at: https://www.faa.gov/news/fact_sheets/news_story.cfm?newsId=10856 [Accessed 10 May. 2018].

[2] Advisory Circular, Department of Transportation, Federal Aviation Administration : Subject: "U.S. terminal and en toute area navigation (RNAV) operations ," AC NO: 90-100A, Initiated by: AFS – 400, April 8, 2014, [Online]. Available: https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_90-100A_CHG_1.pdf

[3] Avrekh I ., Matthews B., and Stewart M., "RNAV adherence data integration system" NASA/TM 2018-[pending]

[4] Stewart, M., Matthews, B., "Objective assessment method for RNAV STAR adherence" IEEE/AIAA 36th Digital Avionics Systems Conference (DASC), St. Petersburg Florida, September 2017

[5] C. M. Johnson, "Human-in-the-loop (HITL) simulation and analysis of optimized profile descent (OPD) operations at Atlanta," The MITRE Corp., McLean, VA, 2009. Contract DTFA01-01-C-00001 Rep. 09-3665

[6] Mercer, J. , Callantine, T. J, & Martin, L. H "Resolving off-nominal situations in schedule-based terminal area operations: results from a human-in-the-loop simulation." 28th International Congress of the Aeronautical Sciences, 2012

[7] Reynolds, H.J.D. and Reynolds, T.G. and Hansman, R.J. "Human factors implications of continuous descent approach procedures for noise abatement in air traffic control," Proc. 6th USA/Europe Air Traffic Management R&D Seminar, Baltimore, MD, June, 2005

[8] Harding Ken, NOAA's National Weather Service, "Thunderstorm formation and aviation hazards," Topeka, KS, July 2011, [Online]. Available: https://www.weather.gov/media/publications/front/11jul-front.pdf

[9] Sprong Kevin R., and Mayer Ralf H., "Analysis of RNAV arival operations with descend via clearances at Phoenix airport" The MITRE Corporation, McLean, VA 26th Digital Avionics Systems Conference October 21, 2007

[10] Commercial Aviation Safety Team (CAST). (2014). SE213: Area Navigation (RNAV) – Safe Operating and Design Practices for STARs and RNAV Departures - SKYbrary Aviation Safety. [online] Available at:https://www.skybrary.aero/index.php/SE213:_Area_Navigation_(RNAV)_%E2%80%93_Safe_Operating_and_Design_Practices_for_STARs_and_RNAV_Departures [Accessed 11 May. 2018].

[11] Commercial Aviation Safety Team (CAST). (2014). SE214: Area Navigation (RNAV) – Procedures and Standards to Improve Path Compliance for STARs and RNAV Departures - SKYbrary Aviation Safety. [online] Available at: https://www.skybrary.aero/index.php/SE214:_Area_Navigation_(RNAV)_%E2%80%93_Procedures_and_Standards_to_Improve_Path_Compliance_for_STARs_and_RNAV_Departures [Accessed 11 May. 2018].

[12] Michael P Matthews, Rich DeLaura, "Modeling convective weather avoidance of arrivals in the terminal airspace," 10th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, Fort Worth, Texas, September 2010

[13] Advisory Circular, Department of Transportation, Federal Aviation Administration : Subject: "Flight procedures and airspace," Order 8260.19G, July 14, 2015, [Online]. Available: https://www.faa.gov/documentLibrary/media/Order/FAA_Order_8260_19G.pdf

[14] Advisory Circular, Department of Transportation, Federal Aviation Administration : Subject: "United States standard for terminal instrument procedures (TERPS)," Order 8260.3C, March 14, 2015, [Online]. Available: https://www.faa.gov/documentLibrary/media/Order/FAA_Order_8260.3C.pdf

[15] Advisory Circular, Department of Transportation, Federal Aviation Administration : Subject: "United States standard for performance based navigation (PBN) instrument procedure design," Change 8260.58A, March 20, 2017, [Online]. Available: https://www.faa.gov/documentLibrary/media/Order/Order_8260_58A_CHG_1_Only.pdf

[16] Chandra, D. C., & Grayhem, R. "Evaluation of a technique to simplify depictions of visually complex aeronautical procedures for NextGen," Proceedings of the Human Factors and Ergonomics Society 57th Annual Meeting, San Diego, CA, 86-90, 2013. doi: 10.1177/1541931213571021.