Predicting the Operational Acceptance of Airborne Flight Reroute Requests using Data Mining

Antony D. Evans, Paul Lee, Banavar Sridhar

Abstract

For tools that generate more efficient flight routes or reroute advisories, it is important to ensure compatibility of automation and autonomy decisions with human objectives so as to ensure acceptability by the human operators. In this paper, the authors developed a proof of concept predictor of operational acceptability for route changes during a flight. Such a capability could have applications in automation tools that identify more efficient routes around airspace impacted by weather or congestion and that better meet airline preferences. The predictor is based on applying data mining techniques, including logistic regression, a decision tree, a support vector machine, a random forest and Adaptive Boost, to historical flight plan amendment data reported during operations and field experiments. Cross validation was used for model development, while nested cross validation was used to validate the models. The model found to have the best performance in predicting air traffic controller acceptance or rejection of a route change, using the available data from Fort Worth Air Traffic Control Center and its adjacent Centers, was the random forest, with an F-score of 0.77. This result indicates that the operational acceptance of reroute requests does indeed have some level of predictability, and that, with suitable data, models can be trained to predict the operational acceptability of reroute requests. Such models may ultimately be used to inform route selection by decision support tools, contributing to the development of increasingly autonomous systems that are capable of routing aircraft with less human input than is currently the case.

Keywords: Air Traffic Management; Airborne Reroute; Data Mining; Operational Acceptability

Table of Figures

Figure 1. Approach to classification of routes as ATC accepted or rejected. The classified routes include DWR advisories and air traffic controller implemented Center route amendments. 23
Figure 2. Common Routes from sector ZFW48 to waypoint ABQ (observed more than 10 times in April 2015). Routes are shown from the first fix after ZFW48 is entered to the final route fix, ABQ. Relative historical route counts are indicated by the thickness of the lines between fixes. High Power VORs are shown in blue. Center and high-altitude sector boundaries are shown in gray. 24
Figure 3. Comparison of (a) full reroute approach and (b) segment approach to identifying reroute usage in historical data. 25
Figure 4. Reroute Event Schematic.. 26
Figure 5. Methodology used for Data Mining. 27
Figure 6. Nested *K*-fold cross validation.. 28
Figure 7. ROC curves for forward search feature selection, applying a random forest in the prediction of ATC operational acceptability of DWR reroute advisories from May to September 2014. Filled circles represent the points with 0.5 discrimination threshold presented in Table 2. (RSS: Reroute Start Sector; D/C: Demand over Capacity) 29
Figure 8. ROC curves for model selection, in the prediction of ATC operational acceptability of DWR reroute advisories from May to September 2014. Filled circles represent the points with 0.5 discrimination threshold presented in Table 3. 30
Figure 9. ROC curve for a random forest with the specified number of trees, in the prediction of ATC operational acceptability of DWR reroute advisories from May to September 2014. Filled circles represent the points with 0.5 discrimination threshold presented in Table 4. 31
Figure 10. ROC curve for nested cross validation results for the prediction of ATC operational acceptability of DWR reroute advisories from May to September 2014. Filled circle represents the point with 0.5 discrimination threshold presented in Table 5.. 32

Table of Tables

Table 1. Common routes (observed more than 10 times in April 2015) originating in ZFW48, filtered for the final route fix ABQ. Route count generated from flight plans and Center route amendments extracted from ASDI data for April 2015. Dash (-) under 'Via' represents a direct route from the route start sector to the final route fix. 33
Table 2. Forward search feature selection results (F-Score), applying a random forest for the prediction of ATC operational acceptability of DWR reroute advisories from May to September 2014. Order of features selected in parentheses in column 1. (RSS: Reroute Start Sector; D/C: Demand over Capacity). 34
Table 3. Model selection results (F-Score) for the prediction of ATC operational acceptability of DWR reroute advisories from May to September 2014. 35
Table 4. Parameter selection results (F-Score) for a random forest with the specified number of trees, for the prediction of ATC operational acceptability of DWR reroute advisories from May to September 2014. 36
Table 5. Nested cross validation results for the prediction of ATC operational acceptability of DWR reroute advisories from May to September 2014. 37

Predicting the Operational Acceptance of Airborne Flight Reroute Requests using Data Mining

Antony D. Evans, Ph.D.[*]
Crown Consulting Inc., NASA Ames Research Center, Moffett Field, CA, 94035

Paul Lee, Ph.D.[†]
NASA Ames Research Center, Moffett Field, CA, 94035

Banavar Sridhar, Ph.D.[‡]
Universities Space Research Association, NASA Ames Research Center, Moffett Field, CA, 94035

1. Introduction

In a safety-critical system such as Air Traffic Management (ATM), system evolution and the introduction of increasing autonomy should be done in stages. Human-system integrated concepts are critical to ensure compatibility of automation and autonomy decisions with human objectives, ensuring acceptability by human operators, and enabling graceful degradation during failures or off-nominal conditions (Edwards and Lee, 2017). Therefore, it is essential that automated support systems work effectively with the humans that operate them. For this reason, automation should be designed to be understandable and acceptable to human operators including airline dispatchers, air traffic controllers and air traffic managers.

For automation tools, including decision support tools, to be successfully implemented in a mixed human-automation environment, there is a need to ensure their outputs have high operational acceptance, i.e., have a high probability of being accepted by human operators. One example is for tools that generate flight reroute advisories. If the reroute advisories generated by decision support tools were more often operationally acceptable, a higher number of them would be accepted by air traffic controllers for implementation in the form of a flight plan amendment, thus increasing the benefit of the tools. Increased operational acceptance would also reduce the workload burden of the tool on the air traffic controller per requested reroute, because fewer reroute modifications would have to be made before it could be implemented. Ultimately, it could also free up air traffic controller's time to work on other problems, with further benefits to the system.

One application for which the generation of operationally acceptable routes is particularly important is Trajectory Based Operations (TBO) (SESAR, 2007; Federal Aviation Administration, 2014). TBO is an air traffic management paradigm that allocates user-negotiated four-dimensional trajectories to flights, which are updated dynamically as conditions change. For success of the concept in a mixed human-automation environment, these routes should be generated in such a way as to maximize the probability of operational acceptance.

A number of studies have examined elements of route acceptability to Air Traffic Control (ATC). Taylor and Wanke (2012) define an optimization approach that generates operationally acceptable reroutes for flights predicted to request deviations from their current routes around weather. The method considers many factors including route deviation distance, conformance of the reroute to historically flown routes, weather impact on the current route, sector congestion, required point-outs[§] and inter-facility coordination.[**] The route network used for the optimization was generated by segmenting historically flown routes into fix-pair arcs. Thus, all arcs in the modeled network consist of previously-flown connections between fixes, so each individual arc in the network has some level, depending on usage, of operational acceptability. Reroutes are constructed from these arcs using an optimization algorithm, and the

---

[*] Senior Scientist I, Aviation Systems Division, M/S 210-6, corresponding author.
[†] Senior Research Associate, Human-Systems Integration Division, M/S 262-4.
[‡] Senior Scientist, Aviation Systems Division, M/S 210-10.
[§] A point-out refers to the requirement that one air traffic controller requests that the air traffic controller of an adjacent sector also monitors a flight that is flying close to the boundary between their sectors.
[**] If a reroute takes the aircraft into a different facility, such as a Center, coordination may be required between the two facilities. This is particularly true if the route by which the aircraft enters the new facility is non-typical.

reroutes that best meet a set of metrics of operational acceptability are presented as potential alternatives to users (Taylor and Wanke, 2009).

Other related research includes Stewart et al. (2012), which describes a concept for reroutes around convective weather that leverages technologies to automate the coordination between traffic managers and air traffic controllers. The concept assumes the use of the Corridor Integrated Weather System (CIWS) and Convective Weather Avoidance Model (CWAM) for weather detection and in reroute generation to avoid the weather, and incorporates route acceptability factors described by Taylor and Wanke (2012). Ballin and Wing (2012) describe the NASA developed Traffic Aware Strategic Aircrew Requests (TASAR) concept, which combines Automatic Dependent Surveillance-Broadcast (ADS-B) and airborne automation to enable user-optimal in-flight trajectory re-planning and to increase the likelihood of ATC approval for the resulting trajectory change request, incorporating traffic, weather, and airspace information in the optimization process.

In the approaches used by Taylor and Wanke (2009, 2012), Stewart et al. (2012), and Ballin and Wing (2012) the factors used to evaluate the operational acceptability of routes are identified through observations and feedback from Subject Matter Experts (SMEs). However, the ability of these factors to predict route operational acceptability has not been validated. Because of the variety of conditions under which reroute requests are made, and the nature of how air traffic controllers make decisions about reroute requests, relying on observations and SME input alone may also have limitations. This is because air traffic controller decision-making is highly complex, based on many different factors and how these different factors interact. Air traffic controller decision-making, therefore, varies significantly under different conditions, and it is difficult, even for SMEs, to explain all the drivers of their decisions to researchers. Operational data, however, if mined effectively, has the potential to capture many of these complexities. A predictor of the operational acceptability of reroute requests, trained on historical operational data, could capture the impact of different factors for which features can be calculated, and how these different factors contribute to the overall acceptance or rejection of a reroute request.

Data mining techniques have been applied to learn trends and correlations from historical data in many fields, and ATM is no exception (e.g., Pfeil and Balakrishnan, 2012; Wang and Kulkarni, 2011; Bloem et al., 2012; Kulkarni et al., 2014; Kuhn, 2016; Li et al., 2016; Balakrishna et al., 2010; Hrastovec and Solina, 2016). However, to the authors' knowledge, data mining has not yet been used to train predictors of route operational acceptability. Given suitable historical data, data mining techniques could provide an effective tool to identify the degree to which the operational acceptance of reroute requests is predictable, and to build a predictor that would estimate the probability that a reroute request would be accepted or rejected by ATC. They also provide the opportunity to identify the dominant drivers of ATC acceptability from the predictor feature set.

The contribution of this paper is (1) the development of a method to quantify the degree to which the operational acceptability of reroute requests is predictable, by developing a sample predictor of route operational acceptability based on the application of data mining algorithms to currently available historical data on route acceptance; (2) demonstration of a proof of concept development of a predictor of route operational acceptability in the Fort Worth Air Route Traffic Control Center (ARTCC, or Center) (ZFW) and adjacent Center airspace, using data mining algorithms; and (3) providing a validation that the features identified in the literature do in fact predict route acceptability, demonstrated by their use in training a predictor of route operational acceptability using data mining algorithms. Limitations do exist, however, on how the developed predictor can be used, because the data used is specific to ZFW and its adjacent centers. With suitable data, the methodology demonstrated in this paper can be used to develop models in future work that predict the operational acceptability of reroute requests in other airspace. These could be used to inform route selection in future decision support tools, ultimately contributing to the development of increasingly autonomous systems that are capable of routing aircraft with less human input than currently required.

Background on NASA and Federal Aviation Administration (FAA) tools and concepts under development to enable TBO, with a particular focus on user-negotiated flight route changes, are described in Section 2. This section also introduces the Dynamic Weather Routes tool, a trial of which at American Airlines provides the training and test data needed for developing the predictors of route acceptability in this paper. The approach for predicting the operational acceptability of route requests is described in Section 3, including details of the data used, the features considered, and the data mining algorithms applied to build a classifier that predicts whether a reroute request is operationally acceptable or not. Results from training the algorithms are presented in Section 4, along with a list of the features identified as being most important, and the performance of the model on a test set. Conclusions are presented in Section 5.

2. Background

A number of tools and concepts are under development by NASA (Ballin and Wing, 2012; Idris et al., 2017; McNally et al., 2012; McNally et al., 2015; Sheth et al. 2016; Gong et al., 2015; Zelinski et al., 2016) and the Federal Aviation Administration (FAA)(Kim and Hansen, 2013) to implement TBO in order to increase the capacity, efficiency and accessibility of the United States' National Airspace System (NAS). Some of these concepts and tools, which operate on a variety of time scales and horizons, identify and recommend changes in flight routes that avoid adverse weather and congestion, or allow increased opportunities for airlines and operators of future aviation concepts, such as unmanned aerial systems and urban air mobility, to request routes that meet their preferences. The effectiveness of each of these tools and concepts could be improved by generating reroute advisories that have increased operational acceptability. Particularly, trials of the Dynamic Weather Routes (DWR) concept at American Airlines (McNally et al., 2012; McNally et al., 2015), provide data for analyzing operational acceptability, and for training classifiers that can predict whether a route will be operationally acceptable or not. Data on when reroute requests are not operationally acceptable are particularly difficult to obtain, as they are not typically recorded, making the DWR trial data highly valuable. DWR also provides an example application of where a predictor, such as that developed in this paper, could be used.

DWR is a trajectory automation system that continuously and automatically analyzes trajectories of flight en-route to find simple modifications to their current routes that can save significant flying time and be easily communicated to pilots and air traffic controllers, while avoiding weather and considering traffic conflicts, airspace sector congestion, blocked Special Use Airspace (SUA), and FAA route restrictions (McNally et al., 2012). DWR users, including airline flight dispatchers, are alerted when a route change for a flight can potentially save more than a user specified minimum amount of flight time. Interactive automation enables users to visualize proposed reroutes, modify them if necessary, evaluate key parameters and provide the route modification to pilots for further consideration. After reviewing and accepting the proposed reroute, the pilot can verbally make the reroute request from the air traffic controller operating the sector in which the aircraft is located at that time. The air traffic controller can then reject, modify or accept the reroute, before communicating it back to the pilot verbally and in the form of a Center route amendment.[††] DWR reroute advisories update every 12 seconds as the most recent Center radar track and flight plan data are received. McNally et al. (2012) includes a complete description of the DWR system, while McNally et al. (2015) includes an updated description of the "Autoresolver" algorithm used in DWR to compute routes around modeled weather.

The DWR system is currently used in ZFW, and processes all enroute flights in the Center airspace as well as those in the first-tier adjacent Center airspace. Potential savings for all flights in ZFW airspace, corrected for savings flights achieved today through pilot requests and air traffic controller clearances, were estimated to be about 100,000 flight minutes for 15,000 flights in 2013. Test results from an operational trial conducted by NASA and American Airlines indicate actual savings of 3,290 flying minutes for 526 American Airlines revenue flights from January 2013 through September 2014 (McNally et al., 2015).[‡‡] However, many of the route advisories generated by DWR – 78% – were never reviewed by dispatchers, in part for reasons of high workload (in this trial the tool was also not always monitored). Furthermore, of those DWR route advisories reviewed by dispatchers, only 65% were rated acceptable by dispatchers, and of these, only 40% indicated actual observed savings for American Airlines flights in the form of a Center route amendment within 30 minutes of the advisory being accepted by the dispatcher, indicating that these advisories were acted upon by ATC. This suggests that as much as 38% of the dispatcher accepted routes were rejected by ATC. In a large majority of cases, once American Airlines dispatchers accept the route advisories, the observed Center route amendments that follow do not exactly match the route accepted by the dispatcher, so were modified by ATC. This was the case for 78% of the Center route amendments analyzed this paper. These results suggest that the operational acceptability of DWR advised routes could be improved, which would likely reduce dispatcher and air traffic controller workload, allowing more DWR advised routes to be evaluated by dispatchers and air traffic controllers. This would likely result in delay savings for more flights.

In this paper a predictor of a proposed route's operational acceptability is developed using data from the DWR trial at American Airlines. Limitations in the data mean that the developed predictor may only be valid for flights in ZFW and its adjacent Center airspace. However, development of such a predictor can quantify the degree to which the operational acceptability of reroute requests is predictable, for the dataset available, and therefore whether a predictor such as that developed, but trained on appropriate data, could be valuable more broadly. The developed predictor may also provide a proof of concept for the development of predictors of route operational acceptability.

[††] A Center route amendment is a flight plan amendment entered into the host computer.

[‡‡] The actual savings in this trial are significantly lower than the 100,000 minute potential savings described earlier because the trial involved American Airlines flights only, and the tool was not in constant use.

3. Approach

As is common in the development of predictors using data mining, the approach used to build a predictor of the operational acceptance of reroute requests included the following steps (MathWorks, Inc., 2018):

1. Extracting data on route acceptance and rejection, which allows for supervised learning algorithms to be used in the development of the predictor;
2. Identifying the set of factors that are thought to affect air traffic controller decision making, for which data are extracted to define the features on which the data mining algorithms are trained. A feature identification algorithm is then used to extract from this list of features which have an impact on air traffic controller decision making in accepting or rejecting reroute requests.
3. Identifying the data mining algorithms that best fit the extracted data using a model selection algorithm.
4. Identifying the parameters associated with the chosen data mining algorithm that best fit the extracted data using a parameter selection algorithm.
5. Validation of the developed model.

The availability of data on the acceptance and rejection of reroute requests by ATC, on which the data mining algorithms can be trained, is described in Section 3-3.1 below. This is followed by a description of the features used by the algorithms, and how they are generated, in Section 3-3.2. Section 3-3.3 describes the predictor development, including feature selection, model selection and parameter selection. This section includes a description of the data mining algorithms applied. Finally, Section 3-3.4 describes how the predictor is validated against historical data.

3.1. Data Availability

Data availability is a key constraint, and one of the biggest challenges when applying data mining techniques to problems in ATM, because of the broad range of information used by air traffic controllers and traffic managers to make decisions, data for which may not all be available for mining. Because of the improved model performance, ideally two classes of observations are desired, allowing models to be trained that can classify new observations as belonging to one of the two classes – called a two-class classifier. In the context of predicting the operational acceptability of reroute requests, this means identifying which reroute requests were rejected by ATC, with associated features, and which were accepted. Accepted reroute requests are easily identified in the form of recorded Center route amendments, but rejected reroute requests are much more difficult to identify, as they are not typically recorded. Single class classifiers, which apply anomaly detection techniques to classify new observations as belonging to a single class or not, may be used to build predictors based on observations of just one of the classes (e.g., accepted requests), but their accuracy is generally poorer than two-class classifiers, so the latter is preferable.

During the DWR trials at American Airlines (McNally et al., 2015), all proposed and dispatcher-accepted reroute advisories generated by DWR were recorded. No data are available on which of these dispatcher-accepted reroute advisories were communicated to the pilot by the dispatcher, and which were requested from ATC by the pilot. However, the ATC response to a reroute request can be surmised based on whether or not the flight had a Center route amendment implemented and recorded within 30 minutes of its DWR reroute advisory being accepted by the dispatcher. This means that dispatcher-accepted DWR reroute advisories that did not result in a Center route amendment, or resulted in Center route amendments that differed from the dispatcher-accepted reroute requested, were also recorded. Both of these advisories can be considered to represent ATC rejected reroute requests, if it is assumed that all dispatcher-accepted reroute advisories were communicated to the pilot and then requested of ATC without change. With this assumption, these data provide a useful dataset for which two-class classifiers can be developed to predict ATC route acceptability.

The approach used in this paper is therefore to use two-class classification to train and test data mining algorithms on the ATC rejected and accepted reroute advisories recorded by DWR over 5 months from May 9 through September 30, 2014, making the assumption that all dispatcher-accepted reroute advisories were communicated to the pilot and then requested of ATC without change. A total of 544 observations were extracted over this period.

Results differ depending on how the acceptance and rejection of reroute requests by ATC are defined. All DWR reroute advisories that had no Center route amendment implemented within 30 minutes of it being accepted by the dispatcher (to ensure enough time for coordination) were classed as rejected by the air traffic controller. An example of such a reroute advisory is shown in the top advisory in Figure 1 (from sector ZFW48 to waypoint DVC, via waypoint TXO) where no Center route amendment (solid line) was recorded within 30 minutes. These types of observations represent only 9% of the observations extracted. All observed Center route amendments that were made within 30 minutes of a DWR route advisory being accepted by a dispatcher were classed as accepted by the air traffic controller. They include observed Center route amendments that match the DWR reroute advisory accepted by the dispatcher, such as that shown in the middle advisory in Figure 1 (direct from sector ZFW48 to waypoint ABQ), and observed Center route amendments that are a modification of the accepted DWR route advisory requested, such as that shown in the bottom advisory in Figure 1 (from ZFW65 to ALS, via ACT and PNH). These represent 13% and 47% of observations, respectively. In the latter case, the original dispatcher-accepted reroute advisory that was modified by the air traffic controller before being implemented (from ZFW65 to ALS, via JEN in the bottom advisory in Figure 1) can be classed as either accepted or rejected by the air traffic controller, depending on how the predictor is to be used. From the perspective of the airline, these reroute advisories did lead to delay savings in the form of Center route amendments, even if the actual reroutes implemented ultimately differed from the dispatcher-accepted advisories. This would suggest that, were the predictor to be used by airlines, it may make sense to class these original dispatcher-accepted reroute advisories as ultimately accepted by the air traffic controller. However, from a system perspective, there could be workload benefits from generating reroute requests that are not only acceptable to air traffic controllers after modification, but acceptable to air traffic controllers before any modification. In this case, for reroute advisories that are modified, only the implemented Center route amendment should be classed as ATC accepted, while both the original dispatcher-accepted reroute advisory that is later modified by the air traffic controller, and reroute advisories that do not result in any Center route amendment, should be classed as ATC rejected, as shown in Figure 1. For the results presented in this paper, we applied this latter definition. Note that, consequently, for a single advisory that is modified by ATC, two observations are available – the original advisory, that is considered ATC rejected, and the implemented Center route amendment, that is considered ATC accepted. If more Center route amendments were implemented for the flight, each is classed as ATC accepted, providing further observations. Reroute advisories that are modified by ATC, and therefore classed as ATC rejected reroute requests, represent 31% of observations.

Place Figure 1 about here

3.2. Feature Identification

A number of different factors have been identified in the literature as being important to air traffic controller acceptance or rejection of reroute requests. Factors considered in designing reroutes by Taylor and Wanke (2009, 2012) and Stewart et al. (2012) included reroute deviation distance, conformance of the reroute to historically flown routes, weather impact on the current route, sector congestion, required point-outs and inter-facility coordination. Idris et al. (2017) identified similar factors impacting air traffic controller acceptability of trajectory change requests based on observations and air traffic controller interviews. The most common reasons for rejecting requests in this study were conflicts with other traffic, violating letters of agreement (LOAs)[§§] and negatively impacting neighboring sector workload, major arrival and departure flows and flow restrictions. Other notable reasons for rejection included: the aircraft already being handed off to another sector, intruding into an active SUA, intruding into another Center, weather, and unfamiliarity with the requested waypoints.

Given these factors identified in the literature, and based on SME feedback from a retired traffic manager familiar with ZFW and the adjacent Center airspace, four groups of features impacting air traffic controller acceptance are included:

1. Features describing historical reroute usage;
2. Features describing congestion along the reroute;
3. Features describing reroute deviation; and
4. Features describing the reroute starting point.

These features, and how they capture the features identified in the literature, are described in detail in the subsections below. It is important to note that one key feature described by Taylor and Wanke (2012) that is relevant to the operational acceptability of reroute requests is weather impact. This is not included in the feature set presented in this paper, because all reroute advisories generated by DWR already avoid the forecast weather. However, it is important to recognize that this is also a critical feature.

3.2.1. Features Describing Historical Route Usage

Two features (with abbreviations underlined) are included in the feature set to describe historical usage on the proposed reroute. The details of how they are calculated and the reason for their use are described below.

1. *Hist. Count (Full Reroute)*: The full reroute count, quantifying the usage of the full reroute, from reroute start point to the return point on the current flight plan, in historical data (June to August 2015);

[§§] LOAs are negotiated by air traffic control when operational or procedural needs require the cooperation and concurrence of other persons, facilities or organizations.

2. *Hist. Reroute Count by Segment*: The reroute count by waypoint-pair segment, in historical data (June to August 2015).

Features describing historical usage capture elements of many of the features listed in the literature, including conformance of reroutes to historically flown routes, the requirement for point-outs and inter-facility coordination, LOAs, major arrival and departure flows, SUA, Center boundary crossings, and familiarity with the requested waypoints. It is difficult to include many of these features explicitly because of limitations in the data used. Many of them, such as LOAs, are only available in text format in various documents across different facilities rather than in a digitally extractable database format in a central location. They are, however, implicitly captured by historically flown routes. For example, historically flown routes implicitly comply with LOAs, so if a route appears historically, it can be assumed to comply with LOA. Historical usage of flown routes is therefore taken as a proxy for these features. For this paper, historically observed routes were extracted from flight plans and Center route amendments in historical Aircraft Situation Display to Industry (ASDI) data from June to August 2015 for all sectors in ZFW and its adjacent Centers. From this dataset, tables of historically observed routes between origin sectors and destination waypoints were generated, with counts of historical usage. These tables will be referred to as *common route tables*.

June to August 2015 was chosen for extracting historically observed routes because it captures periods of high traffic and convective weather activity in the region studied, both of which can lead to the rerouting of flights. There is also no overlap with the period in 2014 for which DWR reroute advisories and Center route amendments were collected (May to September 2014), which are the routes for which historical usage is to be quantified. While the usage counts are not technically 'historical' (2015 is after 2014), they do not include the reroute for which historical usage is being quantified. Reroutes were only extracted for local times between 05:00:00 and 23:59:59 to ensure that the reroutes identified did not represent nighttime operations when a wider variety of routes may be acceptable because of very low traffic.

An example of a partial common route table generated for the month of April 2015 is shown in Table 1, with the listed routes plotted in Figure 2. The table is filtered for routes from ZFW48 (a sector in Fort Worth Center) to the Albuquerque VHF (Very High Frequency) Omni-directional Radio range waypoint (VOR) (ABQ). Note that routes with waypoints in fix-radial-distance (FRD) format or latitude-longitude pairs that did not define the first waypoint on the reroute were discarded, as they are difficult to communicate verbally. Reroute start points in FRD format or latitude-longitude pairs are, however, included in the table counts because reroute start points are often in this format (route start points are not, however, shown in Table 1). Tables were only generated for high-altitude and super high-altitude sectors because DWR only operates in en-route airspace. Some of these routes also represent arrival flows into an airport (in this case Albuquerque International Sunport), which are not applicable to DWR flights.

Place Table 1 about here
Place Figure 2 about here

Table 1 indicates that the most common route observed in April 2015 from sector ZFW48 to ABQ was of the form: PNH.TCC.ACH.CLUMP.ABQ. Note in Figure 2 that PNH is not in ZFW48. It represents the first waypoint on the route after it crosses the ZFW48 boundary. In this case, there is no waypoint on that route inside ZFW48. Also, not all 526 occurrences of this route had the same previous fix to PNH. All routes that passed through ZFW48 and then followed the route described are included in this count. Furthermore, the 8th most common route (MRMAC.IRW.CRUSR.GOONI.PNH.TCC.ACH.CLUMP), includes the most common route (PNH.TCC.ACH.CLUMP). Therefore, the 44 occurrences of the 8th most common route are included in the 526 occurrences of the most common route quoted.

Figure 3 illustrates the two approaches used to identify how often the observed DWR reroute advisories and Center route amendments from 2014 appeared in the ASDI data from 2015. Full reroute historical counts were identified by directly comparing the full reroute, from the reroute start point to the return point on the current flight plan (referred to as the return capture fix), to the historically observed routes listed in the common route tables, and, when identified, recording the historical count for that route from June to August 2015. However, because full reroutes often have relatively low counts in the data, the DWR reroute advisories and Center route amendments from 2014 were also broken up into the component segments linking all the waypoints in the route. These were each compared to the common route tables. For each given DWR reroute advisory or Center route amendment, a new historical reroute count by segment was then generated, based on the counts of the reroute's component segments from the common route tables. Specifically, the reroute count by segment is taken to be the minimum of the counts of all the component segments of that reroute. Reroute counts by segment are always equal to or higher than full reroute counts.

Place Figure 3 about here

3.2.2. Features Describing Congestion

Four features are included that describe congestion levels on the proposed reroute. These are listed below. Details of how they are calculated and the reason for their use are described in the paragraphs following.

3. *RSS D/C Ratio*: The ratio of demand to capacity (D/C) for the sector in which the reroute would start (the reroute start sector, RSS). Demand is defined as the number of aircraft projected to be in the reroute start sector when the flight in question started the reroute, including the rerouted flight. Capacity is defined by the sector Monitor Alert Parameter.[***]
4. *RSS Over Capacity*: Whether the reroute start sector would be over capacity or not when the reroute started.
5. *No. Sectors Over*: The number of sectors between the reroute start point and return capture fix in which projected demand would exceed capacity.
6. *Max D/C Ratio*: The maximum ratio of projected demand to capacity across all sectors between the reroute start point and return capture fix, including the rerouted flight.

Features describing congestion capture elements of some of the other features described in the literature, including sector congestion, conflicts with other traffic, and sector workload. Particularly, one feature that may be of importance is whether or not the requested reroute would require close monitoring to avoid a potential conflict.[†††] This is the case if the reroute took the flight towards another aircraft downstream. Acceptance of such a reroute would require monitoring by the air traffic controller to ensure that no potential conflicts with the other aircraft arise. In this case, a more likely response would be a modification to the requested reroute to route the aircraft away from the other aircraft, which would require less monitoring. Because of limitations in the data available, a feature explicitly capturing this effect was not included in this paper. Instead, the four features listed above were included.

The ratio of demand to capacity for the reroute start sector (*RSS D/C Ratio*), and whether or not demand exceeds capacity (*RSS Over Capacity*), give an indication of the workload level for the air traffic controller that has to make the decision concerning whether to accept or reject the requested reroute. If the ratio of demand to capacity is high, there is also more likely to be a potential conflict that the air traffic controller must resolve if he or she were to accept the reroute. These features represent proxies for the feature described above identifying whether or not the requested route would require close monitoring to avoid a potential conflict. Demand, defined by the number of aircraft projected to be in the reroute start sector when the flight in question started the reroute, was estimated by simulating the traffic from the time when the dispatcher accepted the reroute advisory, instead of using actual observed traffic counts from when the reroute was to start. This provides a more realistic representation of the information available to the air traffic controller when making his/her decision to accept or reject the route request.

The other two features described above (*No. Sectors Over* and *Max D/C Ratio*) account for the congestion and workload levels for downstream sectors, identified as a feature by Idris et al. (2017). The air traffic controller taking the request does not have access to explicit demand and capacity information for the downstream sectors (such as through the Flight Schedule Monitor,[‡‡‡] which is only available to the traffic manager). However, the air traffic controller is likely able to observe and project how busy other air traffic controllers are in their specific area; hence, these features are included. If it was a traffic manager making the decision, these features would likely be more important. Again, if the ratio of demand to capacity is high downstream, there is also more likely to be a potential conflict that a downstream air traffic controller must resolve if the reroute is accepted. Projected demand in each sector was also calculated by simulating traffic from the time when the dispatcher accepted the reroute advisory, providing a more realistic representation of the information available to the air traffic controller when making his/her decision

All congestion related features were generated for each dispatcher-accepted DWR route advisory and Center route amendment in the DWR trial data from 2015 using the Future ATM Concepts Evaluation Tool (FACET), which simulated air traffic, and computes sector counts. Sector counts were projected based on active flight plans at the time when the DWR reroute advisory was accepted by the dispatcher, or when the Center route amendment was made, and the resulting traffic in each sector at the times when the flight in question was projected to be in that sector. Projected sector counts were, however, based on the actual departure times of aircraft that had not yet departed at the time of

[***] The Monitor Alert Parameter (MAP) is a number defining the maximum number of aircraft in a sector that a controller can reasonably handle. Sector counts greater than the MAP can lead to increases in controller workload, and potentially impacts on safety.

[†††] A conflict refers to a loss of minimum separation requirements between two aircraft.

[‡‡‡] The Flight Schedule Monitor is a decision support tool used by airlines and the FAA that presents a graphical and timeline presentation of airport and airspace demand and capacity information.

the projection, and not scheduled departure times, as would be the case in the Flight Schedule Monitor. This is because scheduled departure times were not available in the data used. Actual departure times were extracted from ASDI data.

3.2.3. Features Describing the Reroute Deviation

The features describing reroute deviation were as follows. Details of how they are calculated and the reason for their use are described below.

7. <u>*No. Downstream Sectors*</u>: The number of downstream sectors between the reroute start point and return capture fix;
8. <u>*Direct Route*</u>: Whether the reroute is direct to the return capture fix, or via one or more auxiliary waypoints.

Reroute deviation distance was identified as important in designing reroutes by Taylor and Wanke (2009, 2012). While lateral deviation from the current route was used in those studies, in this paper we quantified the route deviation by the number of sectors traversed by the reroute, prior to returning to the current flight plan at the return capture fix (*No. Downstream Sectors*). This represents how many sectors are likely to be impacted by the reroute, and therefore the number of controllers impacted by the route change, which could impact controller decision making on the reroute. The difference in route distance or duration is likely to be of greater importance to the pilot and airline decision making than the controller decision making. For this reason, the difference in route distance and duration were not included as features in this paper. The nature of the reroute deviation was also quantified by including a feature indicating whether or not the reroute requested is a direct route to the return capture fix, or includes one or more auxiliary waypoints (*Direct Route*). This feature was included because comparisons of historical reroute count by segment indicate that direct routes saw higher historical usage than routes via auxiliary waypoints. This suggests that whether a route is direct or via an auxiliary waypoint could impact its operational acceptance.

3.2.4. Features Describing the Reroute Start Point

Two features were included describing the aircraft location in the reroute start sector. Details of how they are calculated and the reason for their use are described below.

9. <u>*Time to Exit RSS*</u>: The time between when the reroute advisory was accepted by the airline dispatcher, and when the flight exited the reroute start sector;
10. <u>*Dist. to Exit RSS*</u>: The distance between the reroute start point and the reroute start sector boundary, along the requested reroute.

As described above, one cause for the rejection of reroute requests was that aircraft were close to being handed off to another sector (Idris et al., 2017). This suggests that the location of an aircraft in the reroute start sector when the reroute is requested may be relevant to its acceptance or rejection.

A schematic of reroute events is shown in Figure 4. Two features were included to describe the aircraft location in the reroute start sector because of limitations in the data used. The feature of most interest is the time between when the pilot requested the reroute from the air traffic controller (event 5 in Figure 4), and when the flight exited the reroute start sector (event 8 in Figure 4), because this would give an indication of how long the air traffic controller would have to handle the request. However, for this paper, no data were available on when the pilot requested the reroute from the air traffic controller. An alternative time was therefore used instead. The only times that were available were the time when the airline dispatcher accepted the reroute request (event 2 in Figure 4), before communicating it to the pilot, and the predicted time when the aircraft was to exit the reroute start sector, on the reroute (event 8 in Figure 4). The difference between these times was calculated, and included as a feature (*Time to Exit MSS*). However, the location of the reroute start point is set by DWR so as to be far enough downstream to allow for communication between the dispatcher and pilot, communication between the pilot and air traffic controller, and air traffic controller decision making. The location of the reroute start point can therefore be considered to be a limit on when the air traffic controller is expected to be ready to implement a Center route amendment. The distance between the reroute start point and sector boundary, along the requested reroute, is therefore also calculated, and included as a feature (*Dist. to Exit MSS*). In the case of observed Center route amendments, the reroute start point is classed as the location at which the Center route amendment departs from the current flight plan. Note that the dispatcher can accept a reroute advisory when the reroute start point is in a sector downstream of the current sector. In this case, feature 9 is still calculated relative to the time to exit the reroute start sector, and not the current sector. This is to account for the fact that, if the air traffic controller of the current sector were to reject the request because the aircraft is too close to hand-off, the pilot still has the opportunity to make the request of the air traffic controller in the following sector, for which this is less likely to be a problem.

Place Figure 4 aboute here

3.3. Data Mining

Identification of the best model for predicting the operational acceptability of reroute requests requires a number of development steps. These include feature selection, model selection, and parameter selection, in which the subset of features, the data mining algorithm, and parameters required by the data mining algorithm, are identified sequentially to maximize the chosen performance metric, as shown in Figure 5, before model validation. In this section, the chosen performance metric is described, followed by detailed descriptions of the approaches used for feature, model, and parameter selection. Model validation is described in Section 3-3.4.

Place Figure 5 about here

Feature, model, and parameter selection were solved in an iterative loop, ensuring that the parameters identified in parameter selection were also used in model selection, and that the parameters and model identified in parameter and model selection were also used in feature selection. Only one iteration was needed before the identified feature set, model and parameters converged.

3.3.1. Performance Metric

A number of performance metrics exist for the evaluation of data mining algorithms. *Accuracy*, which measures the fraction of correct predictions from all predictions made, is the most intuitive, but can be misleading when datasets are imbalanced. An alternative metric is *F-Score* (also called F_1 *Score* or *F-measure*), which is the harmonic mean of precision and recall, calculated as follows:

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{1}$$

Here *precision* refers to the number of predictions correctly labeled by the model as belonging to the positive class divided by the total number of predictions labeled by the model as belonging to the positive class. *Recall* refers to the number of predictions correctly labeled by the model as belonging to the positive class divided by the total number of predictions that actually belong to the positive class in the data (also called *Sensitivity* or *True Positive Rate*). F-Score varies from 0 to 1, with 1 being best.

One issue with both accuracy and F-Score as metrics for model performance is that they require setting of a discrimination threshold. Most binary classifiers output the probability of an observation falling within one of two classes. In order to classify the observation, this probability must be compared to a discrimination threshold. If the output probability is greater than the threshold, the observation is classed as positive. Otherwise, it is classed as negative. The discrimination threshold can be set to 0.5, giving equal importance to both classes. However, if one class is considered less desirable than the other, the threshold can be increased or decreased so as to decrease the likelihood of a false prediction in that class. If costs are known for misclassification in each class, the optimal threshold can be set accordingly. However, in many cases such costs are not known, making the setting of the threshold difficult. In the prediction of reroute operational acceptability, misclassification of an unacceptable reroute advisory that would have been rejected by ATC as acceptable, could lead to the reroute being requested by the airline, and consequent air traffic controller workload to reject or modify the proposed route. Conversely, misclassification of an acceptable reroute advisory as unacceptable could lead to no reroute being requested by the airline, or a less efficient reroute being requested, because the predictor would incorrectly predict that the acceptable advisory would be rejected by ATC. This could lead to a loss in efficiency benefits. Because it is difficult to quantify either of these costs, and it is not obvious which dominates, a discrimination threshold of 0.5 is applied in this paper.

Additional performance metrics can be calculated by varying the discrimination threshold and plotting the resulting performance metrics. The Receiver Operating Characteristic (ROC) curve plots true positive rate against *False Positive Rate* (the number of predictions incorrectly labeled by the model as belonging to the positive class divided by the total number of predictions that actually belong to the negative class in the data). The area under the ROC curve provides a metric of model performance ranging from 0 to 1 (1 being best), which is not a function of the chosen discrimination threshold. When using normalized units, this area (typically referred to as Area Under the Curve, *AUC*) indicates the probability that a classifier will rank a randomly chosen positive prediction higher than a randomly chosen negative one (assuming positive ranks higher than negative) (Fawcett, 2006). AUC varies from 0 to 1, with 1 being best.

For the purposes of this paper, model development is based on maximizing F-Score, with the discrimination threshold set to 0.5. Accuracy and AUC are also quoted, and ROC curves are plotted. Other metrics quoted include *Misclassification Error*, which measures the fraction of incorrect predictions from all predictions made (i.e., 1 – *Accuracy*); *True Negative Rate*, which refers to the number of predictions correctly labeled by the model as belonging to the negative class divided by the total number of predictions that actually belong to the negative class (also called *Specificity*); and precision, described above.

3.3.2. Feature, model, and parameter Selection

Feature selection involves identifying which combination of features provide the best performance. Any well-performing algorithm can be used, such as a random forest or support vector machine. Alternatively, feature selection can be run iteratively with model and parameter selection, to ensure that the same algorithm and parameters are used for feature selection as are identified in the model and parameter selection. In this paper, a random forest (Breiman, 2001) was used. A random forest is an ensemble learning method for classification that constructs a multitude of decision trees during training and outputs the class that is the majority vote of the classes of the individual trees (Hastie et al., 2008). The algorithm used is as follows (Hastie et al., 2008):

1. For b = 1,2,…B, where B is the chosen number of trees in the random forest:
    a. Randomly draw a sample of size N from the training data (called a bootstrap sample).
    b. Grow a random forest tree T_b from the bootstrap sample by recursively repeating the following steps for each terminal node of the tree, until the maximum tree depth or minimum node size is reached. In this paper maximum tree depth is not constrained, while minimum node size is set to 1.
        i. Select m features at random from the complete set of M features, where $m = \lfloor\sqrt{M}\rfloor$.
        ii. Pick the feature that best splits the sample data.
        iii. Split the node into two daughter nodes.
2. Output the ensemble of trees $\{T_b\}_1^B$.

To make a prediction for a new observation, $\boldsymbol{x}$, let $\hat{C}_b(\boldsymbol{x})$ be the class prediction of the bth random-forest tree:

$$\hat{C}^B(\boldsymbol{x}) = majority\ vote\left\{\hat{C}_b(\boldsymbol{x})\right\}_1^B \tag{2}$$

The number of trees, B, was set to 40 for feature selection, as it was found to maximize model performance in parameter selection, described below.

K-fold cross-validation, a form of 'leave-*p*-out' cross validation, was used to estimate performance with different combinations of features. In *K*-fold cross-validation, the original set of observations is randomly partitioned into K equal sized subsets, or 'folds'. Of the K folds, a single fold is retained as the validation data for testing the model, and the remaining $K - 1$ folds are used as training data. The cross-validation process is then repeated K times, with each of the K folds used exactly once as the validation data. This means that it matters less how the data are divided. Every data point is used in a test set exactly once, and in a training set K-1 times. The larger the value of K chosen, the lower the variance of the resulting estimate. The *K*-fold cross validation algorithm used is therefore as follows:

1. Divide the data in K equal parts.
2. For each k = 1,2,…K,
    a. Fit the model to the other K-1 parts, giving model $\hat{\beta}^{-k}$.
    b. Compute its error in predicting the kth part:

$$E_k = \sum_{i=1}^{p_k} 1\left(y_i \neq \hat{\beta}^{-k}(\boldsymbol{x}_i)\right) \tag{3}$$

    where p_k is the number of observations in the kth part; y_i is the observed class for observation i, and $\boldsymbol{x}_i$ the observed features for observation i.

3. Calculate the cross-validation error as:

$$E_{CV} = \frac{1}{K}\sum_{k=1}^{K} E_k \tag{4}$$

For the results presented in this paper, a greedy forward search algorithm (Deng, 1998) was used for feature selection, reducing the number of feature combinations for which performance is calculated, because only a subset of all possible combinations of features is evaluated. The greedy forward search procedure begins with an empty feature set and, in each step, adds the feature that results in the largest improvement in performance to the current feature set, building up the feature set one feature at a time, until all features have been added. The feature set with highest performance of all those tested was then adopted. In this paper, 10-fold cross validation (i.e., K=10) was used for feature set evaluation. This value was found to produce good model performance in a reasonable computational time.

Model selection involves identifying which data mining algorithm provides the best performance. The model is a two-class classifier, with ATC rejection of a reroute request defined as the 'positive' class, and ATC acceptance of a reroute request defined as the 'negative' class. The positive class is typically labeled as the class for which it is most important to correctly predict results. The importance of ATC acceptance vs. rejection of reroute requests depends on the context in which the question is being asked. In the context of this paper, ATC rejection became the focus because the concept explores increases in automation in part to avoid unnecessary air traffic controller workload, which can be achieved by limiting the number of routes requested that are rejected by controllers. Hence correctly classifying ATC rejection is considered important, and ATC rejection was classed as the ‘positive’ outcome.

A number of algorithms were applied to train the model, using the *R* statistical computing environment:

(1) A logistic regression, which is also known in the literature as logit regression, maximum-entropy classification or log-linear classification. In logistic regression the probabilities describing the possible outcomes of a single trial are modeled using a logistic function. In this paper, logistic regression was applied using the generalized linear model package *glm* in *R*, with the model link function set to *logit*.

(2) A support vector machine, which constructs a hyper-plane to separate different classes of data. A sigmoid kernel (also called a hyperbolic tangent or multilayer perceptron kernel) was found to be most effective for the data in this paper, and takes the form:

$$k(x, y) = tanh(\alpha \cdot x^T y + c) \quad (5)$$

The support vector machine was applied using the *svm* package in *R*, with the kernel set to *sigmoid*. The slope α was set to the inverse of the data dimension, and the constant intercept *c* was set to 0. These values are defaults in the *R* package used, and were found to produce good model performance, so were maintained.

(3) A single decision tree, which uses a decision tree structure to classify data. In these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Observations are classified by making decisions based on feature values through a series of decision points, resulting in a final classification at the final leaf nodes of the decision tree. In this paper, a decision tree was grown using the recursive partitioning tree package *rpart* in *R*. Both the minimum number of observations that must exist in a node in order for a split to be attempted, and the minimum number of observations in any terminal ‘leaf’ node, were set to 0. The maximum depth of any node in the final tree was set to 5. These values are defaults in the *R* package used, and were found to produce good model performance, so were maintained.

and two ensemble methods –

(4) A random forest, as described above. In this paper, the random forest was applied using the *randomForest* package in *R*. The number of trees, *B*, was set to 40, as it was found to maximize model performance in parameter selection, described below.

(5) Adaptive boosting (AdaBoost) (Freund and Schapire, 1997), which combines the output of the other learning algorithms which are typically computationally simple and have relatively poor performance (called 'weak learners') into a weighted sum that represents the final output of the boosted classifier. It is adaptive in the sense that in subsequent iterations, weak learners are modified in favor of those instances misclassified in previous iterations. In this paper, AdaBoost was applied using the *boosting* function from the *adabag* package in *R*, using classification trees as the weak learners. The number of trees used was 40, as identified for the random forest. The AdaBoost.M1 algorithm is used, with the weight updating coefficient α set to $1/2\, ln((1 - err)/err)$, which was found to produce the best model performance of the alternatives available in the *R* package used.

Again, *K*-fold cross validation, with *K*=10, was used to estimate performance using each algorithm.

Imbalances in the number of observations in each class were also considered. For the full dataset, 40% of routes were classed as rejected by ATC, while 60% of routes were classed as accepted. This does not represent a large data imbalance. The impact of class balancing on performance was still explored, using a variety of over- and under-sampling approaches (Seiffert et al. 2010; Menardi and Torelli, 2014; Lunardon et al., 2016; Chawla et al., 2002) but none were found to improve performance, so are not presented in this paper.

Finally, parameter selection involves identifying key parameters in the chosen data mining algorithm that provide the best performance. For example, for a random forest or AdaBoost, the number of trees or weak learners was varied between 20 and 100. In this paper, *K*-fold cross validation, with *K*=10, was used to estimate performance across a range of each of the key parameters of the chosen algorithm.

3.4. Validation

The purpose of validation is to demonstrate that, for the data available, the approach described above produces a classifier that effectively predicts whether a new reroute would be operationally acceptable to ATC or not, in the form of a probability of acceptance, impacting whether the airline would request the reroute or not. Validation is typically done by dividing the original dataset into two sets: one ‘development’ set for developing the model, and one ‘evaluation’ or ‘test’ set, for testing the developed model on different data. This ensures that the data used to train the model and compare model parameters are not used to evaluate the model performance. Once the approach has been validated, a final model can be trained on all the available data.

If the available dataset is small, it is desirable to calculate performance metrics based on all the data, which can be done using K-fold cross validation. Because K-fold cross validation is already used in the model development – to select features, model and parameters – 'nested' K-fold cross validation is needed for validation. Nest K-fold cross validation is illustrated in Figure 6. It means that the data are randomly split into K_1 folds (① in Figure 6), which are used, in different combinations by switching out different folds, to compile K_1 different evaluation sets (made up of one fold each) and K_1 different development sets (made up of K_1-1 folds each) (② in Figure 6). Each development set is then randomly split into K_2 folds (K_1 may differ from k_2) (③ in Figure 6), which are each used for feature, model, and parameter selection using K_2-cross validation. This means using K_2 folds, in different combinations by switching out different folds, to compile K_2 different test sets (made up of one fold each) and K_2 different training sets (made up of K_2-1 folds each) (④ in Figure 6). With the feature set, model, and parameters selected, the resulting models developed using each of the K_1 development sets are then re-trained on their full development set, generating new model coefficients. These K_1 new models are then tested on their respective K_1 evaluation sets. The model performance across all K_1 evaluation sets, which represents the whole dataset, can then be calculated. Note, however, that the feature, model, and parameter selection across all K_1 models may not produce identical results. Results applying this approach are presented in Section 4-4.4 for the prediction of route operational acceptability to ATC.

Place Figure 6 about here

Once suitable validation results have been generated, the model development can be repeated on the full dataset, using K-fold cross validation across the full dataset for feature, model, and parameter selection. The features, model and parameters selected for the prediction of route operational acceptability to ATC are presented in Section 4-4.1, 4.2, and 4.3 respectively. Once features, model and parameters are selected, the final predictor can be trained on all the available data, so as to most accurately represent all the data available.

4. Results

Feature, model, and parameter selection results, applied to the full dataset, are presented below, followed by model validation results using nested cross-validation. Finally, results applying a single class classifier are also presented for comparison.

4.1. Feature Selection

Table 2 and Figure 7 present feature selection results calculated for the full dataset with the positive class defined for routes that are rejected or modified by ATC, and the negative class defined for routes that are accepted by ATC. The maximum number of observations included in all steps of the feature selection is 544. Of these 544, 40% are positive and 60% negative. Because some observations have missing data for some features, each step of the feature selection has a different number of observations, ranging from 317 to 544. The first column of Table 2 shows a list of all features considered. The second column shows the F-Scores for a random forest trained on each of these features, separately. The random forest is trained with 40 decision trees, no constraint on the tree depth, and a minimum terminal node size of 1. The number of features randomly sampled as candidates at each split, with replacement, is the floor of the square root of the total number of features (i.e., $\lfloor\sqrt{p}\rfloor$, where p is the total number of features). The highest F-Score in column 2 (0.674) is for the feature *Hist. Reroute Count by Segment*. This indicates that historical usage is the strongest single contributor to reroute operational acceptance tested. This is supported by a direct comparison of the *Hist. Reroute Count by Segment* of each observed route, and its acceptance or rejection by ATC, which indicates that 97% of Center route amendments made in response to the dispatcher-requested DWR reroute advisories in the 2014 DWR trial data also appeared in the historical data from June to August 2015 (i.e., had an *Hist. Reroute Count by Segment* greater than 0). This indicates that if a route is not observed historically, it is unlikely to be accepted by ATC. However, the opposite is not true. Requesting a reroute that is observed in the historical data does not guarantee ATC acceptance. In fact, 69% of dispatcher-requested DWR reroute advisories that were rejected by ATC in the 2014 DWR trial data also appeared in the data from June to August 2015. This indicates that there are also other reasons for an air traffic controller's acceptance or rejection of a reroute request, hence the inclusion of features other than those that indicate historical usage.

Place Table 2 about here
Place Figure 7 about here

Using the greedy forward search approach to feature selection, *Hist. Reroute Count by Segment* was therefore retained as a feature, and the model performance re-evaluated with each remaining feature added to the feature set separately. Results are shown in column 3 of Table 2. Including *Time to Exit MSS* in the feature set with *Hist. Reroute Count by Segment* yields the highest F-Score for 2 features (0.719), indicating that the location of the aircraft in the reroute start sector when the reroute is requested is relevant to the reroute operational acceptability, as suggested by Idris et al. (2017). Forward search is used to add all remaining features, with results shown in Table 2.

The feature set yielding the highest F-Score (0.815) contains 7 features, as follows:

1. Hist. Reroute Count by Segment;
2. Time to Exit RSS;
3. No. Downstream Sectors;
4. Direct Route;
5. Dist. to Exit RSS;
6. RSS D/C Ratio; and
7. RSS Over Capacity.

The inclusion of any more features than those listed above did not improve the model performance. Note that this feature set includes features describing historical usage (1), location of the aircraft in the reroute start sector (2 and 5), reroute deviation (3 and 4), and congestion (6 and 7). Significantly, the selected features describing congestion only describe congestion in the reroute start sector, and not in downstream sectors. This means that downstream sector congestion did not manifest in the data, suggesting that downstream congestion may not be a significant factor in the operational acceptability of reroute requests initiated by pilots. This may be due to air traffic controllers having limited information on downstream sector congestion when making these decisions. However, it may also be that the features used to describe downstream congestion are not representative of how air traffic controllers evaluate downstream congestion. It is noted that this analysis includes only reroutes initiated by pilots, and does include reroutes initiated by ATC, and particularly by Traffic Managers, who, according to SMEs, do consider downstream sector congestion information in their decision making.

These results provide a validation that most of the features identified in the literature do in fact predict route acceptability. However, they indicate that not all features identified in the literature – particularly only sector congestion in the maneuver start sector, and not in all sectors impacted by the reroute – contribute to predicting route acceptability.

Figure 7 shows the ROC curve for the random forest, for each of the feature sets identified to maximize F-Score in Table 2, from 1 to 7 features. There is a clear progression as features are added, shifting the ROC curve towards the upper left (a perfect model would be in the upper left corner).

With the feature set selected, the data-mining algorithm that maximizes F-Score can be selected.

4.2. Model Selection

Model selection results are presented in Table 3 and Figure 8. The features used are the 7 features listed in Section 4-4.1 above. The number of observations included is 317, of which 48% are positive and 52% negative. Details of the models used are as described in Section 4-4.1. The best performing model in terms of F-Score is the random forest (0.815), with AdaBoost the second best (0.766). The random forest also performs best under all other metrics. These results are generally expected as ensemble learning techniques such as a random forest and AdaBoost typically outperform other machine learning algorithms because the group of classifiers trained performs more accurately than any single classifier, utilizing the strengths of the individual group of classifiers while at the same time circumventing the weaknesses of the individual classifier (Kotsiantis and Pintelas, 2004). However, this is not always the case (e.g., Evans and Lee, 2016). Beyond this observation, it is difficult to identify explicitly why the random forest performs better than, e.g., AdaBoost, for the particular data studied in this paper.

Place Table 3 about here
Place Figure 8 about here

Figure 8 shows the ROC curve for each of the models, where it is clear that the random forest provides the best performance across a range of discrimination thresholds.

4.3. Parameter Selection

In a random forest, a key parameter that must be set is the number of decision trees used. It is important that the number of trees is large to ensure that all observations are used, preferably more than once, and that all features are used, again preferably more than once. If the number of trees is too small the predictive power of the random forest decreases. Because the problem described in this paper has a small number of features (7), and the number of observations is also not large (317, given the feature set used), a very large number of trees is unnecessary. 10-fold cross-validation is used to compare the performance of the model with a range of values for this parameter, from 20 to 100, as shown in Table 4 and Figure 9. In Table 4, the number of trees that maximizes the model F-Score is 40. This value also maximizes all other metrics. The ROC curve for the different parameter values in Figure 9 shows that the results vary very little with number of trees.

Place Table 4 about here
Place Figure 9 about here

Three other key parameters that require setting in random forests are the maximum tree depth for the decision trees, the minimum terminal node size, and the number of features randomly sampled as candidates at each split. Because of the small number of both observations and features, however, no constraint is imposed on the tree depth in this paper, and the minimum terminal node size is fixed at 1. The number of features randomly sampled as candidates at each split, with replacement, is also fixed, at $\lfloor\sqrt{p}\rfloor$, where p is the total number of features. Given p=7, 2 features are randomly sampled at each split.

In summary, given the results presented in Sections 4-4.1, 4.2 and 4.3, the model development suggests that the best performing model for the prediction of ATC operational acceptability of reroute requests in ZFW and its adjacent Centers is a random forest with 40 trees, trained on a feature set including *Hist. Reroute Count by Segment*, *Time to Exit MSS*, *No. of Downstream Sectors*, *Direct Route*, *Dist. to Exit MSS*, *MSS D/C Ratio*, and *MSS Over Capacity*.

4.4. Model Validation

The model results were validated numerically using nested 10-fold cross validation. Results are described in the section below. This is followed by a qualitative discussion of the model results with an SME, in the next section.

4.4.1. Numerical Validation

Table 5 and Figure 10 present the results of the nested 10-fold cross validation, including feature, model, and parameter selection as described in Section 3-3.4. The validation dataset is reasonably well balanced (out of 317 observations, 40% are positive, and 60% negative).

Place Table 5 about here
Place Figure 10 about here

The model performance is reasonable, with an F-Score of 0.767. Other performance metrics such as accuracy, precision and AUC are also reasonable. Accuracy particularly can be compared directly to the percentage of the majority class (60%). This can be considered to be a baseline accuracy which could be achieved with a model that simply specifies all outputs to be the majority class (i.e., ATC accepted), without consideration of any feature values. In comparison, the model developed has an accuracy of 74.4%, which is 14.4% better than this baseline. The true positive rate in Table 5 is particularly high (0.875), indicating that the approach is effective at accurately predicting whether a reroute is likely to be rejected or modified by ATC. As stated above, this is of particular importance in the application considered, because misclassification of route rejection by ATC could lead to increased air traffic controller workload. The true negative rate is lower (0.624), indicating that the approach is less effective at accurately predicting whether a reroute is likely to be accepted by ATC.

Choice of a different discrimination threshold could also alter the tradeoff between true positive rate and true negative rate, moving along the ROC curve shown in Figure 10 (an increase in discrimination threshold would lead to a shift down and left along the curve, while a decrease in discrimination threshold would lead to a shift up and right along the curve). The area under the ROC curve (AUC), however, is high (0.814), indicating good model performance across a range of discrimination thresholds.

4.4.2. Qualitative Discussion with SME

In order to gain operational insights into why the identified features might better predict route acceptability than those rejected, the results were discussed with a retired traffic manager with extensive experience in ZFW and the adjacent Center airspace, who has also contributed significantly to NASA research in traffic flow management (the questionnaire used can be found in Appendix A). The retired traffic manager suggested that the most important factors in determining the operational acceptability of a requested reroute in ZFW while airborne were (1) characteristics of the underlying airspace, such as location of airport arrival and departure flows; (2) the air traffic controller's individual workload levels; and (3) time to hand-off to the next sector. The importance of characteristics of the underlying airspace (1) is consistent with identifying *Hist. Reroute Count by Segment* as a key feature in the feature selection results described in Section 4-4.1. If characteristics of the underlying airspace were accounted for by the feature set explicitly, instead of historical usage, the set of acceptable reroutes may expand, with some routes with low historical usage, but which meet the required characteristics of the underlying airspace, being included. However, it would not be clear if these routes were included because they are genuinely acceptable to air traffic controllers, or because the set of characteristics of the underlying airspace accounted for in the feature set was incomplete. Air traffic controller workload (2) would be impacted by congestion levels in the reroute start sector, which is consistent with identifying *MSS D/C Ratio* and *MSS Over Capacity* as key features in Section 4-4.1. Time to hand-off (3) would be impacted by the location of the aircraft in the reroute start sector, which is consistent with identifying *Time to Exit MSS* and *Dist. to Exit MSS* as key features in Section 4-4.1.

Reroute deviation was considered by the retired traffic manager to be of lower importance. Two of the features identified in Section 4-4.1, *No. Downstream Sectors* and *Direct Route*, describe reroute deviation. *No. Downstream Sectors* may be capturing the need for flight plan waypoints at Center boundary crossings, or similar geographical airspace related requirements. It is more likely that reroutes traversing a high number of sectors would not meet this requirement, requiring reroute modification. *Direct Route* may be capturing the historical usage of the reroute, since direct routes in this dataset saw higher historical usage than routes via an auxiliary waypoint. This is because auxiliary waypoints are primarily used to avoid weather, and weather is highly variable. In the version of DWR fielded, many of the routes used to avoid different weather cells via auxiliary waypoints are unique, not appearing in the historical data at all.

The retired traffic manager also confirmed that air traffic controllers do not have information about downstream congestion, which would make features describing downstream congestion, such as *No. Sectors Over* and *Max D/C Ratio*, of lower relevance. This is consistent with the feature selection results in Section 4-4.1, where these features were not identified as being important.

4.5. Single Class Classifier

It is not immediately obvious how the model developed in this paper can be extended to other regions and applications, because of the limited data for rejected reroute requests on which to train a two-class classifier. Alternative approaches exist, such as single-class classification, for which only a single class of observations are required in the form of Center route amendments. These are widely available. The performance of single class classifiers is, however, significantly poorer than the two-class classifiers developed in this paper. Developing a single-class classifier for the problem considered in Sections 4-4.1 to 4-4.4, using a Support Vector Machine (SVM) with sigmoid kernel (no method could be identified to use a random forest for single class classification), yields a very low F-Score of 0.314 when model results are compared to actual route acceptance. This is significantly lower than the 0.815 model development F-Score for the two-class classifier using a random forest (Table 3), or even the 0.658 F-Score for the two-class classifier using a comparable SVM. The rate at which the single-class classifier predicts reroute acceptance by ATC is actually high – 0.879 – but the prediction of reroute rejection by ATC – at a rate of only 0.211 – leads to the low F-Score. This is expected since it is reroute rejection for which no training data are used. This is a problem if it is the prediction of route rejection by ATC that is considered of highest importance, as discussed in Section 3-3.3. However, in applications for which the prediction of route acceptance is most important, this could provide a viable alternative. Ultimately, most beneficial to the building of an effective predictor of route operational acceptability would be better data for the two-class classification problem, both in terms of geographical coverage (beyond ZFW and adjacent Centers) and in terms of quality. This latter point is of particular importance. The data from the 2014 DWR trial does not include information describing exactly what route was requested of the air traffic controller, or when the route was requested. While proxies for this exist in the data used in this paper, discrepancies between these and what was really seen by the air traffic controller could erode the performance of the model, and impact the effectiveness of the approach when extended to other regions. Processing of ATC voice data could provide such higher quality information, but this is beyond the scope of this paper.

5. Conclusions

In this paper, the authors developed a predictor of operational acceptability for reroute requests in ZFW and its adjacent Center airspace. In order to confirm whether or not the operational acceptability of reroute requests is predictable or not. The predictor was developed using data mining techniques applied to flight plan amendment data and data from a trial of the NASA developed DWR tool at American Airlines in 2014. Routes implemented in the form of a Center route amendment within 30 minutes of the DWR reroute advisory being accepted by the airline dispatcher were classed as operationally acceptable to ATC. Routes are classed as operationally unacceptable to ATC if they are either modified by the air traffic controller before being implemented in the form of a Center route amendment, or do not result in any Center route amendment being implemented at all. Features used in the development of the predictor were taken from the literature and include data on historical usage of the advised route, congestion levels in the reroute start sector and downstream, route deviation from the current flight plan, and the location of the reroute start point in the reroute start sector. Features, models, and key model parameters were selected using 10-fold cross validation, while nested cross validation was used to validate the model performance, giving a reasonable F-Score of 0.77. The model found to provide best performance was a random forest with 40 trees. Importantly, the rate at which the classifier predicts whether a reroute is likely to be rejected or modified by ATC is high: 0.88. This is of particular importance in the application considered, because misclassification of route rejection by ATC could lead to increased air traffic controller workload. Features identified as relevant include (1) the historical usage of the route change, (2) the proximity of the reroute start point to the boundaries of the airspace sector containing the reroute start, (3) the length of the reroute, and (4) the congestion in the reroute start sector.

These results provide a data-based validation that most of the features identified by SMEs in the literature, do in fact predict route acceptability, with the exception of sector congestion downstream of the reroute start sector. Through discussions with a retired traffic manager, operational insights were gained into why the majority of the features identified are relevant to the operational acceptability of requested airborne reroutes in ZFW and its adjacent Centers.

The application of data mining to the DWR trial data presented in this paper provides significant results indicating that the operational acceptance of reroute requests is indeed predictable, and that, with suitable data, models can be trained to predict the operational acceptability of reroute requests. In future work, the operational acceptability of these routes can be traded off with other reroute benefits, such as flight time savings, as demonstrated by Evans et al. (2016), which trades off historical route usage and delay savings from the DWR tool. Models predicting the operational acceptance of flight reroutes may ultimately be used to inform route selection by decision support tools, contributing to the development of increasingly autonomous systems that are capable of routing aircraft with less human input than is currently the case.

It is also important to highlight that one of the advantages of increasingly autonomous systems is that they can enable efficiency improvements that would not be possible without automation. This could include the enabling of control actions that would not be considered operationally acceptable to air traffic controllers today, but which are safe. One example would be to allow reroutes through regions of airspace, like arrival corridors, that are generally protected by air traffic controllers, if it can be confirmed that this would be safe because no arrival traffic is present at that time. In today's operations, air traffic controllers are unlikely to do this because they have limited information about congestion levels in other sectors. The development of a tool solely using data mining based on past operations would miss this opportunity. Data mining should therefore be combined with other approaches, such as optimization, to take full advantage of opportunities for efficiency improvements.

Future work includes improving the quality of the features mined, expanding the training data on which the predictor is developed, and extending the work to other regions, beyond ZFW and its adjacent Centers. This would allow integration of the predictor into tools that generate reroute advisories.

Acknowledgements

This research was funded by NASA under contract number NNA16BD14C. The authors would like to acknowledge helpful comments from Heather Arneson, Deepak Kulkarni and Yao Wang of NASA Ames Research Center, and feedback on the identified features from retired ZFW traffic manager Danny Vincent.

References

Balakrishna, P., Ganesan, R. and Sherry, L., 2010. Accuracy of reinforcement learning algorithms for predicting aircraft taxi-out times: A case-study of Tampa Bay departures. *Transportation Research Part C: Emerging Technologies*, *18*(6), pp.950-962.

Ballin, M. and Wing, D., 2012, September. Traffic aware strategic aircrew requests (TASAR). In *12th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference and 14th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference* (p. 5623).

Bloem, M., Hattaway, D., and Bambos, N., 2012. Evaluation of algorithms for a miles-in-trail decision support tool," *Proc. of International Conference on Research in Air Transportation*.

Breiman, L., 2001. Random forests. *Machine learning*, *45*(1), pp.5-32.

Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, *16*, pp.321-357.

Deng, K., 1998. *Omega: On-line memory-based general purpose system classifier* (Doctoral dissertation, Carnegie Mellon University), CMU-RI-TR-98-33.

Edwards, T.E. and Lee, P.U., 2017. Towards Designing Graceful Degradation into Trajectory Based Operations: A Human-Machine System Integration Approach. In *17th AIAA Aviation Technology, Integration, and Operations Conference* (p. 4487).

Evans, A.D. and Lee, P.U., 2016. Analyzing Double Delays at Newark Liberty International Airport. In *16th AIAA Aviation Technology, Integration, and Operations Conference* (p. 3456).

Evans, A.D., Sridhar, B. and McNally, D., 2016. Improving Operational Acceptability of Dynamic Weather Routes Through Analysis of Commonly Use Routings, In *16th AIAA Aviation Technology, Integration, and Operations Conference* (p. 3600).

Fawcett, T., 2006. An introduction to ROC analysis. *Pattern recognition letters*, *27*(8), pp.861-874.

Federal Aviation Administration, 2014. 4D Trajectory Based Operations (4D TBO). Concept of operations CONOPS (version 1). *Federal Aviation Administration*, Washington, DC, USA.

Freund, Y. and Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, *55*(1), pp.119-139.

Gong, C. and McNally, D., 2015. Dynamic arrival routes: A trajectory-based weather avoidance system for merging arrivals and metering. In *15th AIAA Aviation Technology, Integration, and Operations Conference* (p. 3394).

Hastie, T., Tibshirani, R., and Friedman, J., 2008. *The Elements of Statistical Learning* 2nd ed. New York, NY, USA; Springer, ISBN 0-387-95284-5, pp. 587-588.

Hrastovec, M. and Solina, F., 2016. Prediction of aircraft performances based on data collected by air traffic control centers. *Transportation Research Part C: Emerging Technologies*, *73*, pp.167-182.

Idris H., Enea G., Burke K., and Wing D., 2017. Assessment of Air Traffic Controller Acceptability of Aircrew Route Change Requests", *Proc. of USA/Europe Air Traffic Management Research & Development Seminar*.

Kim, A. and Hansen, M., 2013. A framework for the assessment of collaborative en route resource allocation strategies. *Transportation Research Part C: Emerging Technologies*, *33*, pp.324-339.

Kotsiantis, S. and Pintelas, P., 2004. Combining bagging and boosting. *International Journal of Computational Intelligence*, *1*(4), pp.324-333.

Kuhn, K.D., 2016. A methodology for identifying similar days in air traffic flow management initiative planning. *Transportation Research Part C: Emerging Technologies*, *69*, pp.1-15.

Kulkarni, D., Wang, Y. and Sridhar, B., 2014. Analysis of airport ground delay program decisions using data mining techniques. In *14th AIAA Aviation Technology, Integration, and Operations Conference* (p. 2025).

Li, L., Hansman, R.J., Palacios, R. and Welsch, R., 2016. Anomaly detection via a Gaussian Mixture Model for flight operation and safety monitoring. *Transportation Research Part C: Emerging Technologies*, *64*, pp.45-57.

Lunardon, N., Menardi, G., and Torelli, N., 2016. ROSE: a Package for Binary Imbalanced Learning, *R Journal*, 6:82–92.

MathWorks, Inc., 2018. Mastering Machine Learning: A Step-By-Step Guide with MATLAB, *The MathWorks, Inc.*, Natick, MA, USA.

McNally, D., Sheth, K., Gong, C., Love, J., Lee, C.H., Sahlman, S. and Cheng, J., 2012, September. Dynamic weather routes: a weather avoidance system for near-term trajectory-based operations. In *28th International Congress of the Aeronautical Sciences*.

McNally, D., Sheth, K., Gong, C., Sterenchuk, M., Sahlman, S., Hinton, S., Lee, C.H. and Shih, F.T., 2015. Dynamic weather routes: two years of operational testing at American Airlines. *Air Traffic Control Quarterly*, *23*(1), pp.55-81.

Menardi, G. and Torelli, N., 2014. Training and assessing classification rules with imbalanced data, *Data Mining and Knowledge Discovery*, 28:92–122.

Pfeil, D.M. and Balakrishnan, H., 2012. Identification of robust terminal-area routes in convective weather. *Transportation Science*, *46*(1), pp.56-73.

Seiffert, C., Khoshgoftaar, T.M., Van Hulse, J. and Napolitano, A., 2010. RUSBoost: A hybrid approach to alleviating class imbalance. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *40*(1), pp.185-197.

SESAR Consortium, SESAR D3, 2007. The ATM Target Concept, Document No. DLM-0612-001-02-00a, *EUROCONTROL*, Brussels, Belgium.

Sheth, K., Clymer, A., Morando, A. and Shih, F.T., 2016. Analysis of Multiple Flight Common Route for Traffic Flow Management. In *16th AIAA Aviation Technology, Integration, and Operations Conference* (p. 4207).

Stewart, T., Askey, L. and Hokit, M., 2012, September. A concept for tactical reroute generation, evaluation and coordination. In *12th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference and 14th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference* (p. 5586).

Taylor, C. and Wanke, C., 2009. Dynamic generation of operationally acceptable reroutes. In *9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO) and Aircraft Noise and Emissions Reduction Symposium (ANERS)* (p. 7091).

Taylor, C. and Wanke, C., 2012. Dynamically generating operationally acceptable route alternatives using simulated annealing. *Air Traffic Control Quarterly*, *20*(1), pp.97-121.

Wang, Y., and Kulkarni, D., 2011. Modeling weather impact on Ground Delay Programs," *SAE Journal of Aerospace*, 4(2):1207–1215.

Zelinski, S., Xue, M., and Bassett P., 2016. Optimized Route Capability (ORC) Intelligent Offloading of Congested Arrival Routes. In *16th AIAA Aviation Technology, Integration, and Operations Conference* (p. 4357).

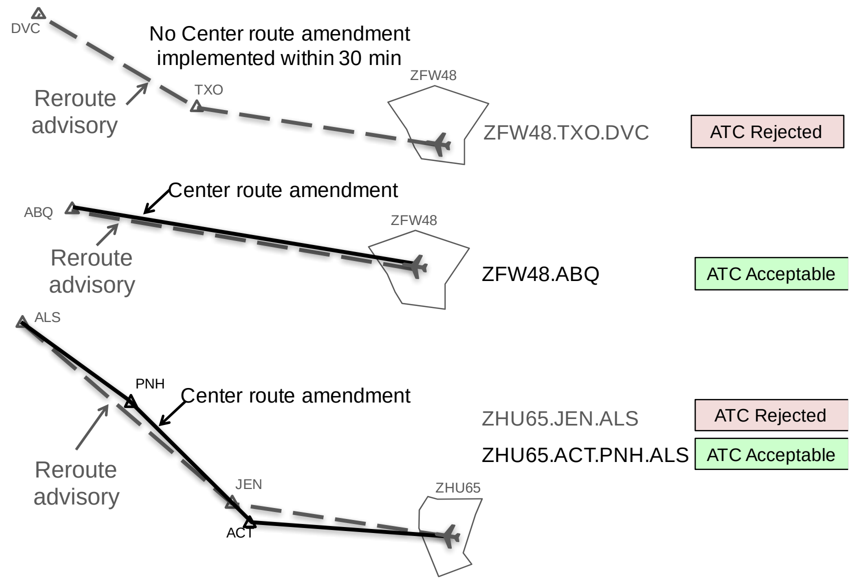


Figure 1. Approach to classification of routes as ATC accepted or rejected. The classified routes include DWR advisories and air traffic controller implemented Center route amendments.

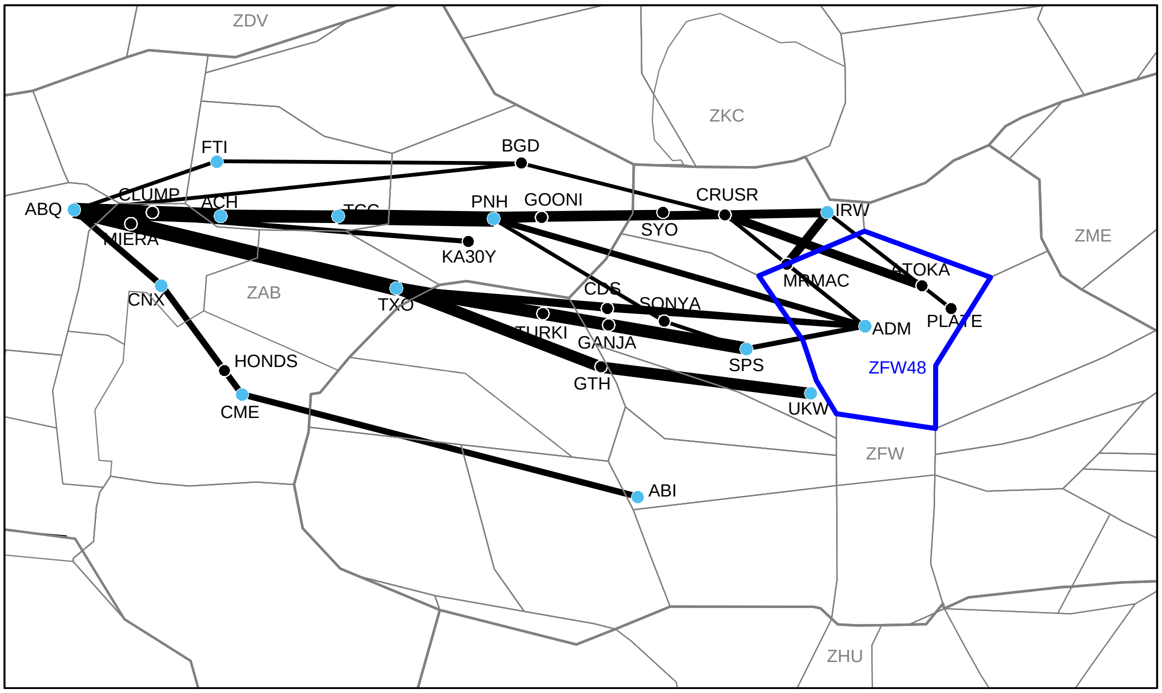
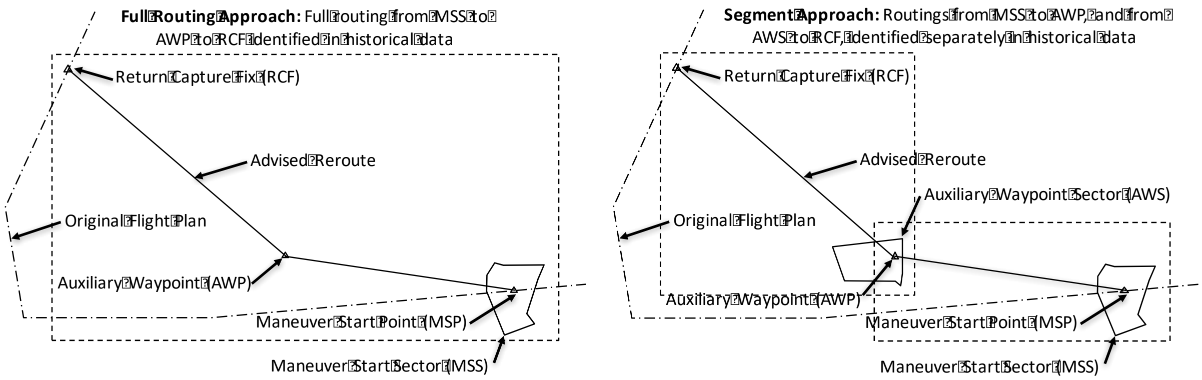


Figure 2. Common Routes from sector ZFW48 to waypoint ABQ (observed more than 10 times in April 2015). Routes are shown from the first fix after ZFW48 is entered to the final route fix, ABQ. Relative historical route counts are indicated by the thickness of the lines between fixes. High Power VORs are shown in blue. Center and high-altitude sector boundaries are shown in gray.



(a) (b)

Figure 3. Comparison of (a) full reroute approach and (b) segment approach to identifying reroute usage in historical data.

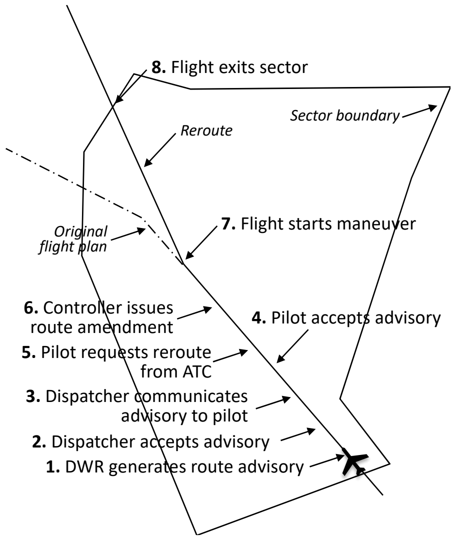


Figure 4. Reroute Event Schematic.

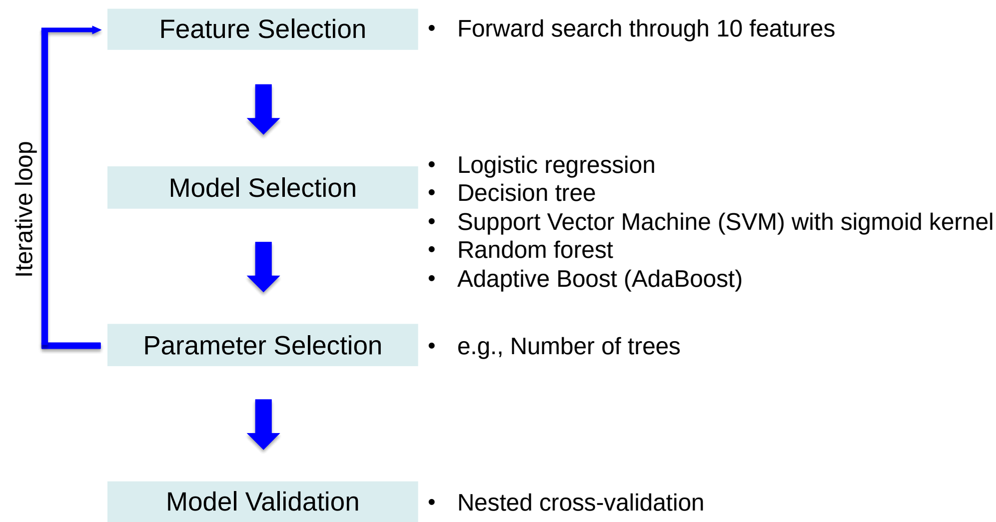


Figure 5. Methodology used for Data Mining.

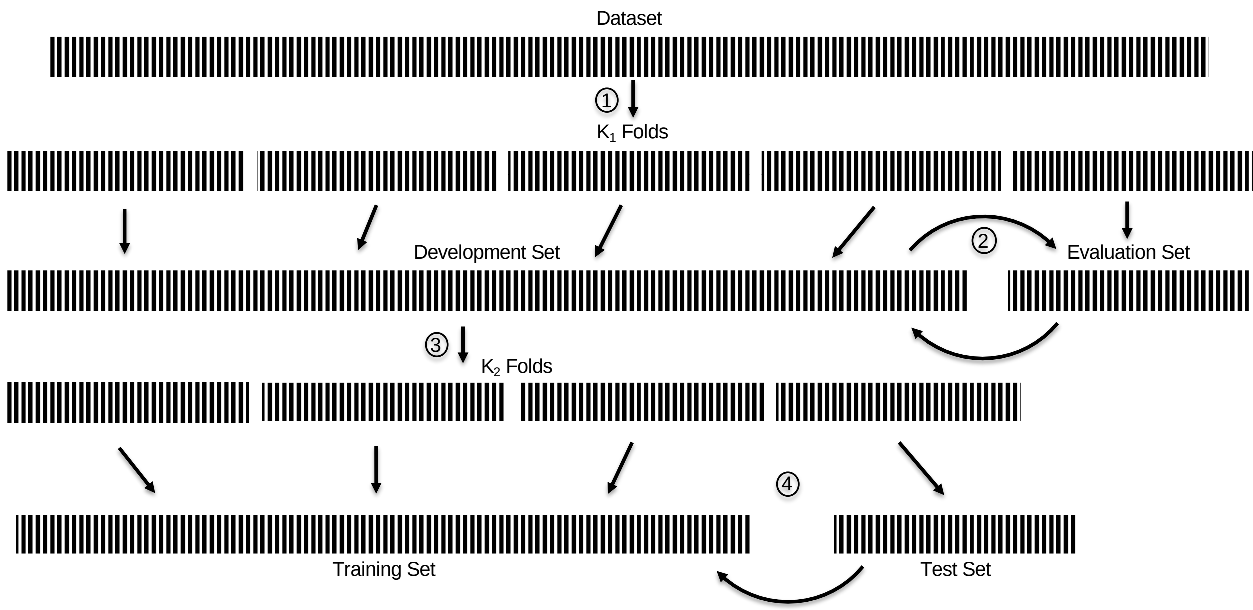


Figure 6. Nested *K*-fold cross validation.

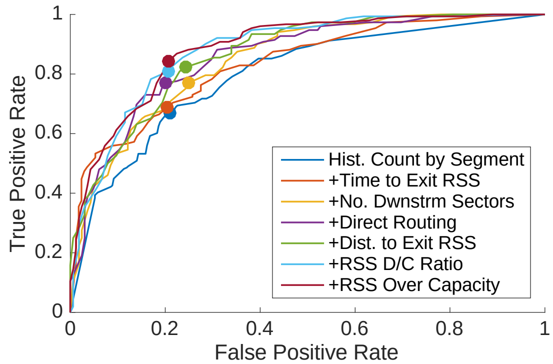


Figure 7. ROC curves for forward search feature selection, applying a random forest in the prediction of ATC operational acceptability of DWR reroute advisories from May to September 2014. Filled circles represent the points with 0.5 discrimination threshold presented in Table 2. (RSS: Reroute Start Sector; D/C: Demand over Capacity.)

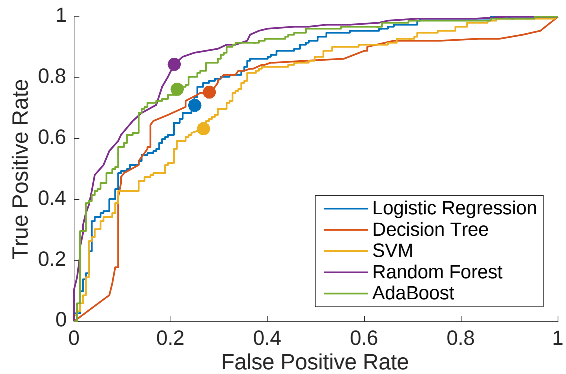


Figure 8. ROC curves for model selection, in the prediction of ATC operational acceptability of DWR reroute advisories from May to September 2014. Filled circles represent the points with 0.5 discrimination threshold presented in Table 3.

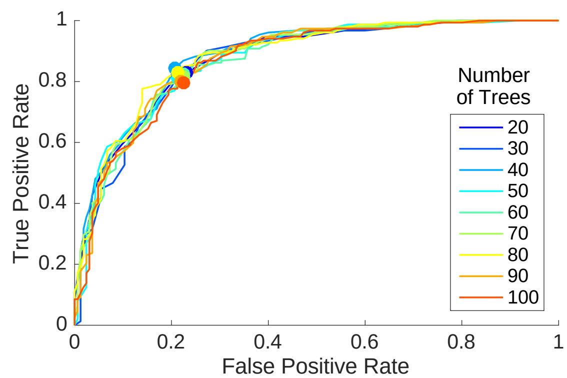


Figure 9. ROC curve for a random forest with the specified number of trees, in the prediction of ATC operational acceptability of DWR reroute advisories from May to September 2014. Filled circles represent the points with 0.5 discrimination threshold presented in Table 4.

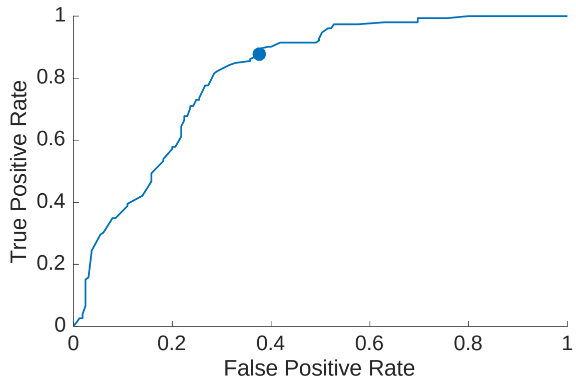


Figure 10. ROC curve for nested cross validation results for the prediction of ATC operational acceptability of DWR reroute advisories from May to September 2014. Filled circle represents the point with 0.5 discrimination threshold presented in Table 5.

Table 1. Common routes (observed more than 10 times in April 2015) originating in ZFW48, filtered for the final route fix ABQ. Route count generated from flight plans and Center route amendments extracted from ASDI data for April 2015. Dash (-) under 'Via' represents a direct route from the route start sector to the final route fix.

Route Start Sector	Via	Final Route Fix	Hist. Count
ZFW48	PNH.TCC.ACH.CLUMP.	ABQ	526
ZFW48	SPS.GANJA.TURKI.TXO.MIERA.	ABQ	373
ZFW48	UKW.GTH.TXO.MIERA.	ABQ	157
ZFW48	TXO.MIERA.	ABQ	109
ZFW48	-	ABQ	101
ZFW48	TCC.ACH.CLUMP.	ABQ	74
ZFW48	PNH.TCC.ACH.	ABQ	54
ZFW48	MRMAC.IRW.CRUSR.GOONI.PNH.TCC.ACH.CLUMP.	ABQ	44
ZFW48	PNH.ACH.	ABQ	37
ZFW48	CRUSR.GOONI.PNH.TCC.ACH.CLUMP.	ABQ	36
ZFW48	ACH.	ABQ	27
ZFW48	ADM.TXO.MIERA.	ABQ	26
ZFW48	GTH.TXO.MIERA.	ABQ	24
ZFW48	ADM.PNH.TCC.ACH.CLUMP.	ABQ	22
ZFW48	ADM.PNH.TCC.ACH.	ABQ	22
ZFW48	ADM.SPS.GANJA.TURKI.TXO.MIERA.	ABQ	21
ZFW48	TXO.	ABQ	17
ZFW48	KA30Y.	ABQ	15
ZFW48	ABI.CME.HONDS.CNX.	ABQ	14
ZFW48	IRW.CRUSR.GOONI.PNH.TCC.ACH.CLUMP.	ABQ	11

Table 2. Forward search feature selection results (F-Score), applying a random forest for the prediction of ATC operational acceptability of DWR reroute advisories from May to September 2014. Order of features selected in parentheses in column 1. Highest F-Score in columns 2-9 in bold. (RSS: Reroute Start Sector; D/C: Demand over Capacity).

	1 Feature	2 Features*	3 Features*	4 Features*	5 Features*	6 Features*	7 Features*	8 Features*	9 Features*	10 Features*
Hist. Count (Full Reroute) **(10)**	0.648	0.695	0.753	0.771	0.764	0.766	0.801	0.775	0.767	**0.780**
Hist. Reroute Count by Segment **(1)**	**0.674**	-	-	-	-	-	-	-	-	-
Direct Route **(4)**	0.387	0.597	0.705	**0.775**	-	-	-	-	-	-
No. Sectors Over **(9)**	NA	0.599	0.693	0.743	0.746	0.766	0.809	0.783	**0.797**	-
Max D/C Ratio **(8)**	0.255	0.664	0.751	0.773	0.769	0.789	0.772	**0.784**	-	-
RSS Over Capacity **(7)**	NA	0.583	0.674	0.744	0.758	0.782	**0.815**	-	-	-
RSS D/C Ratio **(6)**	0.381	0.660	0.749	0.758	0.773	**0.796**	-	-	-	-
No. Downstream Sectors **(3)**	0.484	0.667	**0.755**	-	-	-	-	-	-	-
Time to Exit MSS **(2)**	0.497	**0.719**	-	-	-	-	-	-	-	-
Dist. to Exit MSS **(5)**	0.467	0.665	0.719	0.761	**0.789**	-	-	-	-	-

* Includes feature set with highest F-Score from previous column

Table 3. Model selection results (F-Scores in bold) for the prediction of ATC operational acceptability of DWR reroute advisories from May to September 2014.

	Logistic Regression	**Decision Tree**	**SVM**	**Random Forest**	**AdaBoost**
Accuracy	0.732	0.735	0.685	0.817	0.776
Misclassification Error	0.268	0.265	0.315	0.183	0.224
True Positive Rate	0.711	0.750	0.632	0.842	0.763
True Negative Rate	0.752	0.721	0.733	0.794	0.788
Precision	0.725	0.713	0.686	0.790	0.768
F-score	**0.718**	**0.731**	**0.658**	**0.815**	**0.766**
Area Under ROC Curve	0.818	0.767	0.770	0.886	0.864

Table 4. Parameter selection results (F-Scores in bold) for a random forest with the specified number of trees, for the prediction of ATC operational acceptability of DWR reroute advisories from May to September 2014.

	Number of Trees								
	20	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
Accuracy	0.798	0.801	0.817	0.798	0.798	0.795	0.808	0.792	0.785
Misclassification Error	0.202	0.199	0.183	0.202	0.202	0.205	0.192	0.208	0.215
True Positive Rate	0.829	0.816	0.842	0.809	0.822	0.816	0.829	0.803	0.796
True Negative Rate	0.770	0.788	0.794	0.788	0.776	0.776	0.788	0.782	0.776
Precision	0.768	0.780	0.790	0.778	0.772	0.770	0.783	0.772	0.766
F-score	**0.797**	**0.797**	**0.815**	**0.794**	**0.796**	**0.792**	**0.805**	**0.787**	**0.781**
Area Under ROC Curve	0.877	0.871	0.886	0.875	0.870	0.878	0.883	0.874	0.867

Table 5. Nested cross validation results (F-Score in bold) for the prediction of ATC operational acceptability of DWR reroute advisories from May to September 2014.

	Nested Cross-Validation
Accuracy	0.744
Misclassification Error	0.256
True Positive Rate/Recall	0.875
True Negative Rate	0.624
Precision	0.682
F-Score	**0.767**
Area Under ROC	0.814

Appendix A: Subject Matter Expert Questionnaire

1. What is the process by which a controller responds to a reroute request from a pilot?
2. What information does the controller have access to when making a decision about a pilot reroute request?
    a. With regard to weather?
    b. With regard to air traffic demand?
3. What are the most important factors in determining the operational acceptability of a requested reroute in ZFW?
4. Would you consider the following factors to be relevant to a controller's decision making on a pilot reroute request in ZFW?
    a. Weather?
    b. Whether the controller is familiar with the requested reroute or not?
    c. Characteristics of the underlying airspace, including the location of other traffic flows, route compliance with LOAs, and route requirement for point-outs or interfacility coordination?
    d. Workload for the controller making the decision on the reroute request?
    e. Workload in sectors downstream of the controller making the decision on the reroute request?
    f. Reroute deviation from the original route?
    g. How close the flight is to hand-off?
5. Would you consider the following specific features to be relevant to a controller's decision making on a pilot reroute request in ZFW?
    a. Historical usage of the reroute requested?
    b. Ratio of demand to MAP value for the sector in which the maneuver is being requested to start?
    c. Whether the demand exceeds the MAP value for the sector in which the maneuver is being requested to start?
    d. Maximum ratio of demand to MAP value for the sectors downstream of the sector in which the maneuver is requested to start?
    e. Number of sectors downstream of the sector in which the maneuver is requested to start in which demand exceeds the MAP value?
    f. Number of sectors and centers impacted by the requested reroute?
    g. Whether the requested reroute in a direct routing, or a routing via an auxiliary waypoint?
    h. The time and distance before the flight is to be handed off to the next sector?
6. Any other comments?