Initial Analysis of and Predictive Model Development for Weather Reroute Advisory Use

Heather Arneson *
NASA Ames Research Center, Moffett Field, CA 94035, USA

In response to severe weather conditions, traffic management coordinators specify reroutes to route air traffic around affected regions of airspace. Providing analysis and recommendations of available reroute options would assist the traffic management coordinators in making more efficient rerouting decisions. These recommendations can be developed by examining historical data to determine which previous reroute options were used in similar weather and traffic conditions. Essentially, using previous information to inform future decisions. This paper describes the initial steps and methodology used towards this goal. A method to extract relevant features from the large volume of weather data to quantify the convective weather scenario during a particular time range is presented. Similar routes are clustered. A description of the algorithm to identify which cluster of reroute advisories were actually followed by pilots is described. Models built for fifteen of the top twenty most frequently used reroute clusters correctly predict the use of the cluster for over 60% of the test examples. Results are preliminary but indicate that the methodology is worth pursuing with modifications based on insight gained from this analysis.

I. Introduction

Currently, traffic management coordinators (TMCs) react to flow constraints in the National Airspace System (NAS) by implementing air traffic management initiatives in order to safely and efficiently route air traffic through the NAS. One method of handling traffic is to reroute air traffic around regions of airspace affected by convective weather. Acceptable reroutes are issued by TMCs based on their understanding of weather conditions and their previous experience dealing with similar conditions. However, some of this responsibility of suggesting reroutes is now being given to airlines. A new system for pre-departure rerouting, Collaborative Trajectory Options Program CTOP, is currently in limited use.[1] This program allows airlines to submit multiple preferred routes to the FAA through a trajectory option set (TOS) when severe weather will affect their flights' planned routes. A decision support tool that uses historical data to inform future decisions along with metrics of efficiency will benefit both the FAA and airlines as they develop and evaluate their rerouting solutions.

Previous and ongoing work is focused on finding similarities between days based on weather conditions or NAS performance metrics.[2,3,4] With a vast amount of weather information, the challenge here is to condense the data in such a way that operationally relevant features are preserved (or highlighted) in order to compare days. Operationally relevant features are often difficult to determine and depend heavily on the objective. In Ref. 4, days are clustered base on the Weather Impacted Traffic Index (WITI) values at each Air Route Traffic Control Center (ARTCC or Center). Once similar days are discovered, the use of particular advisory reroutes used on similar days are quantified. Previous research has also involved creating or selecting reroutes for flights which avoid airspace affected by severe weather, for example Refs. 5, 6, and 7. However, little work has been done to predict the use of specific routes in the presence of convective weather. Here, rather than identifying similar scenarios based on weather conditions, we apply reroute usage as labels for supervised learning models and allow the model to discover patterns in the weather data indicating the use of a particular reroute.

This work is a stepping stone in the development of a tool that could be used to inform the TMC and airline reroute decision process. This paper describes and shows the results of determining if a flight followed a reroute and the associated convective weather scenario. A distance metric was used to determine if flight

*Research Aerospace Engineer, Aviation Systems Division, NASA Ames Research Center, P.O. Box 1, Mail Stop 210-6, Moffett Field, CA 94035-0001, AIAA Member

followed a specific reroute. The reroutes are then clustered using the k-means algorithm[8] and this distance metric. A model is then developed using the reroute and convective information to predict whether or not a route will be used based on convective weather data.

There are two major challenges in developing predictive models for this problem. First, the number of weather features available, many of which are irrelevant, is large compared to the number of observations, making it difficult for the model to identify relevant weather features. Second, each reroute is used infrequently making it difficult for the model to extract similarities between observations. Models are at the early stages of development and results are given here to illustrate the use of machine learning methods to provide a starting point for more detailed model development.

Detailed discussions of the process of selecting and issuing reroutes, the CTOP program and methods of comparing routes are given in section II. An overview of the methodology is given in Section III. The route distance metric is presented in Section IV. Reroute implementation and use is quantified in Section V and model development and preliminary results are described in Section VI. Several features were calculated from the raw weather data and an ensemble method[9] was used to create models for clusters of advisory routes.

II. Background

In this section, some background information is given on the use of reroutes and CTOP. Previous work in the areas of route comparison and route use analysis is discussed in more detail. This section ends with a discussion of methods designed to predict the use of reroutes.

Currently, reroutes used in response to weather conditions are specified by TMCs using reroute advisories. These advisories are based on *plays* published by the FAA in the National Severe Weather Playbook.[10] Each play in the Playbook is a set of routes frequently used to react to severe weather. The Playbook currently consists of 237 plays. Advisories, as implemented, are usually based on a play, however, TMCs can and often do, modify the play. Having a set of plays available for use by TMCs aids in communication among traffic managers when planning for a specific event. However, TMCs have the flexibility to modify plays as they see fit, thus creating somewhat custom built advisories. For example, during the study period from June to August 2011, 1,669 weather required advisories were issued, with 735 unique names. These advisories consisted of a total of 34,247 routes with 2,770 unique origin and destination pairs.

In an effort to give airlines a means to state their preference for reroute options, a traffic management initiative known as the Collaborative Trajectory Options Program (CTOP) is currently being developed.[1] CTOP allows airlines to specify and rank acceptable routes for each of their flights through a Trajectory Options Set (TOS). The implementation of this program requires that airlines evaluate and file reroute options which can be used to route around convective weather. Airlines are at different stages in developing their procedures and tools required for their participation in CTOP. Airlines would like to identify geographically diverse and operationally feasible routes as potential members of a TOS. Each potential route must be evaluated based on airline objectives. A tool which can suggest previously used routes for similar weather would assist airlines in identifying potential routes which will likely be acceptable by the FAA.

There is a wealth of information available in historical records of advisory reroutes and flight tracks that can be leveraged for use in a variety of applications. This work often involves comparing routes to find similarities, form clusters, or identify outlier trajectories. Given widely varying spatial and temporal descriptions of routes and flight tracks, each method involves a representation of a route or flight track and an associated distance metric. Distance metrics that have been used for similar problems could be applied here.

De Armon et al.[11] focus on building a queueing model of the NAS and group routes in order to limit the number of paths between a given origin and destination in an effort to simplify the model. They describe each route by the sequence of airspace sectors through which the route passes and use the edit distance between these paths as the distance metric. The edit distance quantifies differences in sequences of "words" (in this case, sector identifiers) by finding the minimum number of manipulations (delete, shift or replace) of words required to morph one sequence into the other. This route representation is at a courser scale than would be useful for the current application.

Enriquez, et al.[12] parameterize flight tracks by time, scaling the duration of the flight to span the interval $[0, 1]$ and interpolating between points to get some fixed number of positions evenly spaced throughout the time interval. The distance between two routes is the sum of the Euclidean distances between corresponding

points. In this work, we are not concerned with a time component of the trajectories. The method of 12 could be modified to generate points along each route that are evenly spaced spatially and then compared using the Euclidean distance. However, a small "dogleg" in one of two nearly identical routes causes a shift in the location of these points throughout the length of the route. This shift results in a distance defined by this metric that could be quite large when in fact, the two routes overlap for most of their lengths.

Distance metrics have been developed which are based on the longest common subsequence (LCS) of the sequences of points used to describe two routes. The calculation of the LCS between two sequences is similar to the edit distance in that it is a measure of difference in the words used in each sequence. However, unlike the edit distance, an LCS method identifies ordered subsequences of words that are common to both sequences. Gariel, et al.[13] focus on identifying typical flight paths in the TRACON (Terminal Radar Approach Control) area based on historical flight paths. As a first step, turning points in trajectories are identified by finding heading changes. These turning points are clustered using the Euclidean distance as the metric to form waypoints. Each route is described as the sequence of waypoints over which the route passes. Routes are then compared to one another using an LCS-based distance metric. A similar method is presented in Vlachos, et al.[14] however, the route description is a time series of locations and is not transformed in any way. A method is presented to compare routes using a LCS method while allowing for spatial and temporal shifts in routes.

In this work, we use a spatially uniform description of routes and use a LCS based metric to quantify overlap between routes. This is quite similar to the method used in 13, however, we do not need to perform the steps of identifying and clustering turning points or identifying the waypoints through which each route passes. Instead, we divide the NAS into a grid and represent each route as the sequence of grid elements through which it passes.

III. Methodology

The main objective of this work is to predict whether or not an advisory route was used during specific weather scenarios. There are several intermediate pre-processing steps required to get to a point at which models can be developed to predict advisory route usage. At a high level, these steps involve analyzing the use of reroutes, clustering reroutes and extracting relevant features from weather data. These steps are described in more detail in the following paragraphs.

First, we would like to understand reroute use. It is possible for a route to be issued as part of an advisory but not used by a flight during the valid time range of the advisory. We will refer to routes that are issued simply as *issued routes*. We focus on routes that were flown with the intent of identifying operationally relevant routes. Those will be called *used routes*. A route is considered to be used if at least one flight which departs during the valid time range of the advisory follows that route. Flight plans could be used to identify flights that follow a particular route, however, flight plans can be amended throughout the flight. Rather than checking each flight plan for conformance to a route throughout the flight, the actually flown flight track is used to determine adherence to the route. Given the difference in spatial resolution of the advisory route description and flight track, this comparison is not straight forward. The method of describing and comparing routes is given in Section IV. Generally, the route and track descriptions are transformed into the same spatially uniform grid. Then a distance metric is defined based on the fraction of overlapping segments of the flight track and route advisory.

As mentioned previously, many of the routes used are spatially similar. When looking for trends in routing decisions, we are less concerned with the particular route used than we are with the path the route defines. That is, we are interested in predicting the use of spatially similar routes. Groups of similar routes are identified by clustering routes. Routes are first grouped by origin and destination Center or airport pairs. The routes of each group are then clustered using the k-means algorithm. The problem is then transformed into predicting whether or not a *cluster* of routes is used during a given time window. A route cluster is considered to be used during a time window if any of its members was used during that time window.

We would then like to predict the use of route clusters. Advisory routes inbound to ZNY from June to August 2011 were used for this work as part of a NASA project to improve arrival operations in that area. This time range is broken into 2,208 one-hour time windows. Routing decisions are made hours in advance of the time that weather is predicted to affect airspace and advisories have ranges of one to fifteen hours. Thus, working with one-hour intervals is an appropriate resolution given the prediction lead times and valid durations of advisories. Each of these one-hour time windows is considered an observation. Each

observation consists of weather data, referred to as *features*, and an indication of whether or not each route cluster was used during that time window, referred to as the *class* of the observation. If a route cluster is used, the observation is said to be a positive observation for that cluster, otherwise it is said to be a negative observation. The problem is then posed as a classification problem in which machine learning methods are used to predict the class (positive or negative) of each observation.

A challenge of working with this type of data is that there is a large imbalance of positive and negative observations. Routes used during this time period were grouped into 253 clusters. The top twenty most frequently used clusters were used during 50 to 240 one-hour time windows. The number of positive observations is much less than the total number of observations. The class label that appears most frequently in an imbalanced data set is referred to as the *majority class* while the underrepresented class is referred to as the *minority class*. In this case, negative observations are the majority class and positive observations are the minority class. Classification problems with imbalanced data are challenging as machine learning algorithms rely on statistics of a large amount of data to be able to extract patterns and accurately classify data. Given a large imbalance of examples of each class, machine learning algorithms will not be able to identify the minority class accurately. A minority class over-sampling technique is used to address this imbalance.

Weather information is used as the features of each observations. The features make up a set of data for each observation that is used by a machine learning algorithm to generate models to predict the class of the observation. Weather predictions and actual data for the NAS are available at high spatial and temporal resolutions making the size of available data impractical for use as is in model generation. For this work, observed cloud echo top height measurements are used to describe convective weather conditions. Those data can be used to quantify flight deviations based on the comparison of echo top and the flight altitude of the aircraft. The number of echo top readings which cover the NAS is 57,540. In order to extract high level trends in the data, the echo data are first down-sampled spatially and averaged over the one hour time windows. This reduces the number of weather features covering the NAS to 1,000. The number of weather features is still quite large, with many being irrelevant to the use of a particular cluster. Eighteen hand selected features are then calculated from these data with an attempt to reveal weather conditions at locations that are key to determining whether or not a route cluster was used.

Finally, with the above organization of cluster use and weather features, machine learning models are developed. A model is built for each cluster individually which takes weather features of observations as input and predicts the class of each observation. The models are built using labeled data, that is, a set of observations including weather features and class label.

IV. Comparison of Paths

Advisory reroutes are based on plays, but TMCs do have the flexibility to modify these routes. Given that there are various ways to specify and implement rerouting, there may be hundreds of unique routes used to react to similar whether conditions, but their differences may be small enough to be considered practically equivalent. An integral part of finding trends in the methods used to react to similar weather is having a means to identify similar routes and flights that follow those routes. A first step in this work is to develop a distance metric which can be used to compare routes to one another and routes to flight tracks. Since this comparison can be made between advisory routes and flight tracks, routes and tracks will be referred to generically as *paths*.

Routes are described in advisories and plays by a text description of the waypoints, fixes and jet routes used to define the route. A comparison of these text descriptions can be used to identify identical or similar routes. However, even if the text descriptions do not match, two routes may be spatially similar. Since traffic is routed around certain regions of airspace to avoid severe weather, the geographical location of routes is important. Thus, the physical proximity of two routes should be quantified, regardless of the text description used to define the routes. In Section IV.A gives more detailed descriptions of routes and flight tracks. The distance metric used is presented in Section IV.B.

IV.A. Path Description

Each reroute advisory or play consists of several routes. The routes are described by an origin, destination and route between the origin and destination. Origins and destinations are specified as an airport or Air Route Traffic Control Center (ARTCC or Center). The route consists of a series of departure fixes, waypoints,

jet routes and arrival fixes which are specified by alphanumeric identifiers. Jet routes themselves are defined as a series of fixes and waypoints. Each fix and waypoint has a location specified by latitude and longitude coordinates. During the months of June through August 2011, 1,669 weather required advisories were issued, with 735 unique names. These advisories consisted of a total of 34,247 routes with 2,770 unique origin and destination pairs. Each advisory route is described by up to 100 coordinates. There are 3,234 unique coordinates used in the description the routes issued during the study period.

Flight track data are available in Aircraft Situation Display To Industry (ASDI) data. RADAR systems are used to track flights and the speed, altitude and position of flights are calculated. Updates are provided every 12 seconds for flights in enroute airspace and every 5 seconds for flights within a TRACON. This information is then made available through ASDI.

Both advisory routes and flight tracks can be described by a sequence of points, specified by latitude and longitude coordinates. The mathematical description of paths used in the development of the distance metric is given in Appendix A.A.

IV.B. Distance Metric

The points used to describe a given path are not necessarily evenly spaced. An overlapping stretch of two paths may be described by a different number of points, thus a point-by-point comparison of two paths would not yield a meaningful distance. Furthermore, when tracks are compared to paths, the problem is further complicated due to the continuous nature of aircraft flight position measurements.

The objective in the development of a distance metric for this application is to design a distance metric which:

1. Provides a meaningful measure of spatial distance between paths
2. Is robust to path specification
3. Can be used to assess partial overlap between paths.

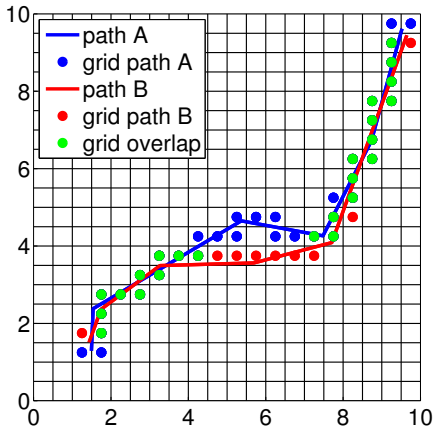


Figure 1: Example paths, grid representations and overlapping grid elements.

In order to capture spatial properties of each path while being robust to the path description, a grid based representation of paths is used. The path description is spatially uniform, with airspace divided into a grid. A path is described by the sequence of grid elements through which it passes. Distances between paths are based on the overlapping portions of the paths. This can be used to test for full or partial overlap of paths. The grid size is specified by the user and affects the tolerance of the metric. The specific grid size choice used for the problem addressed here is discussed in Section V. An illustrative example is shown in Fig. 1 in which two paths are plotted along with their respective grid representations. The overlapping grid elements are highlighted in green. The mathematical development of the distance metric calculation is given in Appendix A.B.

IV.C. Characteristics of Metric

An advantage of this route description and distance metric comes from the spatially uniform description of paths. However, the choice of Δlat and Δlon will affect the distances between paths. Larger Δlat and Δlon will allow larger deviations to be considered "identical" while smaller Δlat and Δlon requires "identical" routes to be closer together. A downside of this technique is that two parallel path segments may be separated by an amount less than Δlat or Δlon but have grid representations that do not overlap.

The user is given the choice of the denominator in the fractional difference calculation to allow for different objectives. A flight track is likely to extend beyond the reroute description, thus, we would like to find the fraction of the flight track overlapping the advisory route, and not vice-versa. The minimum length of the two paths compared is used as the denominator for this case.

If routes do not use overlapping grid elements, the distance is 1, thus, there is no distinction between routes that are close and parallel and routes that may be on opposite sides of the country. To address this issue, this metric could be used at different levels, that is, start with a relatively large Δlat, Δlon to identify

roughly close routes. Then, a smaller grid Δlat, Δlon can be used to get a more precise distance between routes identified as close using the coarser grid description.

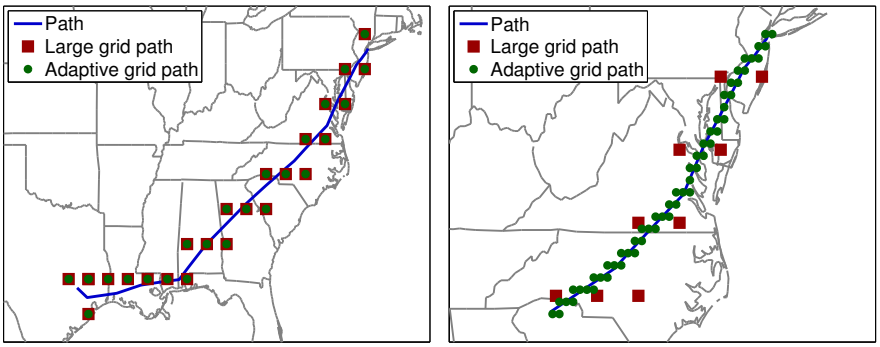
V. Advisory Route Use Analysis

Here, flight tracks are matched to advisory routes. The results presented here are for weather required advisories issued from June 2011 through August 2011 that have their destination within New York Center (ZNY). The distance metric described above is used to compare advisory routes with flight tracks. The procedure used to identify matching flights and routes is detailed below, followed by analysis results.

For each flight, all advisories active at the time of departure are identified. The origin and destination airports of the flight are then compared to each active advisory route origin and destination. Recall that the origin and destination of a route may be a Center or an airport. If the origin (destination) is specified as an airport, the flight origin (destination) matches the route origin (destination) if the airports match. If the origin (destination) is specified as a Center, the flight origin (destination) matches the route origin (destination) if the flight origin (destination) airport is within this Center. Flights and active routes which match both the origin and destination are referred to as *origin-destination matches*.

Next, a coarse distance is calculated between each flight and each advisory route in its set of origin-destination matches using grid elements of size 1.50° latitude by 1.50° longitude, roughly 90 nmi by 70 nmi. If the coarse distance is less than 0.5, a finer distance is calculated using a smaller grid size. An adaptive grid size is used as the smaller grid size, varying from 0.25° latitude by 0.25° longitude, roughly 15 nmi by 12 nmi, to 1.50° latitude by 1.50° longitude. If the advisory route consists of 10 or fewer grid elements in the large grid representation, the small grid size is used for the second comparison. If the advisory route consists of 25 or more grid elements in the large grid representation the second comparison is not made. Grid size is varied linearly between the extremes based on the number of large grid elements needed to describe the advisory route. The extreme values of the grid size and the function used to calculate the grid size for the finer comparison are user specified. For this study, values were chosen based on visual inspection of various routes and comparisons performed with various grid sizes.

The extremes of the grid element sizes are illustrated in Fig. 2. Two advisory reroutes are shown. Figures 2a and 2b show the waypoint descriptions of two advisory routes overlaid with their large grid representations and adaptive grid representations. The route shown in 2a is described by 25 large grid elements, with a corresponding adaptive grid size equal to the large grid size. The route shown in 2b is described by 10 large grid elements, with a corresponding adaptive grid size equal to the small grid size.



(a) Path A (b) Path B

Figure 2: Figures 2a and 2b show the waypoint descriptions of two advisory routes overlaid with their large grid representations and their respective adaptive grid representations.

If the distance between the advisory route and the flight tracks (using the minimum number of grid elements used to describe advisory route or the flight tracks as the denominator) is less than a user defined threshold then the flight is considered a possible match for that advisory route. Adjusting the threshold will affect the results of matching flights to advisory routes and will be considered in future work, as illustrated in Fig. 3. Adaptive grid path distances were only calculated when the large grid path distance was at or below 0.5, thus, the asymptote of the adaptive grid curve is the fraction of flights that match at least 50%

of the route using the large grid.

For the purposes of this work, this threshold was set to 0.15 (marked in Fig. 3) for the adaptive grid distance, requiring that a flight track overlap an advisory route for at least 85% of the route to be considered a potential match. Each flight is matched to at most one route, selected based on the distance between the flight and all of the potential route matches. These matches are referred to as *path matches*. Routes with at least one path match are considered used routes.

Flight tracks and advisory matching was performed using the above procedure for flights inbound to ZNY, with results shown in Table 1. Results are presented for flights, advisories and routes. Of the flights inbound to ZNY during the period of the study, 3.71% are an origin-destination match to at least one advisory route. Of those origin-destination matches, 29.07% are also path matches of the advisory route. Looking at advisories, 92.47% of the issued advisories consist of at least one route which is an origin-destination match of a flight. Of these advisories, 78.29% consisted of at least one route matched by a flight. This indicates that, while a path match may not be found for every route within an advisory, most advisories consisted of routes that were matched by at least one flight. Of the individual routes issued, 44.17% have an origin-destination match, and only 20.22% of these routes are matched by a flight.

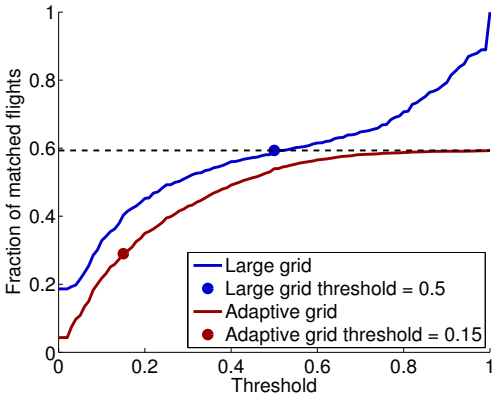


Figure 3: The fraction of flights matching the origin and destination of an advisory route that also match the path as a function of the cutoff threshold for path matching.

In summary, we found that roughly 29% of flights which match the origin and destination of an advisory route actually use an advisory route and that only 20% of routes between an origin destination pair matching at least one flight are actually used. There are several factors that could contribute to the resulting low conformance of flights to advisory routes. Here, we consider a flight track to match a route if their paths overlap for more than 85% of the shorter path. Segments of routes can be omitted for flights departing from particular airports, allowing these flights to join up with the route at some intermediate point. Additionally, advisories can be canceled or modified and thus, comparisons made here could include inactive routes.

Table 1: Flights and Advisories inbound to ZNY.

	Total	Origin - Destination		Path	
		Matched	% Matched	Matched	% Matched
Flights	190,046	7,060	3.71%	2,052	29.07%
Advisories	279	258	92.47%	202	78.29%
Advisory routes	4,476	1,977	44.17%	905	20.22%

VI. Modeling the Use of Advisory Reroutes

The focus of this section is on the development of a predictive model of advisory routes inbound to ZNY from June to August of 2011. The time range is broken into 2,208 one-hour time windows. The objective is to predict the use of specific routes during a given hour based on weather conditions. Each time window is considered an observation, consisting of a set of features and a class label. The features consist of weather data. The class label is positive if the advisory cluster was used at least once during the time window, otherwise, the label is negative. There are two main challenges in developing a model for this particular data: the large number of features available compared to the number of observations and the low ratio of positive to negative examples. When the number of features is large compared to the number of observations, models tend to capture random error and can easily be over-fit to the training data. The number of times each route is used throughout this time period is much less than the number of times it is not used. Building models from an imbalanced data set tends to result in large error rates in the prediction of the minority class.

VI.A. Clustering Routes

Over 4,000 advisory routes were issued during the time frame of interest. Many of these routes are identical or practically identical. In order to understand reroute use, we would like to first group identical or similar reroutes.

The first step in the clustering is to group routes by origin and destination. Given that the path portion of the advisory reroute description may lead from a Center rather than a specific airport, several routes have paths which overlap but originate in different Centers. Airports within 100 nmi of each other were considered to be identical for the purpose of this grouping. A total of 72 groups were formed at this stage. The grid-based metric with the large grid size is then used to compare paths. Clustering within each origin-destination group was performed using the k-means algorithm. The total number of resulting clusters is 253. The top 20 most frequently used clusters are used during 50 to 240 of the 2,208 one-hour time windows. Formulating the prediction problem as predicting the use of route clusters rather than individual routes increases the number of positive examples in the data set. However, even after clustering, the imbalance between positive an negative observation for each cluster is large. This imbalance is addressed in the model formulation.

VI.B. Weather Features

Next, we process weather data to create features. The specific weather data used in this work are measurements of convective cloud top altitude, referred to as echo tops. These values are estimates of the top of clouds based on radar measurements. The values are discrete altitude levels, 0 ft to 50,000 ft at 5,000 ft intervals for a total of 11 levels. Readings of echo tops are available every 2.5 minutes at a spatial resolution of 4.65 nmi. The number of echo top readings which cover the NAS is 57,540. If echo top values at this spatial and temporal density were naively used as features in a predictive model it would likely lead to over-fitting the training data. In order to reduce the number of features, echo top values are averaged and then used to create a smaller set of features.

Values were averaged at the center of grid elements of size 1.25° latitude by 1.25° longitude, roughly 75 nmi by 58 nmi (1,000 total grid elements of this size cover the NAS), over one-hour time windows. These values were then binned into 11 bins to mimic the echo top data.

This is a large number of features, most of which are likely unrelated to the use of a particular cluster of routes. In order to reduce the number of features used to generate the predictive models and attempt to extract information relevant to the use of a particular cluster, several features were calculated from the gridded echo top values. These features were created with the intent of condensing the large feature space of the echo top data to fewer, more relevant features. The focus was on capturing echo top information along the route cluster and direct from origin to destination. The motivation behind this is to create features which can be used to compare weather conditions on the direct path to an optional reroute. If the weather along the direct path is mild, there would be no need to issue a weather required advisory between that origin and destination. Here, we use the direct path between origin and destination. Details of the calculation of the features used are given in Appendix B.

VI.C. Model Formulation

At this point, we have organized weather and reroute use information into observations consisting of weather features with class labels for each reroute cluster and we are ready to create models of reroute cluster use. Models were generated for each advisory route cluster with the objective of determining the time windows in the test set in which the advisory cluster was used.

We use a random forest method[9] for this classification problem. A random forest method is an example of an ensemble method. Ensemble methods consist of many simple classifiers that may result in poor classification results individually. The predictions of these weak classifiers are then given a weighted vote to give the final prediction of the ensemble method. The predictions of each weak learner are weighted based on the fraction of misclassified examples in the training set. The specific type of classifier used as the weak learners in a random forest method are decision trees.

The weak learners are developed using only a subset of the training observations and a subset of the available features. Observations are selected uniformly at random from the training set for use in the development of each weak learner. The observations are selected in a way that increases the ratio of positive to negative examples compared to the ratio found in the full training set. However, oversampling of the positive

examples in this way can lead to over-fitting the model to the training set. We use Synthetic Minority Oversampling Technique (SMOTE)[15] to create additional positive examples. In SMOTE, features of additional positive observations are created by selecting a positive observation and taking a convex combination of that observation and one if its nearest neighbors. These observations are referred to as "synthetic" since they do not naturally appear in the data. This process can be repeated using other neighbors of the selected positive observation. This forces the decision region of the positive examples to be more general and thus less likely to over-fit the positive examples in the training set than simply oversampling the positive observations. In this work, synthetic positive observations are created from the features listed in section VI.B.

Each of the weak learners exhibits poor performance. The predictions from these models are given a weight $\alpha \in \mathbb{R}$ based on their respective prediction errors ε. Here, we use a function of ε to calculate α. This function of ε has a value of 0 at $\varepsilon = 0.5$ and increases as ε decreases. Learners with prediction errors $\varepsilon \geq 0.5$ are given a weight of $\alpha = 0$. Using this weighting function, weak learners with low error are given high weights. The ensemble prediction for the test data set is a weighted vote from each of the weak learners.

There are several parameters of the algorithm which the user must specify. These parameters are: the number of synthetic positive observations to be created from some number of nearest neighbors, the fraction of positive examples in the training set for each weak learner, the number of features used for each weak learner, and the tree depth of each weak learner. Typically, tree depths of one or two are used as deeper trees tend to over-fit the test data.

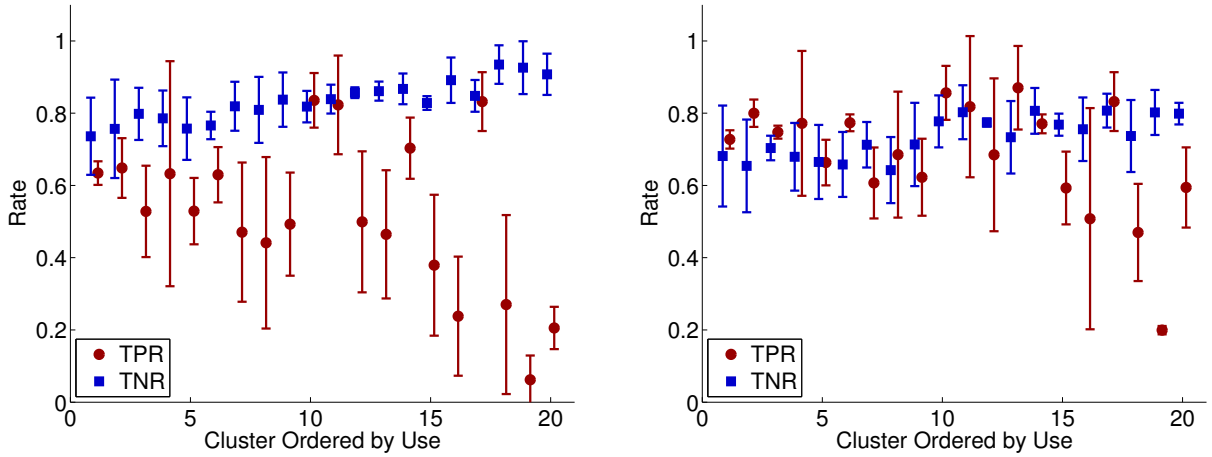
In order to test the performance of each ensemble model, we need to apply it to a testing set of observations that were *not* used in the development of the model. Here, we use K-fold cross-validation method to assess the performance of the ensemble models. In this method, the original data set is divided into K equal sized subsets, where K is a user specified value. The cross-validation is then performed K times for $k = 1, 2, \ldots K$. For each k, the k^{th} subset is held out from the rest of the subsets. The holdout set is referred to as the testing set, the other observations make up the training set. The ensemble model is built using *only* the data available in the training set. Once the model has been built, it is used to predict the classes of the observations in the testing set. The testing set prediction results for each value of k are used to describe the performance of the modeling method.

VI.D. Results

Predictive models were developed for each route cluster using the method described in Section VI.C. Models were generated for each cluster individually using the same parameters of the algorithm for each model. The particular parameters used were hand selected through trial and error. The data were divided into three sets for three-fold cross validation. The same test and training sets were used for the ensemble models of each cluster. SMOTE was used to increase the number of positive examples in the training set by a factor of nine. Positive and negative examples were selected uniformly at random for the sub-training set with a ratio of 3:2. Ten features were selected uniformly at random for each weak learner. Decision trees of depth two were used as the weak learners.

Prediction results are shown in Fig. 4 for models which do not make use of SMOTE and models that do. True positive rates (TPR) and true negative rates (TNR) are plotted for each of the top 20 most frequently used clusters. TPR indicates the fraction of positive observations that were correctly classified as positive and TNR indicates the fraction of negative observations that were correctly classified as negative. A perfect classifier would have both TPR and TNR equal to one. The top 20 most frequently used clusters were selected based on the number of time windows in which a cluster was used. These particular clusters were used in 50 to 240 time windows. Figure 4a shows results for classifiers built to predict the use of advisory routes in a given time window without using SMOTE, while figure 4b shows corresponding results from classifiers built using SMOTE. Results shown in fig. 4a indicate that prediction performance degrades considerably as the number of positive observations in the data set decreases. Including additional positive features using SMOTE greatly improved the classification results. A total of 15 of 20 clusters have average TPR and TNR above 0.6 for the classifiers which made use of SMOTE compared to only 8 for models which did not make use of SMOTE. While TPR and TNR over 0.6 is only marginally better than random guessing, this does indicate that the use of machine learning methods for this problem could provide useful results. These results show that even a relatively simple predictive model is able to extract patterns and trends from the data to predict whether or not a route cluster will be used.

To better understand the models and prediction results, we will take a more detailed look at a specific cluster. Members of the ninth most frequently used cluster and its cluster center are shown in Fig. 5 are



(a) Test set prediction results for models built to predict the use of advisory route clusters without the use of SMOTE.

(b) Test set prediction results for models built to predict the use of advisory route clusters with the use of SMOTE.

Figure 4: The mean and standard deviation over each cross validation set of true positive rates (TPR) and true negative rates (TNR) are plotted for the top 20 most frequently used clusters. The results are ordered such that the results for the most frequently used cluster are on the left and results for the least frequently used cluster are on the right.

examined in more detail.

Figure 6a shows the distribution of prediction results for each weak learner used in the ensemble classifier for one of the cross validation sets. Notice that both true positive and true negative rates vary greatly and are quite poor in certain instances. However, the ensemble model which uses the weighted vote of the predictions of each of the weak learners achieves a TPR and TNR of 0.69 and 0.79, respectively.

As mentioned above, there are several parameters which need to be set in order to generate models of this type. Varying any one of these parameters will affect the prediction results. Here, we examine one of these parameters in further detail. The fraction of positive examples used in the sub-training set for the models used to generate the prediction results shown in Fig. 4 was 0.6. We examine the affect of varying this value by building models with various values of this fraction. The fraction of positive examples in the sub-training set were varied from 0.1 to 0.9, with results shown in Fig. 6b. When the fraction of positive to negative observations is 0.1, little information is given to the classifier to distinguish positive observations from negative. With more negative observations in the training set, the classifier is biased towards predicting the negative class. Looking at Fig. 6b, it can be seen that for low values of this fraction, TNR is close to 1 while TPR is 0. As this ratio is increased, the classifier is better able to distinguish positive and negative observations and more balanced prediction results are seen. At the other extreme, when the ratio of positive to negative observations is 0.9, most examples are positive and the classifier has little information about negative observations, thus, the classifier is biased toward positive predictions. Understanding how TPR and TNR vary with this parameter could help a user tune this parameter to achieve desired results. For example, if the mis-classification of a positive example is costly or dangerous, it may be desirable to select a higher ratio of positive examples in the sub-training set to correctly classify more of the positive examples if the increase in incorrectly classified negative examples is acceptable.

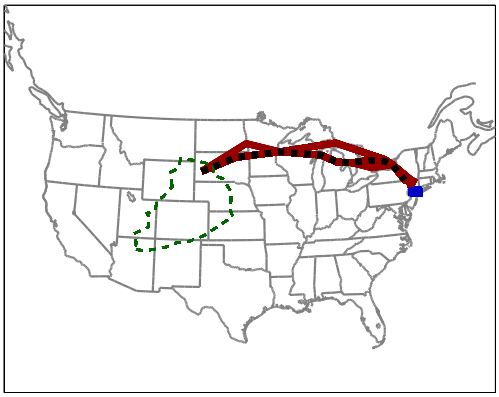
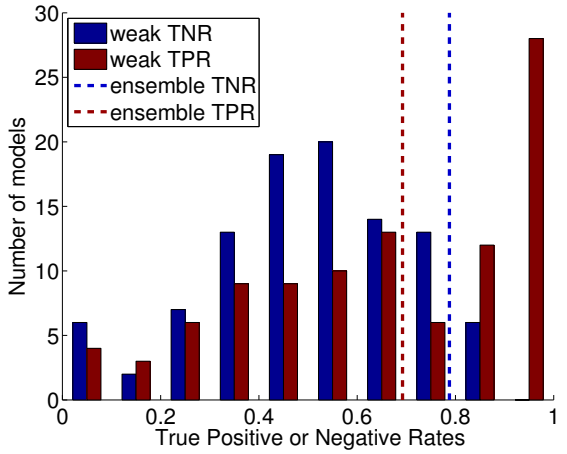
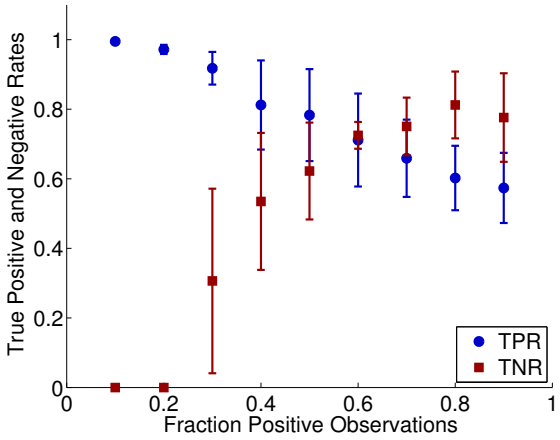


Figure 5: Example reroute advisory cluster. Member routes are shown in red with the cluster center shown in black cluster origin Center is outlined in green.



(a) Intermediate TPR and TNR for the weak learners for one cross validation set and the resulting TPR and TNR of the ensemble model.



(b) True positive and negative rates as a function of the ratio of positive to negative observations used in the sub-training sets for each of the weak learners.

Figure 6: Additional prediction results for advisory reroute cluster shown in Fig. 5.

VII. Conclusions and Future Work

A method of processing advisory reroute data, flight track data and weather data was presented and a model was developed to predict weather or not particular advisory reroutes were used given weather conditions. While the prediction of true positive and negative rates are lower than what would be useful in the generation of a decision support tool, they do indicate that this line of research is worth pursuing with modifications. While we have data describing weather and traffic conditions, there are many factors being considered by the TMC that are not represented in this data and are not well documented. Furthermore, the translation of the weather data that we do have into an operationally relevant form is an art requiring knowledge of the domain and factors influencing routing decisions. While several hand selected features were used in this work, a more systematic method of data reduction should be developed.

A method of comparing advisory routes, and flight paths was presented and used to analyze the historical use of advisory reroutes. A machine learning technique was then developed to predict the use of reroute advisory clusters based on convective weather information. Model development involved addressing two fundamental challenges inherent in working with the reroute and weather data, namely the abundance of features and relatively low number of positive examples in the data set. Fifteen out of the twenty clusters for which models were built result in average true positive and true negative rates above 60%. While these results are preliminary, they do indicate that applying machine learning techniques to the problem of predicting advisory reroute use could give meaningful and useful results. Several ways in which analysis and modeling could be modified to improve prediction results are discussed in the following paragraphs.

The path comparison method could be improved to better indicate the proximity of paths when they do not have grid elements in common but are nevertheless close spatially. Modifying this distance metric based on the techniques presented in Ref. 14 or the using the Hough transform[16] could add more flexibility in route matching. The choice of the threshold used to specify the distance between paths that are considered a match should be further studied. Partial matches may be operationally significant if, for instance, a close match occurs near the departure or arrival airport where traffic density is higher than in enroute airspace.

Features were created from weather data in order to highlight convective weather along the routes and along the direct path from origin to destination. Further work can be done to generate more features, focusing on the locations of important fixes or jet routes along the reroute and along typical direct routes. Additionally, a finer grid size could be used at key locations along the route, such as arrival fixes or within the TRACON of the destination airport. This finer resolution would help in determining which among several routes would be most appropriate to react to weather conditions.

There are many types of weather data and forecasts available. In this work, the convective weather data used was actual echo tops. These values indicate the tops of could. While this is important in determining

if flights are able to fly above a storm or around it, more information is needed to understand the impact of the storm on flights. Vertically Integrated Liquid (VIL) combined with echo tops gives a better indication of the intensity of convective weather. The Convective Weather Avoidance Model (CWAM)[17] model makes use of both echo and VIL to model the probability of pilot deviation around weather. The use of CWAM should give a better indication of weather or not a flight would follow a specific path.

The use of weather forecast data to predict the implementation and use of reroutes is a necessary step in the development of a tool that could be used to select reroutes as reactions to specific weather forecasts. TMCs use the Collaborative Convective Forecast Product (CCFP) during planning. CCFP is a forecast that combines weather predictions from various sources. The forecast indicates the predicted intense convection activity and is updated every 2 hours with forecasts made for 2, 4, and 6-hour periods. Using the CCFP forecast that TMCs were using to make reroute advisory decisions will be useful in predicting whether or not a reroute was issued.

There are a variety of ways in which the machine learning method could be improved. There are several parameters of the model that the user must specify. Future work will include a more systematic approach to setting these parameters for the models of each cluster individually. Minority class oversampling was performed using SMOTE to address the imbalance of the data set. Expanding the data set to include summer months of additional years would add more positive observations and could also improve the modeling results. Rather than predicting whether or not a route cluster will be used, one could develop models to predict how likely it is that a route cluster would be used. For instance, if many flights follow the same route cluster in a given hour, it may be more likely that a route from that cluster would be used during similar weather conditions. The number of flights matching a path along with a varying threshold for path matching to distinguish the level of conformance to a path could be used in models designed to predict the probability of route use.

Acknowledgments

I would like to thank Michael Bloem, Shon Grabbe, Deepak Kulkarni, Avijit Mukherjee, and Banavar Sridhar for conversations and recommendations regarding this work.

References

[1] "Collaborative Trajectory Options Program (CTOP)," http://www.nbaa.org/ops/airspace/tfm/tools/ctop.php.

[2] Hoffman, B., Krozel, J., Penny, S., Roy, A., and Roth, K., "A Cluster Analysis to Classify Days in the National Airspace System," *AIAA Guidance, Navigation and Control Conference*, 2003, p. 11.

[3] Asencio, M., "A clustering approach for analysis of convective weather impacting the NAS," *Integrated Communications, Navigation and Surveillance Conference (ICNS), 2012*, April 2012, pp. N4–1–N4–11.

[4] Mukherjee, A., Grabbe, S., and Sridhar, B., "Classification of Days Using Weather Impacted Traffic in the National Airspace System," *Transportation Research Board 92nd Annual Meeting*, Washington DC, January 2013.

[5] Wanke, C. and Greenbaum, D., "Sequential Congestion Management with Weather Forecast Uncertainty," *AIAA Guidance, Navigation and Control Conference and Exhibit*, No. AIAA 2008-6327, August 2008.

[6] Taylor, C. and Wanke, C., "Dynamic Generation of Operationally Acceptable Reroutes," *Proceedings of 9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO)*, No. 2009-7091, AIAA, September 2009.

[7] McNally, D., Sheth, K., Gong, C., Love, J., Lee, C. H., Sahlman, S., and Cheng, J., "Dynamic weather routes: a weather avoidance system for near-term trajectory-based operations," *28th international congress of the aeronautical sciences (ICAS), Brisbane*, 2012.

[8] MacQueen, J. et al., "Some methods for classification and analysis of multivariate observations," *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Vol. 1, Oakland, CA, USA., 1967, pp. 281–297.

[9] Tan, P.-N., Steinbach, M., and Kumar, V., *Introduction to Data Mining*, China Machine Press, 2005.

[10] "National Severe Weather Playbook," http://www.fly.faa.gov/PLAYBOOK/pbindex.html, 2013.

[11] DeArmon, J. S., Taylor, C. P., Masek, T., and Wanke, C. R., "Air Route Clustering for a Queuing Network Model of the National Airspace System," *14th AIAA Aviation Technology, Integration, and Operations Conference*, American Institute of Aeronautics and Astronautics, 2015/05/15 2014.

[12] Enriquez, M. and Kurcz, C., "A Simple and Robust Flow Detection Algorithm Based on Spectral Clustering," *Submitted to 5th International Conference on Research in Air Transportation (ICRAT 2012)*, 2012.

[13] Gariel, M., Srivastava, A., and Feron, E., "Trajectory Clustering and an Application to Airspace Monitoring," *Intelligent Transportation Systems, IEEE Transactions on*, Vol. 12, No. 4, Dec 2011, pp. 1511–1524.

[14] Vlachos, M., Kollios, G., and Gunopulos, D., "Discovering similar multidimensional trajectories," *Data Engineering, 2002. Proceedings. 18th International Conference on*, 2002, pp. 673–684.

[15] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P., "SMOTE: Synthetic Minority Over-sampling Technique," *Journal Of Artificial Intelligence Research*, Vol. 16, 2002, pp. 321 – 357.

[16]Ballard, D., "Generalizing the Hough Transform to Detect Arbitrary Shapes," *Pattern Recognition*, Vol. 13, No. 2, 1981, pp. 111–122.

[17]DeLaura, R., Robinson, M., Pawlak, M., and Evans, J., "Modeling convective weather avoidance in enroute airspace," *13th Conference on Aviation, Range, and Aerospace Meteorology, AMS, New Orleans, LA*, Citeseer, 2008.

[18]Bergroth, L., Hakonen, H., and Raita, T., "A survey of longest common subsequence algorithms," *String Processing and Information Retrieval, 2000. SPIRE 2000. Proceedings. Seventh International Symposium on*, 2000, pp. 39–48.

A. Path Comparison

A.A. Path Description

Each path can be described by a sequence of points, specified by latitude and longitude coordinates, for example

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{p}_1 \\ \boldsymbol{p}_2 \\ \cdots \\ \boldsymbol{p}_m \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \\ & \cdots \\ p_{m1} & p_{m2} \end{bmatrix}$$

where m is the total number of points used to describe the path.

The latitude and longitude coordinates of each waypoint are rounded to the nearest multiple of Δlat, Δlon, effectively dividing the NAS into a grid of size Δlat by Δlon. Grid elements between waypoints are included by assuming a straight line trajectory between waypoints. The grid based description of route $\boldsymbol{P}$ is

$$\hat{\boldsymbol{P}} = \begin{bmatrix} \hat{\boldsymbol{p}}_1 \\ \hat{\boldsymbol{p}}_2 \\ \cdots \\ \hat{\boldsymbol{p}}_{\hat{m}} \end{bmatrix} = \begin{bmatrix} \hat{p}_{11} & \hat{p}_{12} \\ \hat{p}_{21} & \hat{p}_{22} \\ & \cdots \\ \hat{p}_{\hat{m}1} & \hat{p}_{\hat{m}2} \end{bmatrix},$$

such that

$$\hat{\boldsymbol{p}}_{i+1} = \hat{\boldsymbol{p}}_i + \begin{bmatrix} u_i \Delta\text{lat} & v_i \Delta\text{lon} \end{bmatrix}, \forall\ i = 1, 2, \ldots, \hat{m} - 1$$

where $u_i \in \{-1, 0, 1\}$, $v_i \in \{-1, 0, 1\}$ and $|u_i + v_i| = 1$.

A.B. Distance Metric

The distance between two paths is based on the overlap of the grid representations of the paths. The distance is defined as the fractional difference between the number of grid elements in common to both paths and the number of grid elements used to describe one of the paths being compared. The calculation of the distance between two routes, $\boldsymbol{P}$ and $\boldsymbol{Q}$, according to this metric can be described in steps, as follows.

First, find the grid representations $\hat{\boldsymbol{P}}$ and $\hat{\boldsymbol{Q}}$ of routes $\boldsymbol{P}$ and $\mathbf{Q}$. Next, find subsets of $\hat{\boldsymbol{P}}$ and $\hat{\boldsymbol{Q}}$ that consist of sequential points within the original vectors and are common to both routes. A sequential subset of route $\hat{\boldsymbol{P}}$ is described as

$$\begin{bmatrix} \boldsymbol{p}_a \\ \boldsymbol{p}_{a+1} \\ \boldsymbol{p}_{a+2} \\ \cdots \\ \boldsymbol{p}_{a+l} \end{bmatrix}$$

for some $a, l \in \mathbb{N}$ such that $1 \leq a \leq a + l \leq \hat{m}$.

The longest common sequential subsets of $\hat{\boldsymbol{P}}$ and $\hat{\boldsymbol{Q}}$ are

$$\text{common}(\hat{\boldsymbol{P}}, \hat{\boldsymbol{Q}}) = \{\boldsymbol{S}_1, \boldsymbol{S}_2, \ldots, \boldsymbol{S}_{\boldsymbol{R}}\}$$

where $R \in \mathbb{N}$ is the number of common sequential subsets and

$$\boldsymbol{S_i} = \begin{bmatrix} \boldsymbol{s}_{i_1} \\ \boldsymbol{s}_{i_2} \\ \cdots \\ \boldsymbol{s}_{i_{l_i}} \end{bmatrix}$$

such that, for $i = 1, 2, \ldots, R$,

$$\boldsymbol{s}_{i_j} = \hat{\boldsymbol{p}}_{a_i+j-1} = \hat{\boldsymbol{q}}_{b_i+j-1}, \text{ for } j = 1, 2, \ldots, l_i$$

and the common subsets do not overlap, that is

$$1 \leq a_{1_0} \leq a_{1_1+l_1-1} \leq a_{2_1} \leq \ldots \leq a_{R_1+l_R-1} \leq \hat{m}$$
$$1 \leq b_{1_0} \leq b_{1_1+l_1-1} \leq b_{2_1} \leq \ldots \leq b_{R_1+l_R-1} \leq \hat{n},$$

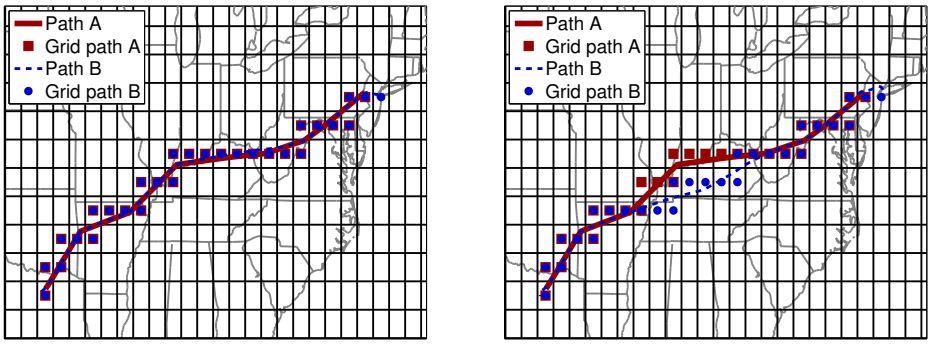
where $\hat{m}$ and $\hat{n}$ are the number of grid elements of $\hat{\boldsymbol{P}}$ and $\hat{\boldsymbol{Q}}$, respectively. There are various algorithms available to compute the longest common subsequence amongst a set (or in this case, pair) of sequences.[18]

The distance between $\hat{\boldsymbol{P}}$ and $\hat{\boldsymbol{Q}}$ is

$$\text{dist}\left(\hat{\boldsymbol{P}}, \hat{\boldsymbol{Q}}\right) = 1 - \frac{\text{length}\left(\text{common}\left(\hat{\boldsymbol{P}}, \hat{\boldsymbol{Q}}\right)\right)}{c} = 1 - \frac{\sum_{i=1}^{R} l_r}{c}$$

where c may be $\min(\hat{m}, \hat{n})$ or $\max(\hat{m}, \hat{n})$, depending on the objective of the comparison. Note that the resulting distance value is unitless and represents the fraction of overlapping grid elements. If $c = \min(\hat{m}, \hat{n})$ is chosen, the distance between $\hat{\boldsymbol{P}}$ and $\hat{\boldsymbol{Q}}$ will be zero if the two paths overlap for the entire length of the shorter path, regardless of whether the longer path extends beyond the shorter. Alternatively, c could be chosen as $\max(\hat{m}, \hat{n})$ to compare the shorter path to the longer.

An advisory reroute is compared to two different flight tracks and shown in Fig. 7. Using this distance metric, the overlap error of the flight track in Fig. 7a is 0, while the overlap error of the two paths in Fig. 7b is 0.21. Both difference are calculated with respect to the minimum number of grid elements required to describe each path.



(a) Grid representations of advisory route (red) and flight track (blue).

(b) Grid representations of advisory route (red) and flight track (blue).

Figure 7: Two flight tracks are compared to the same advisory route. The destination of this particular route is specified as an ARTCC, specifically New York Center (ZNY), and thus ZNY is outlined (dashed red). A grid element size of 0.25° latitude by 0.25° longitude is used for the grid representation.

B. Weather Features

For each cluster of advisory routes grid elements along the route specified as the cluster center μ were identified. For a given cluster i, this set of grid elements is $\mathcal{G}_{\mu_i}$ Additionally, grid elements along the direct route between the origin and destination of the cluster center were identified. For a given cluster i, this set of grid elements is $\mathcal{G}_{\text{dir}_i}$ The complement of these sets, that is, all grid elements not in each of these sets, will be represented by $\mathcal{G}^c_{\mu_i}$ and $\mathcal{G}^c_{\text{dir}_i}$, respectively.

The full list of features used in model generation are:

1. The hour of the beginning of the time window
2. Mean of grid element echo values in $\mathcal{G}_{\mu_i}$ ($\mathcal{G}_{\text{dir}_i}$)

3. Mean of grid element echo values in $\mathcal{G}^c_{\mu_i}$ ($\mathcal{G}^c_{\text{dir}_i}$)
4. Item 2 divided by Item 3
5. Maximum echo value of grid elements in $\mathcal{G}_{\mu_i}$ ($\mathcal{G}_{\text{dir}_i}$)
6. Maximum echo value of grid elements in $\mathcal{G}^c_{\mu_i}$ ($\mathcal{G}^c_{\text{dir}_i}$)
7. Item 5 divided by Item 6
8. Standard deviation of echo values of grid elements in $\mathcal{G}_{\mu_i}$ ($\mathcal{G}^c_{\text{dir}_i}$)
9. Number of grid element values in $\mathcal{G}_{\mu_i}$ ($\mathcal{G}_{\text{dir}_i}$) at each binned echo top value
10. K nearest neighbor label based on Euclidean distance between the above features for each observation

The mean and maximum values along the reroute and direct route give an indication of the severity of weather along those paths. The ratios of these values with the values calculated for the complement of these sets gives an indication of the conditions along the paths compared to conditions in the rest of the NAS. The number of grid element values along the reroute and direct route provides an indication of the convective weather coverage of the path. The K nearest neighbors label is used to provide a rough guess at the class label of the observation based on the similarity of the weather features in this observation compared to observations in the training set.

C. Model Development

An ensemble model was built for each route cluster individually. Before describing the procedure used to create the ensemble model, we need to introduce several terms. The number of weak learners used in each ensemble model is M. The set of all observations is denoted by $\mathcal{S}$. The set of all calculated features is denoted by $\mathcal{F}$. The vector x_i consists of the feature values of observation i. Let $x_i^{\mathcal{F}_m}$ denote the features in $\mathcal{F}_m \subseteq \mathcal{F}$ of observation i. And finally, y_i is the class label of observation i. Model development was performed for each route cluster using the following procedure.

1: Divide $\mathcal{S}$ into $\mathcal{S}_{\text{train}}$ and $\mathcal{S}_{\text{test}}$ sets based on dates of observations
2: $\mathcal{P}_{\text{train}} = \{i \in \mathcal{S}_{\text{train}} | y_i = +\}$
3: Generate synthetic positive observations from $\mathcal{P}_{\text{train}}$ using SMOTE
4: $\mathcal{S}'_{\text{train}} = \mathcal{S}_{\text{train}} \cup \{\text{synthetic positive observations}\}$
5: **for** $m = 1$ to M **do**
6: Subdivide the training data $\mathcal{S}'_{\text{train}}$ into $\mathcal{S}'_{\text{sub-train}}$ and $\mathcal{S}'_{\text{sub-test}}$ by days
7: $\mathcal{P}'_{\text{sub-train}} = \{i \in \mathcal{S}'_{\text{sub-train}} | y_i = +\}$
8: $\mathcal{N}'_{\text{sub-train}} = \{i \in \mathcal{S}'_{\text{sub-train}} | y_i = -\}$
9: $\mathcal{S}_{\text{sub-test}} = \{i | i \in \mathcal{S}'_{\text{sub-test}} \cap \mathcal{S}\}$
10: Create $\mathcal{S}'_{\text{mod-train}}$ by selecting p observations from $\mathcal{P}'_{\text{sub-train}}$ and n observations from $\mathcal{N}'_{\text{sub-train}}$, uniformly at random
11: Create $\mathcal{F}_{\text{m}}$ by selecting f features uniformly at random from $\mathcal{F}$
12: Build a decision tree, $C_m(\cdot)$ using observations $\mathcal{S}'_{\text{mod-train}}$ and features $\mathcal{F}_{\text{m}}$
13: Calculate the error of the prediction results, $\varepsilon_m = \frac{1}{N} \sum_{j \in \mathcal{S}_{\text{sub-test}}} \delta\left(C_m(x_j^{\mathcal{F}_{\text{m}}}) \neq y_j\right)$ where $N = |\mathcal{S}_{\text{sub-test}}|$ and $\delta(\cdot)$ is 1 if the argument is true, and 0 otherwise
14: If $\varepsilon_m < 0.5$, set the importance factor α_m according to $\alpha_m = \frac{1}{2} \ln \frac{1-\varepsilon_m}{\varepsilon_m}$, otherwise $\alpha_m = 0$
15: **end for**
16: Predict the class of observation i for some $i \in \mathcal{S}_{\text{test}}$ according to $C^*(x_i) = \text{argmax}_y \sum_{m=1}^{M} \alpha_m \delta(C_m(x_i^{\mathcal{F}_m}) = y)$

Each of the weak learners exhibits poor performance. The predictions from these models are given a weight α based on their respective prediction errors ε. If the prediction error is less than 0.5, α is calculated using Step 14. The particular choice of α comes from a formulation of the adaptive boost algorithm.[9] This function of ε has a value of 0 at $\varepsilon = 1$ and increases as ε decreases. Using this weighting function, weak learners with low error are given high weights. The ensemble prediction for the test data set is a weighted vote from each of the weak learners.