Prediction of Terminal-Area Weather Penetration Based on Operational Factors

by

Yi-Hsin Lin

B.S., Massachusetts Institute of Technology (2009)

Submitted to the Department of Civil and Environmental Engineering in partial fulfillment of the requirements for the degree of

Master of Science in Transportation

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2013

Author ....
Department of Civil and Environmental Engineering
May 17, 2013

Certified by ....
Hamsa Balakrishnan
Associate Professor of Aeronautics and Astronautics
Thesis Supervisor

Accepted by ....
Heidi M. Nepf
Chair, Departmental Committee for Graduate Students

Prediction of Terminal-Area Weather Penetration Based on Operational Factors

by

Yi-Hsin Lin

Submitted to the Department of Civil and Environmental Engineering
on May 17, 2013, in partial fulfillment of the
requirements for the degree of
Master of Science in Transportation

Abstract

As demand for air transportation grows, the existing air traffic control system is being pushed to capacity. This is especially true during weather events. However, the degree to which weather impacts airspace capacity, particularly within the terminal region, is not well understood. Understanding how weather impacts terminal area capacity will be important for quantifying the uncertainty inherent in weather forecasting and developing an optimal mitigation strategy.

In this thesis, we identify and analyze operational features that may impact whether a pilot chooses to fly through severe weather. In doing so we build upon the work done at MIT Lincoln Laboratory on terminal area Weather Avoidance Fields (WAF) for arriving aircraft. This model predicts the probability of pilot deviation around weather, based solely on weather features. The terminal area WAF was calibrated based on historical pilot behavior during weather encounters near the destination airport. Our model extends the WAF by incorporating operational factors such as prior delays and existing congestion in the terminal airspace. Instead of predicting the probability of deviation, our model will predict the maximum WAF level penetrated by the pilot, using the operational features as input. The thesis combines predictive modeling with case studies to identify relevant features and determine their predictive skill.

An understanding of how operational factors impact weather avoidance will allow researchers to better quantify weather forecasting uncertainty and to understand when precision in forecasting is important. In turn, this will improve our ability to find optimal strategies for delay mitigation.

Thesis Supervisor: Hamsa Balakrishnan
Title: Associate Professor of Aeronautics and Astronautics

Acknowledgments

I would like to thank my advisor, Professor Hamsa Balakrishnan, for her support and guidance throughout this project. Thanks also to Rich DeLaura and Mike Matthews at MIT Lincoln Laboratory for their advice and insights and for providing the weather data used in this thesis. Thanks also to my fellow ICAT students and alums, particularly Diana, Yannis, Lanie, and Harshad, for their help at various stages of the project.

I am grateful to my friends for their ongoing support during my time at MIT. Thanks to Sunny, Simmy, Eric, Niki, and many others for listening to me ramble about aviation, technical support, shared meals, and so much more. Thanks also to Bill Cutter and all my friends and colleagues in the MIT Music Department for keeping me sane through stressful times.

Finally, I would like to thank my family for encouraging me to pursue a graduate degree, and for cheering me on throughout the process.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the last decade, high demand for air travel in the United States has pushed the capacity of the National Airspace System (NAS) to its limits. As a result, any reduction in capacity due to weather or other unforeseen circumstances can result in significant delays. This is especially true of summer convective weather, which can grow and decay rapidly and is difficult to forecast.

Although it is clear that convective weather reduces airspace capacity, the degree to which capacity is reduced as a result of weather is not well understood. While there has been significant research into the types of weather that cause reroutes and the effect of convective weather on controller workload, these studies typically treat all flights as equal. It is known that while pilots typically avoid severe weather, some pilots do penetrate severe weather cells, both enroute and within the terminal area. This thesis explores operational factors that may differentiate pilot behavior, focusing primarily on weather penetration behavior within the terminal area.

1.1 Background

In this thesis, we rely heavily on work previously done at MIT Lincoln Laboratory on convective weather avoidance. Specifically, this thesis indirectly builds on the Convective Weather Avoidance Models (CWAM) developed at Lincoln in the last seven years [2, 3, 4]. These models produce Weather Avoidance Fields, which identify the areas impenetrable to aircraft as a result of weather.

1.1.1 Convective Weather Avoidance Model (CWAM)

The first CWAM, developed in 2006, analyzed the planned and actual trajectories for approximately 500 enroute weather encounters in Indianapolis center (ZID). Weather indicators for this study were derived from VIL (measure of precipitation intensity), echo tops (storm height), and lightning strike counts. (The first two data sources are discussed in more detail in Chapter 2.) CWAM2, developed in 2008, expanded the dataset to about 2000 flights in ZID, ZOB, and ZDC. Another study in 2010 refined the earlier models to improve detection of non-weather related deviations, such as shortcuts, and further increasing the dataset to about 5000 flights.

All of these models identified the difference in altitude between the flight and the echo top height as the primary determinant of pilot deviation in enroute airspace. In other words, pilots frequently overfly weather. A secondary determinant was the fractional VIL coverage over Level 3 in the vicinity of the trajectory. The exact kernel size varied between different versions of the model; 16- and 60-km kernels are typical sizes.

Once the important indicators have been identified, the model returns the probability of deviation for a pilot encountering a particular type of weather. This probability is based on historical pilot behavior. In this sense, the result of model can be thought of as a probability lookup table: for any given altitude, echo top height, and local VIL coverage, the model stores a probability of deviation. This deviation probability is referred to as the Weather Avoidance Field, and can be computed on a pixel by pixel basis given the echo tops and VIL images for a given time.

In recent years, a version of the CWAM specific to the terminal area has been developed. This version found that due to descent, pilots were not typically able to overfly storms, and that the absolute echo top height was a better indicator than the difference between altitude and echo top height. Furthermore, because there is less room to deviate within the terminal area, a 4-km kernel was found to be optimal.

1.1.2 Weather Avoidance Field (WAF)

To understand how the Weather Avoidance Field identifies areas of weather impenetrable to air traffic, it is instructive to look at an example. Figure 1-1 shows the VIL and echo tops in the ORD terminal area on June 12, 2008, around 20:17:30Z. Figure 1-2 shows the analogous plot of the WAF. The arrival trajectories are color-coded according to the maximum WAF penetrated in the terminal area; all departures are grey. The most obvious effect is that the WAF eliminates much of the light rain that has no little to no effect on aviation. Furthermore, not all VIL pixels of Level 3 or above translate to high WAF: some of the smaller cells have relatively low echo tops, which the CWAM has found to be more commonly penetrated by pilots; these are accordingly assigned a lower WAF.

1.1.3 Comparison with actual weather

A question which naturally arises is whether pilots are actually flying through severe weather when they penetrate high WAFs. Since some low level VIL pixels will correspond to high WAFs simply because of proximity to higher VIL levels, it is possible that pilots flying through high WAFs are not actually penetrating weather at all. Another possibility is that pilots are overflying weather, since the terminal WAF does not account for echo top height relative to altitude.

To check whether this is true, we can plot the distribution of actual VILs penetrated by each pilot, sorted by the WAF value. In other words, for all pilots who flew through a WAF of, say, 70, what VIL levels did this WAF correspond to? Figure 1-3 contains the result of this analysis. The left bar in each pair simply indicates the VIL distribution; the right bar removes those cases where a flight was at or above the echo top height.

Figure 1-3 justifies the use of WAF in this thesis. First of all, we do not have to worry about overflying skewing the results. In over 95% of cases, pilots do not overfly weather in the terminal area. This is likely because they are descending and are not typically high enough to do so. Second, there is a strong correlation between high WAFs and high VIL levels. Pilots who penetrate WAFs of 80 or above have a greater than 80% chance of actually penetrating Level 3 VIL or above. Even pilots who flew through WAFs of 80 or above and only penetrated Level 1 or 2 WAF were necessarily within 4 km of intense weather.

1.1.4 Terminal-area operations

The Weather Avoidance Field encapsulates pilot willingness to penetrate severe weather on the basis of the weather features themselves. However, it is possible that there are operational factors that may influence pilot decision-making. Instead of creating a new version of CWAM examining operational features, this thesis takes a different approach. We

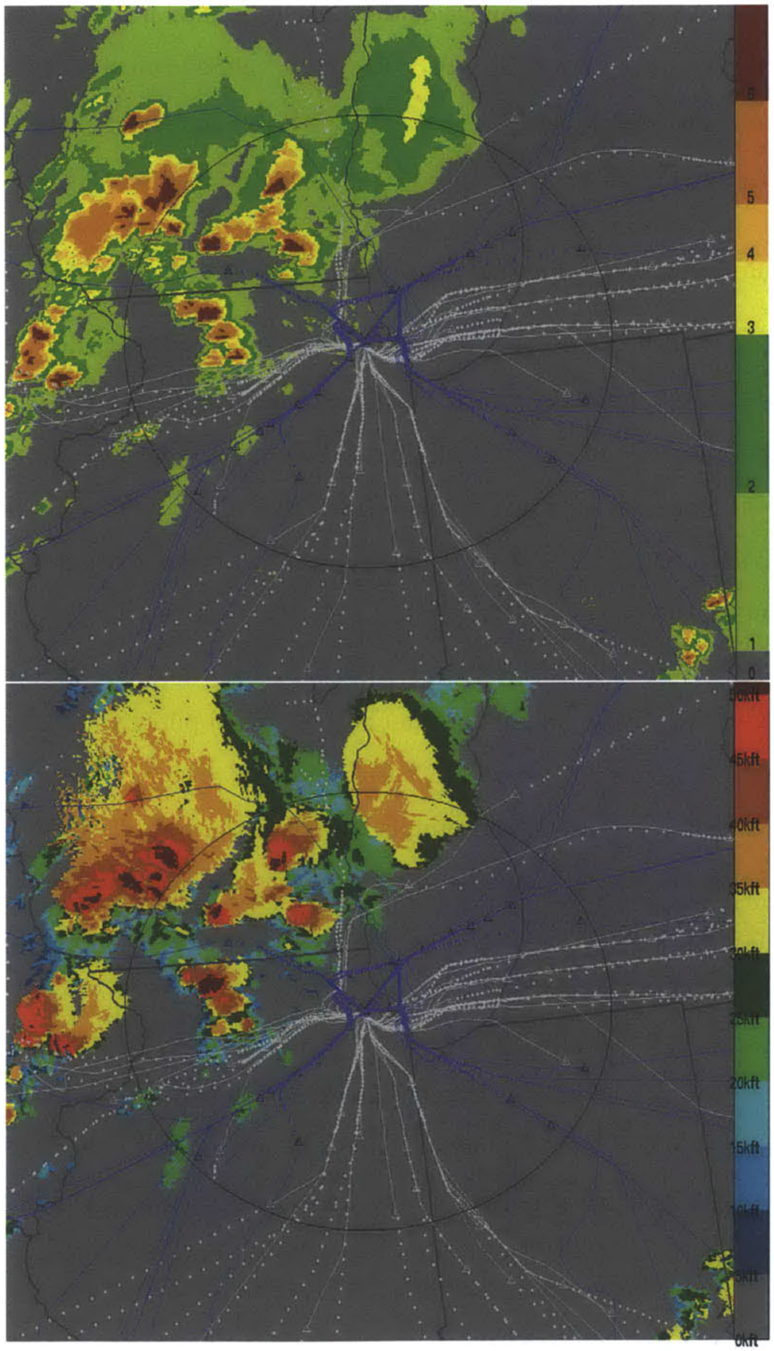

Figure 1-1: VIL and echo tops in the ORD terminal area with overlaid trajectories on June 12, 2008, at 20:17:30Z.

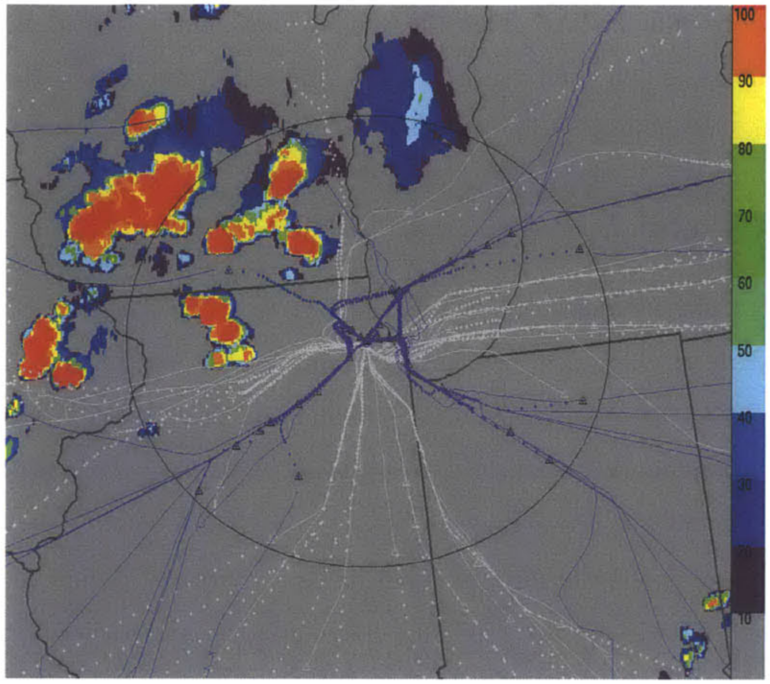

Figure 1-2: WAF in the ORD terminal area with overlaid trajectories on June 12, 2008, at 20:17:30Z.

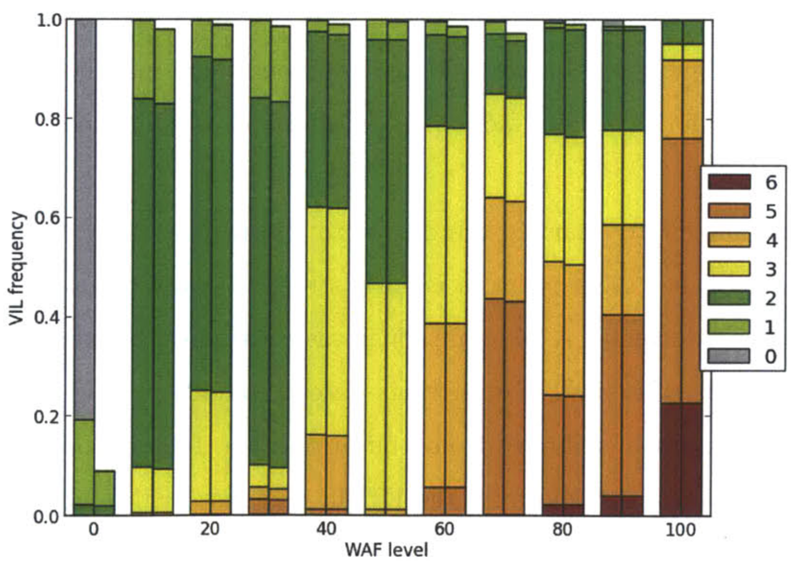

Figure 1-3: VIL distribution for each WAF value. The right bar in each pair removes the flights that were at or above the echo top height.

instead attempt to predict on the basis of operational features when flights will be willing to penetrate regions of high WAF. This allows us to focus on the operational features rather than on determining which combination of VIL and echo tops poses a danger to pilots as this information is already given by the WAF.

Terminal area definition

This thesis specifically focuses on pilot behavior near the arrival airport. This region is not precisely defined. Most major airports have Terminal Radar Approach Control (TRACON) facilities which serve the area immediately surrounding the airport. Using the TRACON boundary is one possible definition. However, TRACONs can vary in size and shape, and a simpler, more general definition is desirable.

In this thesis, we define the terminal area to be the circle of radius 200 km around the airport. Although TRACONs are typically irregularly shaped, a circular region simplifies analysis. To choose the radius, we must consider what characteristics define the terminal area and why pilot behavior in this region might be different from pilot behavior during the enroute portion of the flight. The primary difference is that aircraft trajectories are far more constrained both vertically and horizontally. Enroute pilots are frequently observed to overfly convective weather; a pilot that has already begun its descent sequence may not be able to do this. Furthermore, approach paths are fairly specific, especially at major congested airports such as Chicago O'Hare, and controllers may be less willing to allow pilots to deviate. Both of these factors could affect pilot willingness to penetrate severe weather.

While it is difficult to quantify the degree to which pilots have horizontal latitude to deviate from established flight paths, it is fairly straightforward to determine when arriving aircraft begin their descents. A 200-km radius was chosen as the distance at which aircraft landing at the airport typically begin their descent sequence. Figure 1-4 plots aircraft altitudes as a function of distance from the airport for 100 randomly chosen flights on July 2, 2008. This data is taken from the ETMS database, which will be described in the following chapter. Excluding flights with origins within 300 km, most flights seem to begin their descent sequences between 200 and 250 km away from the airport. (The downward spikes are presumably due to data loss, resulting in an altitude of zero being erroneously recorded in the ETMS database.)

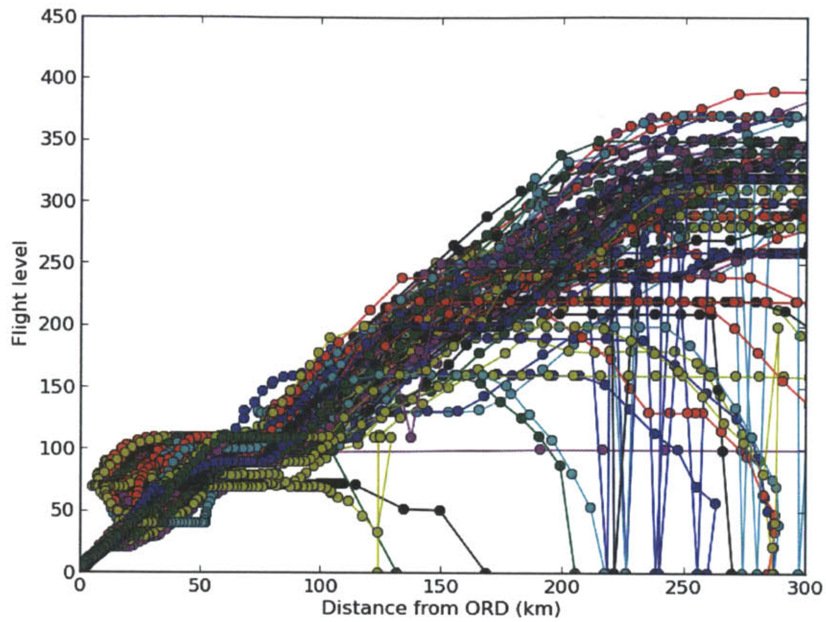

Figure 1-4: Flight altitudes as a function of distance from airport on July 2, 2008.

1.2 Thesis development

The approach taken in this thesis is twofold. Due to the limited number of days with severe weather, a combination of case studies and predictive modeling is used. Chapter 2 discusses the data sources for this study. These include weather data from MIT Lincoln Laboratory for all days in 2008 when Chicago O'Hare was affected by convective weather; 2008 ETMS data from the Volpe National Transportation Center; and ASPM data maintained by the FAA.

Chapter 3 discusses the case studies undertaken, focusing on cases when pilots penetrated severe weather. Commonly observed themes and occurrences are noted and discussed. Along with observations from air traffic controllers and other researchers studying pilot behavior regarding weather, these case studies inform which features were examined and extracted for the predictive model.

Chapters 4 and 5 describe each of the features included in the predictive model. Chapter 4 contains most of the features that can be directly extracted from one of the databases included in this thesis. Chapter 5 discusses trajectory-based features, such as flow analysis and influence of previous pilots.

Chapter 6 describes the two predictive models used in this study. Decision trees were chosen due to their transparency and applicability to small datasets. Random forests were explored as a simple extension to decision trees. Finally, Chapter 7 discusses the implications of the thesis and plans for future work.

Chapter 2

Overview of Databases

Three main data sources were used in this thesis: weather data from Lincoln Laboratory, trajectory data from the Enhanced Traffic Management System (ETMS) database provided by the Volpe National Transportation Center, and flight data from the FAA's Aviation System Performance Metrics (ASPM) database.

2.1 Weather data

Since 1998, the FAA's Aviation Weather Research Program has been seeking ways to integrate and streamline the various competing aviation weather forecast systems into a single cohesive weather forecast covering the entire continental United States (CONUS). MIT Lincoln Laboratory has been at the forefront of this research, along with NCAR, NASA, NWS, and several other research institutions.

A preliminary requirement for this research is to have high-resolution real-time weather data. To this end, much research has been done at Lincoln Lab to integrate sensor data from a network of individual radars, including NEXRAD, TDWR, and Canadian radars. The resulting mosaic is filtered to remove extraneous noise while preserving the weather information. Some motion compensation is also required to avoid radar echoes during fast-moving storms. We will refer to the output of this process as CoSPA, although this name specifically refers to the forecast products that are based on the real-time weather products described here.

The resulting CoSPA images provide a reliable record of the weather as it moves across the continental United States, with very little volatility. They have been shown to be

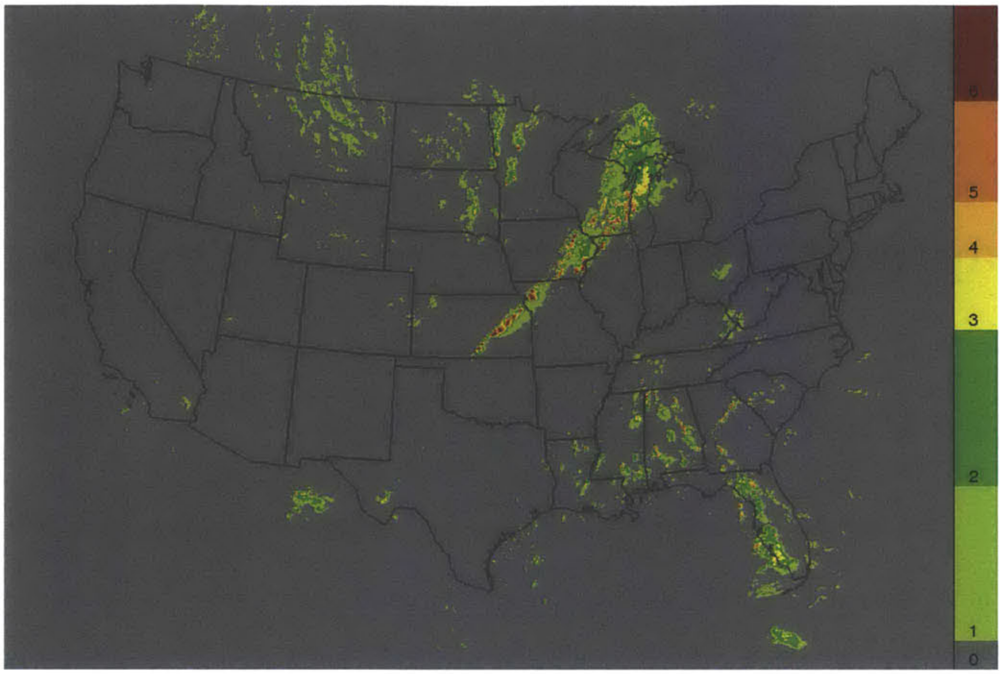

Figure 2-1: An example VIL image from June 13, 2008, at 0000Z.

better at distinguishing between types of severe weather that might be relevant to pilots than previously existing weather products [5]. The images have 1-km spatial resolution, updating every 2.5 minutes. The main CoSPA real time weather products are the Vertically Integrated Liquid and the echo tops. These are the weather inputs considered in this thesis.

2.1.1 Vertically Integrated Liquid

The Vertically Integrated Liquid (VIL) represents the total amount of liquid in a vertical column of the atmosphere. During the image conversion process, the raw reflectivity values returned by the radar are converted to a 0-255 scale, where 255 represents a null value. The VIL indicator has been found to differentiate heavy weather events better than previously used weather variables. For display purposes, this scale is divided into 6 VIL levels; the exact cutoffs are not equally distributed and were chosen to correspond to pilots' perceived threat levels in previously used weather displays. Level 3 VIL corresponds to a 'yellow' threat level; Level 6 corresponds to a 'red' threat level.

Figure 2-1 shows an example VIL image from June 13, 2008, at 0000Z. The weather at this particular time provides a useful demonstration of the common types of weather encountered in the US. First, there are intense, scattered thunderstorms in the Southeast.

Second, there is a large-scale line developing over the Midwest. Note the contrast between the large-scale line developing over the Midwest and the scattered cells in the Southeast. Both of these contain cells of severe weather, but the overall texture is quite different. The thunderstorms in the Southeast are associated with summer convection. These tend to be isolated, smaller cells that grow and decay quickly and are very difficult to forecast. In contrast, the line of storms across the Midwest is associated with a cold front. Both of these weather types contain cells of Level 5 and 6 VIL and can be severely disruptive to aviation. Finally, there is scattered light rain across the Northwest; this consists mostly of Level 1 VIL and is very unlikely to affect aviation. The VIL thresholds were chosen to make these distinctions readily apparent.

2.1.2 Echo tops

While the VIL gives a good representation of where severe weather cells are located in the horizontal plane, it does not provide any information on where weather is located within the vertical column. This is crucial since it is well known that pilots flying at sufficiently high altitudes can simply overfly even very severe weather cells.

The echo tops indicator was developed in direct response to this observation. The echo tops indicates roughly the maximum height of the clouds containing the weather. Note that the echo top images tell us nothing about the minimum height of weather. If an aircraft's altitude is greater than the echo top height, we can be certain that they are flying above the weather. If the aircraft's altitude is below the echo top height, it is reasonable to assume that the pilot is flying through the weather, though in rare cases it is possible that the pilot is flying below the weather cells.

Figure 2-2 shows an example echo tops image from the same time as the VIL image above, June 13, 2008, at 0000Z. Comparing this image to the VIL image above, it is apparent that the highest echo tops generally correspond to the areas with high VIL. This is because stronger convective cells typically extend higher into the atmosphere. However, there are occasionally intense storms that occur lower in the atmospheric column that enroute pilots at altitudes of 40-50 kft can easily overfly; as such, these storms may pose little or no disruption to enroute traffic, but may cause problems for descending or ascending aircraft.

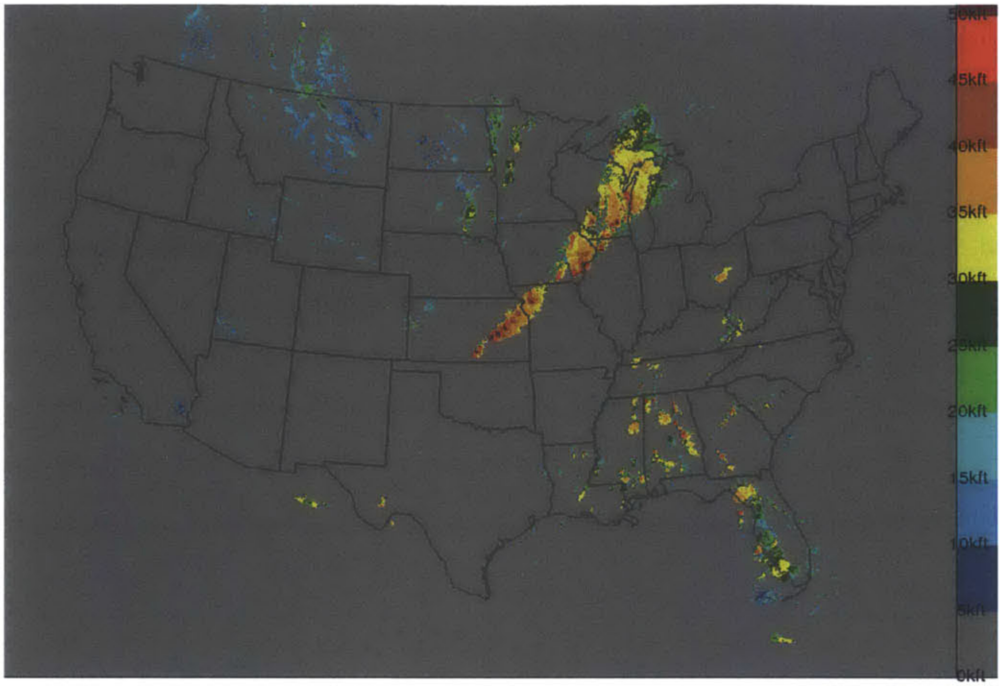

Figure 2-2: An example echo tops image from June 13, 2008, at 0000Z.

2.1.3 Lambert azimuthal projection

The projection used for all of the weather data used in this thesis is the Lambert azimuthal equal area projection. The scripts for converting between latitude/longitude points and the grid points in each weather image were also provided by MIT Lincoln Laboratory. This projection preserves area within contours, but does not preserve angles. Angular distortions are minimal at the center point and increase with distance away from the center point. The center point for the particular projection used in the aviation weather system is (38N, 98W), a point in central Kansas. This point was chosen to minimize distortion over the entire CONUS. Since our study primarily concerns Chicago O'Hare International Airport (ORD) at (41.98N, 87.90W), it is important to keep in mind that the angular and distance distortions there are not insignificant.

The distortion due to the Lambert projection can be seen in Figure 2-3. The small blue circle in the center indicates the location of Chicago O'Hare International Airport. The state outlines are shown in the background for reference. The two rings indicate radii of 50 and 100 km, respectively, in the grid projection. The two crossing lines indicate lines of constant latitude and longitude across two degrees. Due to the angular distortion in the

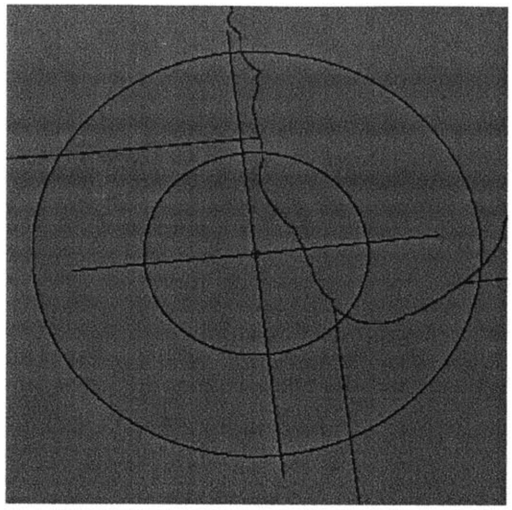

Figure 2-3: Local distortions due to the Lambert azimuthal equal area projection near Chicago O'Hare Airport.

Table 2.1: List of Chicago O'Hare case days from summer 2008.

Date	Start	End
2008-06-12	2008-06-12 10:35	2008-06-14 04:00
2008-06-25	2008-06-25 10:40	2008-06-26 00:25
2008-07-02	2008-07-02 00:10	2008-07-03 07:30
2008-07-07	2008-07-07 10:00	2008-07-09 06:00
2008-07-10	2008-07-10 10:00	2008-07-11 04:00
2008-08-04	2008-08-04 10:15	2008-08-04 20:35
2008-08-22	2008-08-22 10:00	2008-08-23 00:10

projection, these lines are not precisely horizontal and vertical relative to the grid imposed by the projection. This should be kept in mind throughout the thesis, particularly in the discussion of approach angles in Chapter 4.

2.1.4 Case days

The weather archives for June, July, and August 2008 were reviewed. All of the days with severe weather lasting more than one hour in the Chicago O'Hare terminal area were included in the dataset provided by MIT Lincoln Laboratory. Overall, in summer 2008 there were seven weather impacts over eight days. These are outlined in Table 2.1. All times are given in UTC (Chicago's local time zone in summer is UTC-0500).

2.2 ETMS database

The Enhanced Traffic Management System (ETMS) provides trajectory data on a flight-by-flight basis. This data is derived from the Aircraft Situation Display to Industry (ASDI) feed provided by the FAA to various other sites. The data is automatically generated by transponders on each aircraft and sent as real-time messages to the ASDI feed. The ETMS database consists of two main tables. The first contains basic information about each flight, and the second contains the message history sent by the transponder, if any exist. The ETMS flight database includes the flight ID (airline code and flight number for commercial flights, tail number for general aviation flights), scheduled, planned, and actual origin and destination airports, scheduled, planned, and actual departure and arrival times, and aircraft type. ETMS assigns a flight key to each flight, and uses this key to link the flight data to the messages associated with each flight. Positional messages include the message time, latitude, longitude, altitude, and current center. ETMS then derives an average speed from the position and time data. Messages are sent approximately once a minute during the enroute portion of the flight, and approximately once every 15-20 seconds in the ascent and descent portions of the flight. All times in the ETMS database are given in UTC.

All ETMS data from calendar year 2008 was provided by Volpe National Transportation Center for this thesis.

2.2.1 Processing of ETMS data

Filling in missing flight data

Much of the flight data in the ETMS database is incomplete. In the data received from Volpe, many of the fields in the flight database contain null values for many or even most flights. Since this data is derived from the ASDI feed, it is unclear whether there is a data transfer problem or if the data was simply never reported by the airlines. In many of these cases, the missing data is irrelevant to this thesis, and thus does not pose a problem.

The most significant examples that do become relevant are the six fields for origin and destination airports. The ETMS flight database has fields for scheduled, planned, and actual origin and destination airports, but it is frequently the case that only one or two of the three departure and arrival fields are filled in. In these cases, we assume that if no two existing arrival or departure fields directly contradict one another, then no unusual circumstances such as a diversion or a departure from an unscheduled airport took place, and populate the remaining fields accordingly. A diversion to a busy airport like ORD during a weather event would be quite unusual in any case, since it is far more likely that a flight unable to land at a large airport with reduced capacity would divert to a smaller airport. We do not find any cases of diversions to ORD from a different airport during the time periods studied, though several flights depart ORD and divert back to the airport.

Defining dataset

Each flight in the ETMS database is uniquely identified by the combination of flight key and flight date. In practice, it appears that flights are nearly uniquely identified solely by flight key, but they are occasionally reused on different days. The flight date generally appears to be when the flight plan was first filed for that flight, causing it to appear in the ASDI feed. In most cases this is the same day as the departure date of the flight, but in rare cases the flight date can be one or even two days before the flight actually departs.

Since the full ETMS database is very large (over 300GB), a subset was selected and stored in a separate table for faster queries. This subset consisted of all flights actually landing at Chicago O'Hare after filling in missing destination airports as described in the previous section with one of the flight dates listed in Table 2.2. These are simply all of the dates included in the weather dataset, though the flights are not restricted to the precise times for which we have weather data. This set of 21,500 flights will be referred to as the ETMS flightset, and is a superset of the flights analyzed later in thesis.

2.2.2 Verifying ETMS trajectory data

Missing trajectories

Unfortunately, the message database recording each aircraft's trajectory is incomplete or nonexistent for approximately 17% of the flights in the ETMS flightset. The database file provided by Volpe contain all of the ETMS data for the entire year; we therefore assume that this data was simply not reported. ETMS position data is not required by the FAA; most general aviation aircraft are not outfitted with the appropriate equipment. It does not seem to be specific time intervals that are missing; rather, approximately the same fraction of flights seems to be missing at all times, with some variation.

While only a small fraction of flights in the ETMS flightset are general aviation aircraft, the most plausible explanation is that many commercially flown aircraft are also not outfitted with ETMS equipment, and that the precise percentage varies by airline and by aircraft type. This is consistent with the distribution of missing flights across airlines, aircraft type, and origin airports.

Figure 2-4 contains normalized histograms for these three factors. These are separately normalized within each category. Thus, a red bar of same height as a blue bar indicates that the ratio between the categories is precisely the mean. A higher red bar indicates that more flights are missing than average; a higher blue bar indicates the opposite. These charts show no clear correlations between missing data and any of these factors, though there are variations. For example, the data would imply that MD80s are not outfitted with ETMS equipment; this seems perfectly plausible given that the MD80 is an older aircraft model. International flights do not seem overall more or less likely to be missing trajectory data.

While the ultimate cause of these missing flights is unclear, there appear not to be any overriding factors that would significantly bias the dataset in one way or another. As such, flights with missing trajectories are simply discarded from the ETMS flightset.

Trajectories not completed at ORD

Trajectories are also verified to end near the destination airport. The distance from Chicago O'Hare to the last latitude/longitude position reported by the aircraft transponder is computed; the trajectory is discarded if this distance is greater than 10. In most cases the exact threshold is irrelevant; the verification detects spurious flights rather than minor positioning problems. The vast majority of flights arriving at ORD have final positions within 1 km of the airport. Trajectory continuity is verified in the same step to ensure that there are no significant gaps in the messages once the flight has entered the terminal area.

Spurious trajectories are far less common than missing trajectories. Only 267 flights were discarded from the database in this step of the process.

Diversions back to ORD

Of the 21,500 flights in the ETMS flightset, 22 were flights that departed from ORD and then were diverted back to ORD for some reason. These flights were also excluded from

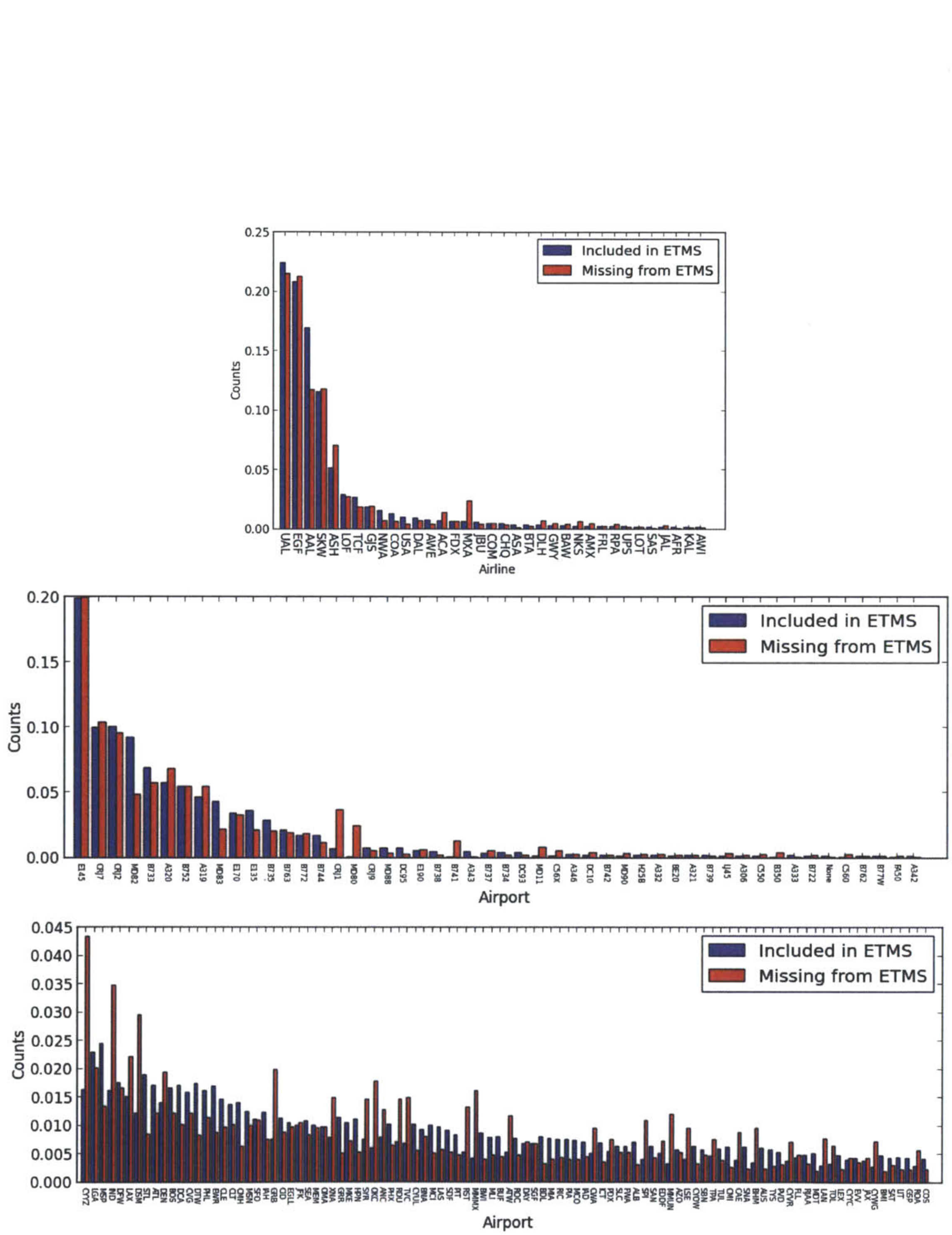

Figure 2-4: Flights with missing ETMS trajectory data by airline, aircraft type, and origin airport.

Table 2.2: List of flight dates included in the ETMS-ASPM flightset.

Date	All	Verified in ETMS	With ASPM match	
2008-06-12	1522	1263	1224	1125
2008-06-13	1634	1191	1155	1134
2008-06-14	1320	1179	1137	42
2008-06-25	1623	1338	1297	983
2008-06-26	1463	1225	1190	0
2008-07-02	1542	1268	1243	1224
2008-07-03	1459	1317	1290	86
2008-07-07	1636	1326	1289	1191
2008-07-08	1484	1324	1309	1294
2008-07-09	1468	1314	1288	76
2008-07-10	1605	1332	1292	1070
2008-07-11	1542	1259	1221	35
2008-08-04	1698	918	873	411
2008-08-22	1504	1307	1283	994
Total	21500	17561	17091	9665

our study. The primary reason is that some of the features examined do not necessarily make sense for a flight departing and arriving at the same airport. Although categorically excluding certain flights could bias the results, the small number of flights in question is unlikely to have a significant effect. Furthermore, while in principle diverting flights might behave differently than regular flights, a perusal of these in particular indicates that most of them diverted within an hour of take-off, and only one flew through any kind of weather. Therefore, these ORD "arrivals" indicate irregular operations and it is reasonable to exclude them from the database.

We are left with 17,561 flights in the ETMS flightset. These are summarized in Table 2.2.

2.3 ASPM database

The Aviation System Performance Metrics (ASPM) database is maintained by the FAA and provides detailed flight data for 77 airports and 22 carriers in the US. This includes Chicago O'Hare. The flight data includes the flight number, scheduled and actual departure and arrival times (in local times), various delay metrics including pushback, wheels-off, wheels-on, and gate arrival delays, and aircraft data. It includes a single departure airport and a single arrival airport; these are assumed to be the actual departure and arrival airports.

The primary reason the ASPM database is used in this study is that it provides more information than the ETMS database about each flight. To do this it is first necessary to match each flight in the ETMS flightset to a flight in the ASPM database.

2.3.1 Matching flights with the ETMS database

Flights from the ETMS flightset were matched with flights in the ASPM database based on arrival day, FAA carrier (airline), flight number, and arrival airport (ORD). The departure airport and approximate arrival time was also verified; flights were more than a two-hour deviation from the ETMS arrival time (the last message time) were verified by hand. Because the ASPM database records most times as local times, it is easier to check arrival times rather than departure times.

Of the 17,561 flights in the ETMS flightset that were not eliminated due to problems with the ETMS data, 17,091 of these were successfully matched to flights in the ASPM database. This set of 17,091 flights will be referred to as the ETMS-ASPM flightset. Finally, flights outside the case periods defined in Table 2.1 are eliminated from the flightset. The variable numbers of flights are mostly explained by how much of each day is included in the study. The flight counts for each flight date are summarized in Table 2.2.

Chapter 3

Case Studies

Case studies have been extremely helpful in guiding the development of this thesis, both for understanding the evolution of weather throughout the day and how this affects the terminal, and for identifying features that affect pilot behavior. This chapter is divided into two sections. The first section charts the evolution of weather and its effects on traffic flows in the Chicago O'Hare terminal area throughout the day on an arbitrarily chosen day from the case set. The second section describes recurring "themes" that are frequently observed in the case studies.

3.1 Case: July 2-3, 2008

This section describes one of the eight case days in our dataset, July 2-3, 2008. It is interleaved with snapshots of the WAF in the terminal area throughout the day. (Recall that this model converts weather features into a single probability of pilot deviation, and thus removes the need to examine multiple weather images at each time period.) This will hopefully provide the reader with a better understanding of how weather affects arrival traffic and some of the measures the FAA can take to mitigate the impact.

Most of our weather cases concern summer convective weather. Convective storms are common in many parts of the United States, particularly the southeast, and typically form in the afternoon. The ground re-radiates heat absorbed from the sun, warming the air layer immediately above and causing it to rise. As these air masses rise, excess water vapor condenses, forming clouds. More energetic masses tend to rise higher, which is why high VILs and high echo tops are generally correlated. Convective storms are associated with heavy rain and severe turbulence. Because they grow and decay relatively quickly, they can be difficult to forecast.

Air traffic managers have a variety of tools they can use to mitigate weather and other delays. In addition to rerouting, there are two programs that are frequently used to reduce incoming traffic to manageable levels; namely, ground delay programs and ground stops. A ground delay program is frequently used when arrivals to a particular airport need to be throttled for any reason; a GDP delays all flights that are still on the ground at the origin airport at the time that the GDP is in effect. Depending on the severity of the bottleneck, GDPs can be set to assign higher or lower average delays. If the delays are severe, a ground stop is sometimes used. These hold all flights bound for the specified airport still on the ground at the origin airport indefinitely, until the ground stop has been lifted. Unlike GDPs, which are supposed to be planned in advance, ground stops are often used as in a more reactionary way once congestion in the terminal area has becomes very severe.

3.1.1 Overview of case day

As is fairly common during the summer, convective weather affected the terminal area from about 15Z (late morning local time) until about 3Z the next day (late evening local time). Figure 3-1 gives a brief overview of the case day. The bars indicate the number of flights landing at the airport at any given time. They are colored according to the maximum WAF penetrated by the aircraft within the terminal area. (Grey indicates that the aircraft did not fly through any weather at all.) The red line indicates the percentage of the terminal area containing WAFs of 80 or greater; this is roughly the severe weather coverage in the terminal area.

Generally speaking, as the amount of weather in the terminal area increases, the total number of flights drops. During the peak of the weather impact, from about 22Z-0Z, traffic drops quite severely. Furthermore, most of these flights penetrate severe weather. While this drop can be attributed to the ground delay and ground stop programs put into place at ORD on the day in question, the relatively large fraction of flights that penetrate weather during the convective weather event indicate that the airspace was likely at full capacity.

It is not universally true that flights are more likely to penetrate severe weather when there is greater severe weather coverage in the terminal area, though these quantities are obviously correlated. Fewer flights penetrate the lower WAF levels in the early part of

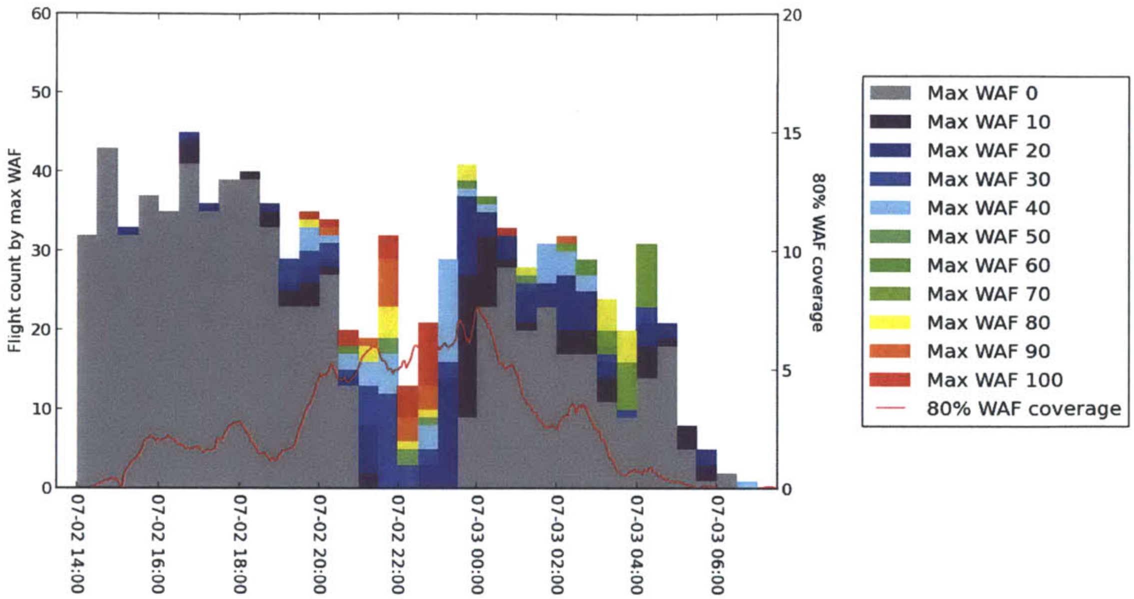

Figure 3-1: Overview of the July 2, 2008, case day. Bars indicate flight counts; the line indicates the fraction of the terminal containing WAFs of 80 or greater.

the day, while the weather is still growing in coverage and intensity. In the latter part of the day, while weather is decaying and moving out of the terminal area, more flights are penetrating high WAFs, even WAFs of 80 and above, despite there being fewer total flights. It is unclear whether this is because of the weather itself (perhaps decaying weather poses less of a threat) or if there are operational reasons that could explain this.

The detailed case analysis will help shed light on how weather and terminal properties interact to determine pilot willingness to penetrate severe weather.

3.1.2 Detailed chronology

On July 2, 2008, Chicago O'Hare was affected by several small lines of convective weather cells which caused severe disruption to air traffic. Weather first enters the terminal area around 14Z (Figure 3-2); however, this first line stays mostly to the north of the airport. While the northbound departure routes out of ORD are affected, the arrival routes to the northwest and northeast are only minimally affected. Although the weather covers a substantial portion of the terminal area and had WAFs of 100, only three arriving pilots penetrate WAFs over 80. In most cases, the weather was simply did not pose an obstacle to standard arrival routes due to its position (Figure 3-3).

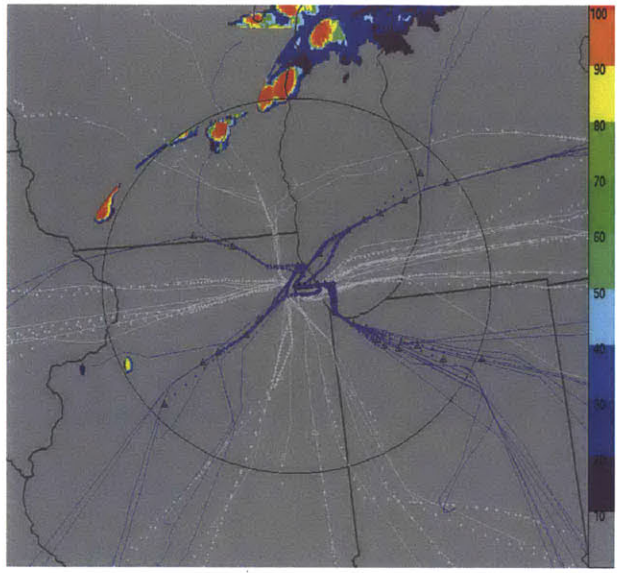

Figure 3-2: ORD terminal area at 15Z.

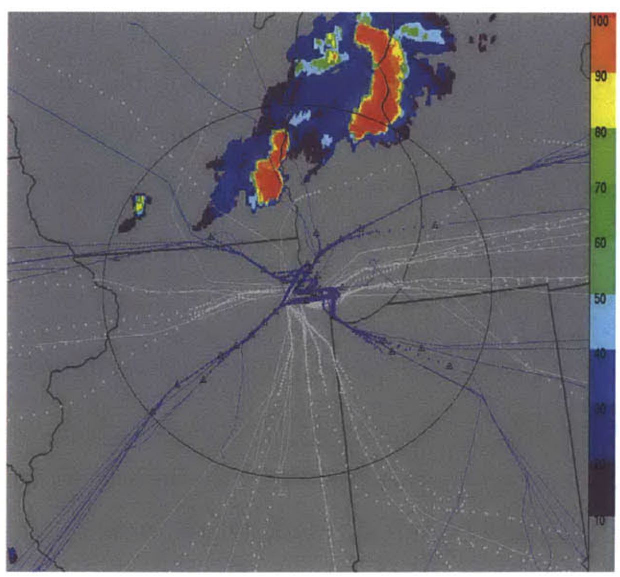

Figure 3-3: ORD terminal area at 17Z.

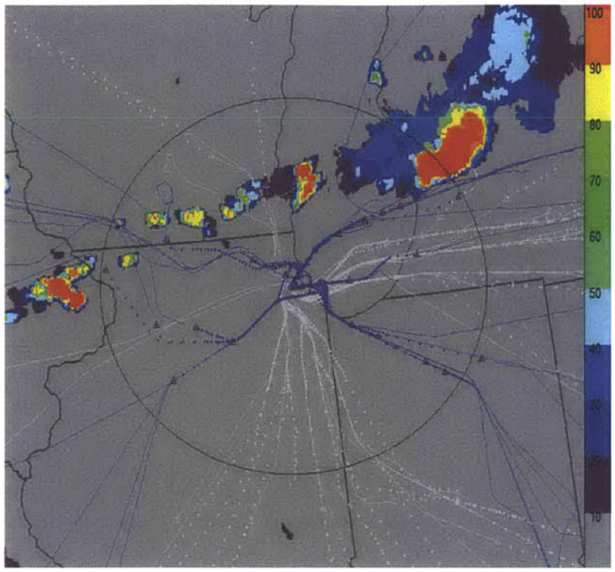

Figure 3-4: ORD terminal area at 19Z.

Air traffic flow managers noted the impending arrival of a second series of convective cells which would pass almost directly over the airport. These cells were less severe than the first line, but would prove to be a far greater disruption due to their location. At 1534Z, well before the arrival of the second series of cells, air traffic managers accordingly planned a ground delay program (GDP). This GDP was scheduled to take effect at 18Z, approximately when the second series of cells would enter the terminal area.

The second series of storms entered the terminal area around 1915Z due west of the terminal area, traveling due east (Figure 3-4). These storms had cells of Level 5 and Level 6 VIL, and correspondingly had WAFs of 100. They also grow in size as they move eastwards towards the airport. Note that the departures streams shift south in response to this weather. Furthermore, because of the ground delay program which took effect at 18Z, the number of flights attempting to land at ORD has begun to decrease. This is evident from the overview shown in Figure 3-1.

At the same time, the first line of weather to the north begins to impact the northeast arrival streams. This streams respond by first shifting south in an attempt to fly around the weather, and then eventually rerouting behind the storm through the lower level WAFs. Although this change in routing is clear, it is unclear to what extent individual pilots are picking their way through breaks in the storm as opposed to being actively vectored through the storm in a continuous flow. As the weather intensified, fewer and fewer pilots approach

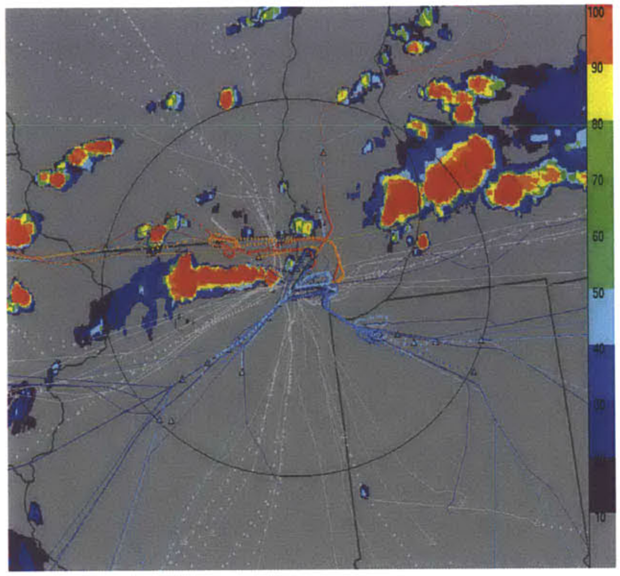

Figure 3-5: ORD terminal area at 21Z.

from the northeast, until this cornerpost is shut down entirely.

Meanwhile, the second line of cells from the west have been gradually approaching the airport, reaching ORD shortly after 21Z (Figure 3-5. This line is smaller in extent than the first, though of similar intensity. Nevertheless, as this storm approached the airport, over twenty pilots penetrated severe weather within the terminal area. This situation quickly becomes untenable, and a ground stop is put into place at 2103Z and extended at 2121Z. It is downgraded to a ground delay program at 2151Z.

These ground stops were no doubt triggered at least in part by the unusually high number of pilots penetrating severe weather. In most cases, it is evident from the images that they simply had no other choice if they wanted to land at ORD, particularly once the storm was directly over the airport as it was around 22Z (Figure 3-6). In some cases, pilots attempted to deviate around a large cell, but perhaps brushed the edge at some point. This behavior is in part because of the need to approach the airport from a specific direction: Aircraft arriving from the west have to go around the storm in order to land on the active runway.

After 22Z, the weather cells begin to decay and move out of the terminal area (Figure 3-7). Nevertheless, various individual pilots continue to penetrate severe weather cells. While some streams are still visible, the overall patterns are loose enough that pilots are likely choosing their own paths through weather. For example, at 01Z (Figure 3-8) we observe

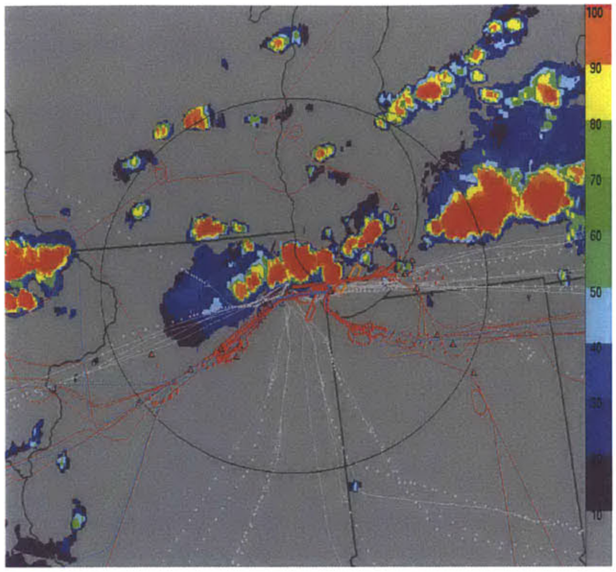

Figure 3-6: ORD terminal area at 22Z.

one pilot who briefly flies through a WAF of 100, even while most of the pilots in the same stream do not. There is clearly a range of risk tolerances in how close pilots are willing to fly to severe weather cells, though it is not clear from this particular case whether it is random.

Even as the weather is decaying, we see that the location of the weather in relation to the airport is important. At 01Z, when the weather is farther from the terminal, relatively few pilots penetrate severe weather. Most of those that do skirt the weather rather than flying straight through it. Yet when the same weather reaches the airport around 03Z (Figure 3-9), pilots who wish to land at the airport again have no choice but to penetrate severe weather in order to do so. This partially explains why there are more severe weather penetrations while the weather is decaying relative to earlier in the decay.

Despite these weather penetrations, by 0240Z the weather in the terminal area has decayed enough that the ground delay program is cancelled early. Furthermore, the decreased levels of traffic mean that each individual flight has greater flexibility to deviate around weather. Given these factors, it is even more surprising that so many flights penetrate WAFs between 40 and 100. It is particularly in periods like this that we would like to determine if there are operational factors that may influence why certain flights penetrate severe weather while others avoid it.

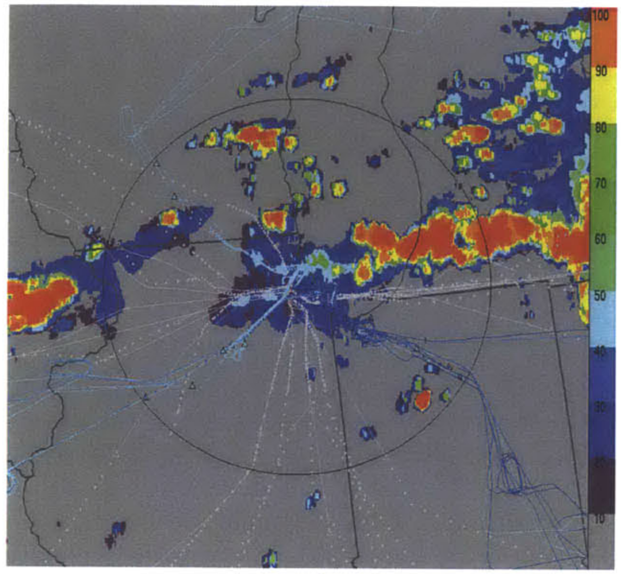

Figure 3-7: ORD terminal area at 23Z on 2008-07-02.

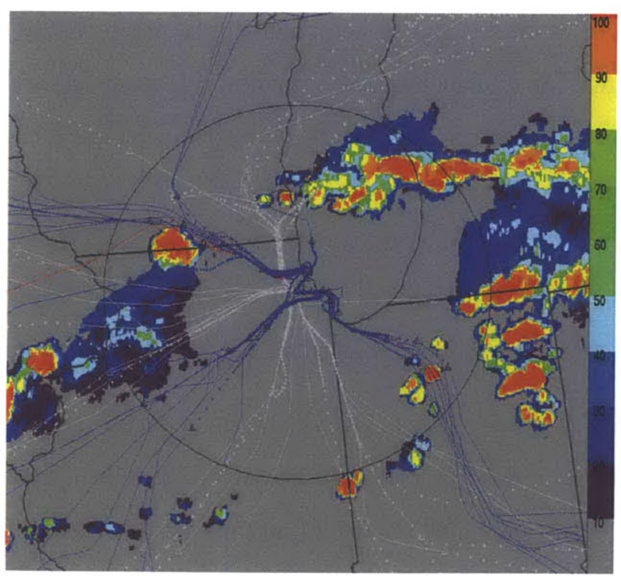

Figure 3-8: ORD terminal area at 01Z on 2008-07-03.

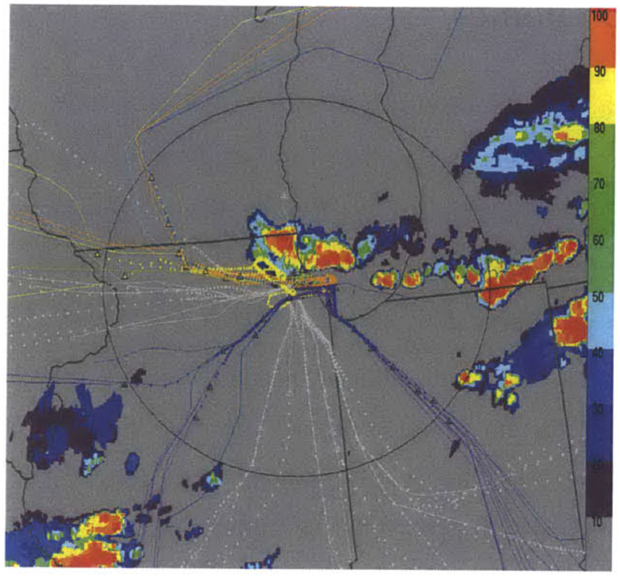

Figure 3-9: ORD terminal area at 03Z.

3.1.3 Case study conclusions

Many factors influence whether a pilot in the terminal area chooses to deviate around weather; indeed, many factors influence whether a pilot will even be in the terminal area at the time of a weather event. First, weather mitigation programs put in place by the FAA may reduce the number of flights into ORD during the worst weather. This would naturally influence the behavior of the remaining pilots, since they may have greater flexibility to deviate. How and when these programs are implemented can give us insight into the capacity of the airspace. For example, the ground stop that was implemented at 2103Z indicates that the situation at that point in time was not sustainable and that air traffic controllers felt the need to reduce the number of flights in the terminal area.

Second, it is clear that proximity of severe weather to the airport is an important factor in determining whether pilots choose to deviate. When the weather is unavoidable, it seems that many pilots are willing to penetrate severe weather, but generally prefer not to otherwise.

Third, we have identified at least one time period in which pilots make a variety of different choices about whether to penetrate severe weather. It is in these cases that we hope operational factors may provide some insight into why some pilots are willing to penetrate severe weather while others are not.

3.2 Recurring themes

Instead of providing a detailed description of all eight case days, we describe representative examples of recurring themes that are commonly observed. This list is certainly not exhaustive but includes observations that have guided the feature analysis. Each of the following plots is a single snapshot in time. The black circle indicates what we define as the terminal area, a circle of radius 200 km around Chicago O'Hare Airport. The solid lines are the trajectories up to that time, and the dotted lines are the future trajectory points. A triangle indicates the current position of each aircraft. Gray trajectories are departures. Arrivals are color coded according to the highest WAF that pilot penetrates within the terminal area. (The color distribution for the trajectories contains slightly different shades of each color relative to the WAF colormap so that they could be distinguished; the color bar to the right is the color bar for the background WAF.)

Pilots fly very close to weather

We see many cases where pilots fly very close to weather, in some cases coming within a few kilometers of very heavy storms. For example, in Figure 3-10 we can see several trajectories passing within a few kilometers of the cell in the northwest quadrant of the terminal area. Note that although these pilots are following very similar trajectories around the weather, one gets close enough to pass through a WAF of 80 (orange), while the others pass through only WAFs below 50 due to only a small deviation in flight path. It is unclear whether this pilot did in fact fly through heavier weather of if the discrepancy in WAFs is due to a timing issue. Nevertheless, it is clear that all of the trajectories are flying very close to severe weather, and that this is an extremely frequent occurrence when severe weather occurs in the terminal area (and enroute, though this is outside the scope of our study).

Pilots take advantage of gaps in weather (and will sometimes pick their way through staggered storms)

It is not uncommon to see pilots flying through lines of severe storms, deviating as necessary to avoid the worst weather. In this example (Figure 3-11), there is actually a gap in the line, but in some cases there is a continuous line of weather with discontinuous heavy cells. In this example, we observed almost no arrivals from the northwest quadrant in a three-

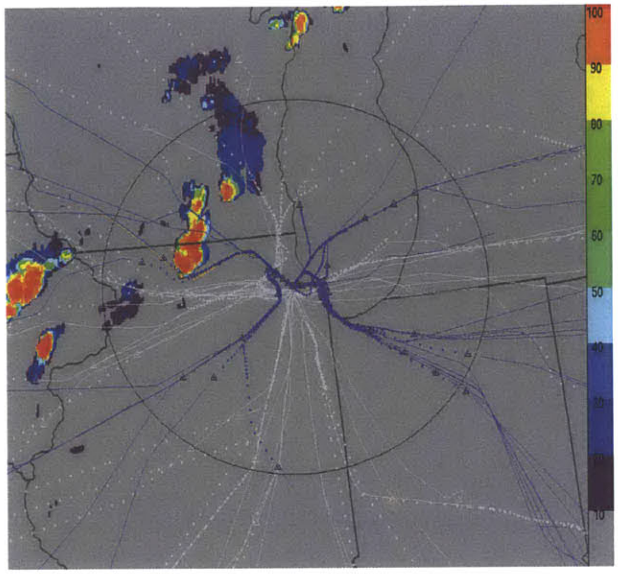

Figure 3-10: Pilots fly very close to weather. Image from 2008-06-12 17:42:30Z.

hour period, except for two episodes where gaps in the weather allowed for a few flights to get through. Pilots or air traffic controllers are able to observe and take advantage of these gaps. Such incidents usually result in pilots flying through middle-range WAFs, but not extremely high WAFs, depending on the particular weather situation.

Weather will perturb arrival paths; pilots eventually begin to fly behind the weather

As weather moves across the terminal area, it's common to see flows bend slightly to avoid flying through the weather. Eventually, pilots begin flying behind the weather rather than in front. For example, the two images in Figure 3-12 are two snapshots of the terminal area 25 minutes apart. The cell in the northwest quadrant has begun to impact the northwest arrival routes, and pilots gradually deviate further and further northeast in order to avoid the weather, which is moving eastwards. Eventually, pilots begin to fly behind the weather. A similar pattern can be observed in the southeast quadrant. It is not apparent from the data whether it is a pilot who prompts the route change or an air traffic controller who directs pilots around the weather.

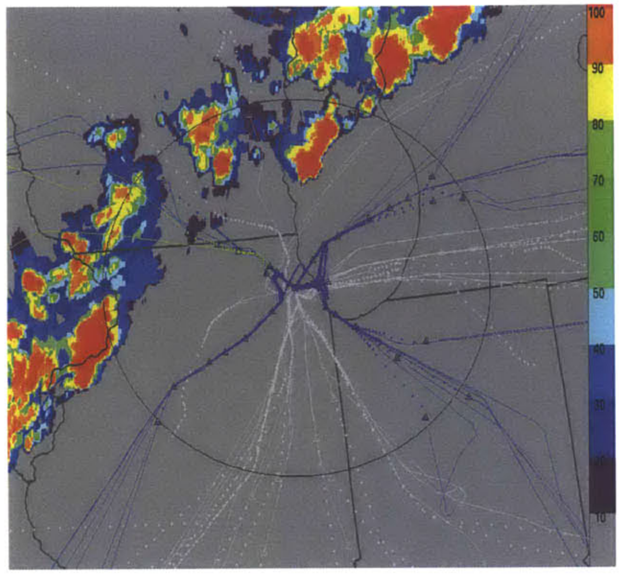

Figure 3-11: Pilots pick their way through gaps in the weather. Image from 2008-06-13 03:07:30Z.

Weather near the airport cannot be avoided

When severe weather is very close to the airport, it is very difficult for the pilot to avoid flying through weather, especially when the weather is along the arrival route. In this July 2 case, we see relatively few severe weather penetrations until the weather moves directly over the arrival paths. When this happens, almost all of the pilots landing are forced to fly through WAFs of 100. As the weather moves off to the east, the number of weather penetrations sharply drops off again.

Significant numbers of severe weather penetrations are not sustainable

Among the seven case days, there are only two periods in which more than a few pilots fly through WAFs over 90, including the July 2 case described in the previous section. Both of these periods result in ground stops and lengthy ground delays, indicating that such activity is unsustainable and undesirable. Although pilots continue to fly through and land in such weather, it is an indication that the airspace capacity is being significantly exceeded.

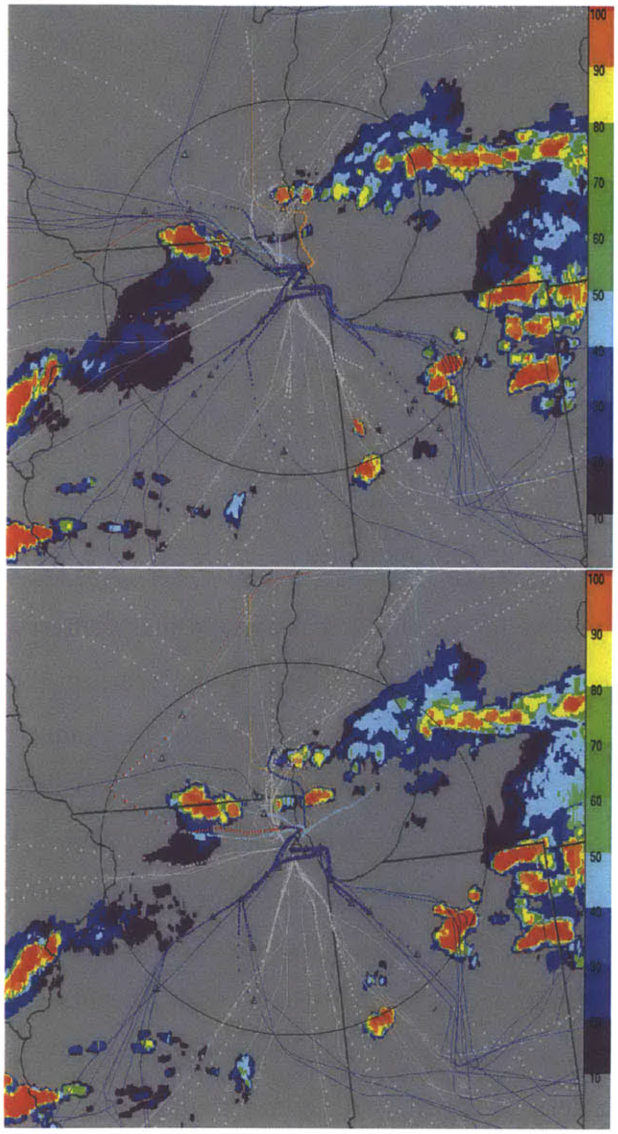

Figure 3-12: Flows become perturbed as weather moves across them; eventually pilots begin to fly behind the weather. Image from 2008-07-03 01:20:00Z and 01:35:00Z.

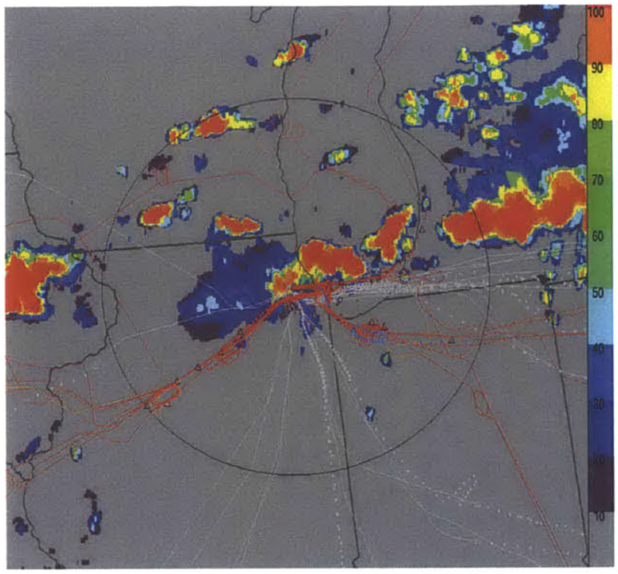

Figure 3-13: Weather very close to the airport cannot be avoided. Image from 2008-07-02 22:20:00Z.

Pilots avoiding weather deviate irregularly; fixed holding patterns are mostly observed during periods of fair weather

Contrary to expectations, we more commonly observed pilots entering into holding patterns inside the terminal area during periods of little to no weather. For example, in the top image of Figure 3-14 we see multiple flights holding due to volume prior to landing at the airport despite there being only mild weather in the terminal area. While congestion is to be expected despite fair weather, it is somewhat surprising that there are relatively few cases of similar holding patterns during heavy weather events. This is likely due to several factors. First, during weather events many flights probably have been delayed or diverted prior to arrival at the terminal area during severe weather events. Second, flights that need to deviate or hold in the terminal area during weather events typically cannot do so in an established holding pattern. Instead, their paths are far more irregular, as shown in the bottom image of Figure 3-14.

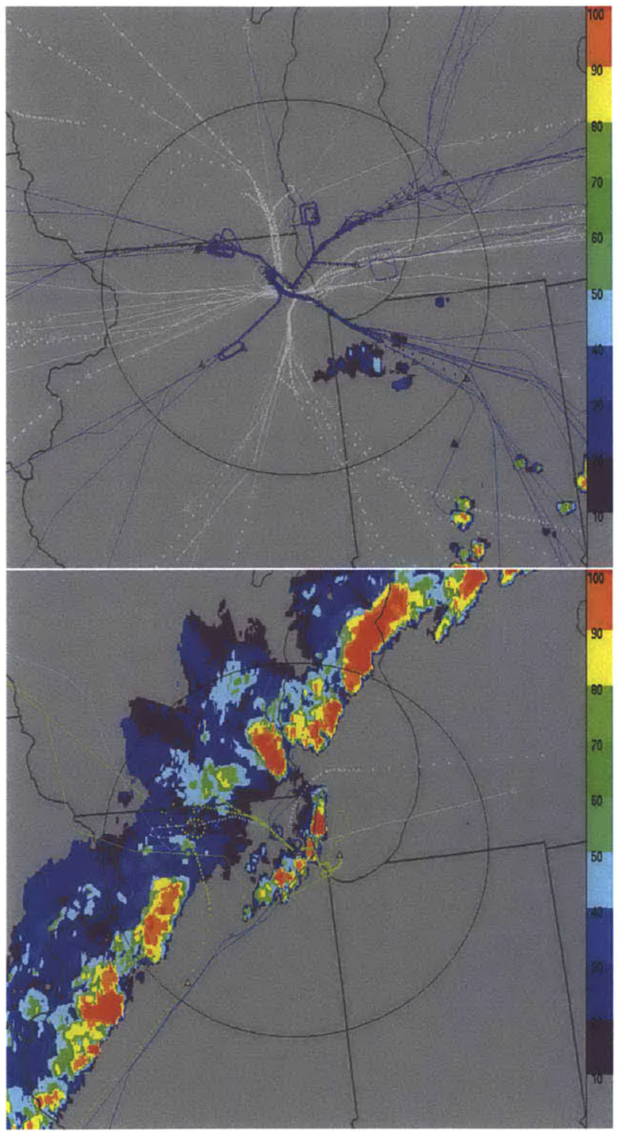

Figure 3-14: During weather events, pilots may deviate irregularly to avoid severe weather, but standard holding patterns are mostly observed during fair weather. Image from 2008-06-25 18:57:30Z and 2008-06-13 06:35:00Z.

Chapter 4

Feature Identification

In this chapter, we describe in detail the features examined for how well they predict pilot penetration of severe weather. (Recall that severe weather penetration is defined as flying through a WAF of 80% or above within 200 km of arriving at Chicago O'Hare.) These features can be divided into five broad categories: on-time performance, properties of the flight, features of the terminal, weather features, and features relating to whether the flight was part of a flow. For each feature, the procedure to extract the feature is described. A normalized histogram showing the distribution of flights that did and did not penetrate severe weather is plotted for each feature. When analyzing these histograms, it is important to keep in mind that the two distributions are normalized indepedently and that there are far fewer flights that penetrate severe weather relative to the total number of flights. Specifically, only 144 flights in the dataset fly through severe weather, compared to over 9500 that do not. Finally, the feature's individual skill and possible explanations for such skill is discussed.

4.1 Flight-based features

The most basic features to include are static features of each flight. These are, generally speaking, easily determined from either the ASPM or ETMS database.

4.1.1 Origin-destination distance

The origin-destination distance for each flight is determined based on the first and last trajectory points. While it might theoretically be more accurate to use the latitudes and longitudes of the origin and destination airports, the ETMS position data is generally more reliable than the airport data. (See Chapter 2 for a more detailed discussion.) Under the assumption that the first recorded position is very close to the origin airport, we consider the origin-destination distance to be the great circle distance between the first and last position points. Note that we previously verified that the last position is within 10 km of ORD, discarding all trajectories where this was not the case. The fact that less than 1 % of flights were discarded in this manner suggests that the position data is reasonably reliable.

Figure 4-1 contains the normalized histograms for pilots that did and did not penetrate severe weather as a function of origin-destination distance. The two plots show the same data with slightly different bins; the second places all flights with OD distance greater than 3000 km into the rightmost bin. The first plot indicates that OD pairs more than 2000 km apart are correlated with severe weather penetration. For reference, San Francisco International Airport (SFO) and Chicago O'Hare (ORD) are approximately 3000 km apart. The flights in the 7000 km range are mostly trans-Pacific flights coming from Asia. One possible explanation is that these longer routes are flown by larger aircraft more able to withstand weather. Another possibility is that midrange flights are disproportionately delayed or canceled during weather events at the destination airport due to ground delay programs or ground stops that do not affect longer range flights that have either already taken off or are not within the scope of the ground delay.

A positive correlation can also be seen in the flights arriving from within 500 km of ORD. This can be more clearly seen in the second plot, which breaks down the data into smaller bins. Recall that we are specifically tracking weather penetration within 200 km of the airport; for many of these flights, it may simply be the case that they do not have sufficient flexibility to avoid the region of the terminal containing severe weather. This is supported by the case studies described in Chapter 3.

4.1.2 Air carrier

While every airline's pilots adheres to safety restrictions and would not subject their flight to dangerous weather, there may be institutional variation in how much turbulence pilots are willing to tolerate. Since turbulence is anecdotally one of the primary factors in determining whether a pilot chooses to fly through severe weather, it is plausible that certain airlines' pilots would be more likely to fly through severe weather.

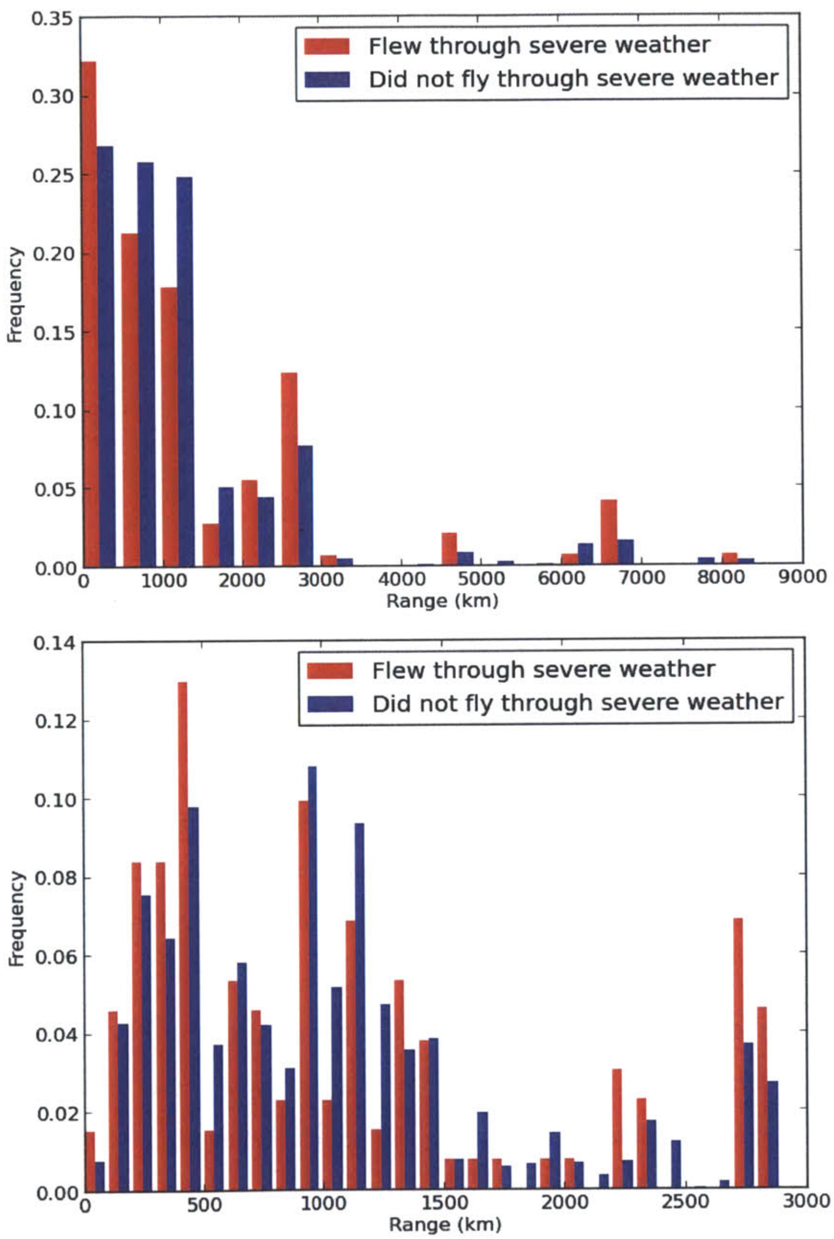

Figure 4-1: Normalized histogram of severe weather penetration by origin-destination distance.

Table 4.1: List of airlines servicing ORD. Airlines with fewer than 100 flights are not considered in the model; airlines with fewer than 25 flights are not listed.

Airline	Code	Flights
United	UAL	2208
American Eagle	EGF	2124
American	AAL	1650
SkyWest	SKW	1116
Mesa	ASH	514
Trans States	LOF	275
Shuttle America	TCF	258
GoJet	GJS	179
Northwest	NWA	149
Continental	COA	130
US Air	USA	95
Delta	DAL	87
Atlantic Coast	ACA	67
American West	AWE	59
Mexicana	MXA	57
JetBlue	JBU	49
Chautauqua	CHQ	46
Comair	COM	37
Atlantic Southeast	ASA	33
Lufthansa	DLH	32
FedEx	FDX	32
ExpressJet	BTA	26

Due to the distribution of airlines servicing Chicago O'Hare, not all airlines appear in the flightset sufficiently often to be statistically significant. As such, only airlines with at least 100 flights in the flightset are included. These are listed along with their identifying codes in Table 4.1.

While there are distinct differences in weather penetration behavior across different airlines, as shown in Figure 4-2, it is unclear how significant these differences are or even whether they are due to airline management. It appears that of the legacy airlines, Continental pilots are the most likely to penetrate severe weather and United pilots are the least likely. (Recall that the flightset is from 2008, well before the United-Continental merger.) Other than American Eagle (EGF), it appears that the other regional airlines are less willing to penetrate severe weather. This makes sense given that they are likely to be flying turboprops and regional jets instead of larger aircraft. American Eagle's different behavior may simply be because their pilots have significantly more experience dealing with ORD specifically, and are more willing to fly through weather as a result.

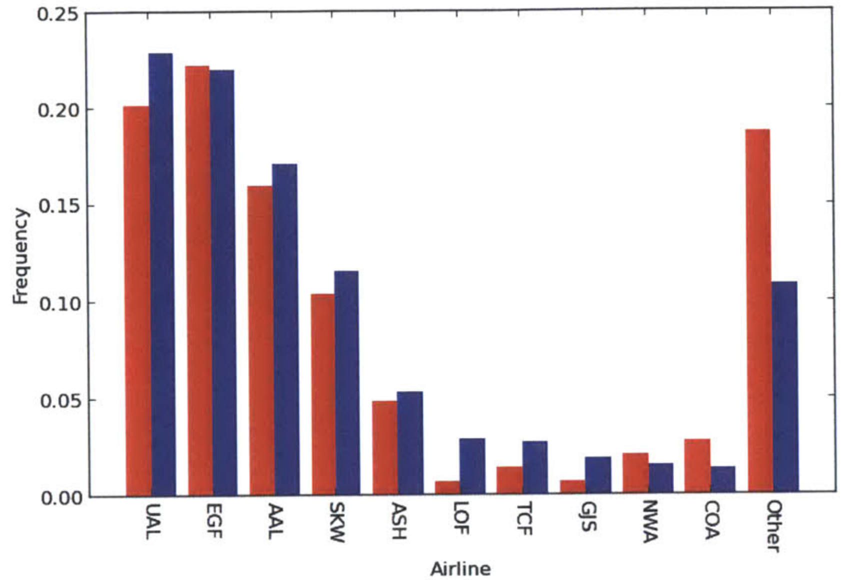

Figure 4-2: Normalized histogram of severe weather penetration by airline for airlines with at least 100 flights in the flightset.

Table 4.2: Aircraft weight classes as outlined by FAA Order 7110.65.

Weight Class	Maximum takeoff weight
Heavy	More than 255,000 pounds
Large	Between 41,000 and 255,000 pounds
Small	41,000 pounds or less

4.1.3 Aircraft size

In previous sections we have mentioned aircraft size as a potential influencing factor that is strongly correlated with other features. It is logical to consider aircraft size directly as well. While there are too many different types of aircraft to consider different models directly, there are many measures of aircraft size to consider. We use the aircraft weight class. These are summarized in Table 4.2.

While Heavy aircraft are more likely to fly through weather (Figure 4-3), the fact that the overwhelming majority of aircraft landing at Chicago O'Hare are Large aircraft means that this feature has only marginal skill.

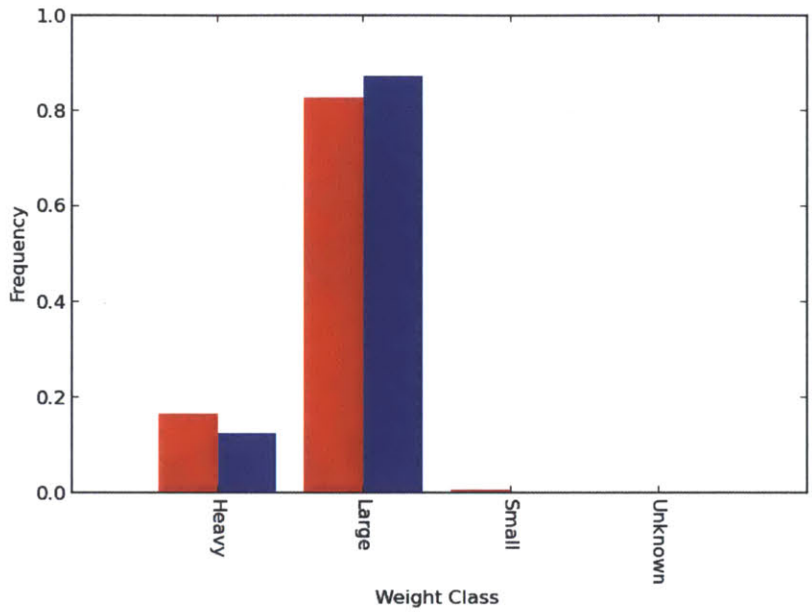

Figure 4-3: Normalized histogram of severe weather penetration by aircraft weight class.

4.2 Delay-based features

An early hypothesis was that a flight's on-time performance would be a determining factor in the pilot's willingness to penetrate severe weather. If a flight were delayed, for example, a pilot might be more willing to endure turbulence in order to land on time. Similarly, a pilot arriving at the destination airport earlier than expected might be more willing to deviate in the terminal area in order to avoid weather since the flight would still land on time.

There are many possible indicators of delay, and it is unclear a priori which would be the most relevant. Therefore, we consider three different features: pushback delay, wheels-off delay, and airborne delay.

4.2.1 Pushback delay

Pushback or gate delay is defined as the number of minutes after the scheduled gate pushback time that the aircraft actually pushes back from the gate. Pushback delay based on the flight plan is given directly by the ASPM database in field `DLAFPOUT`. The delay is set to zero in the event the flight pushes back early.

Note that we are using the delay based on flight plan and not the delay based on the

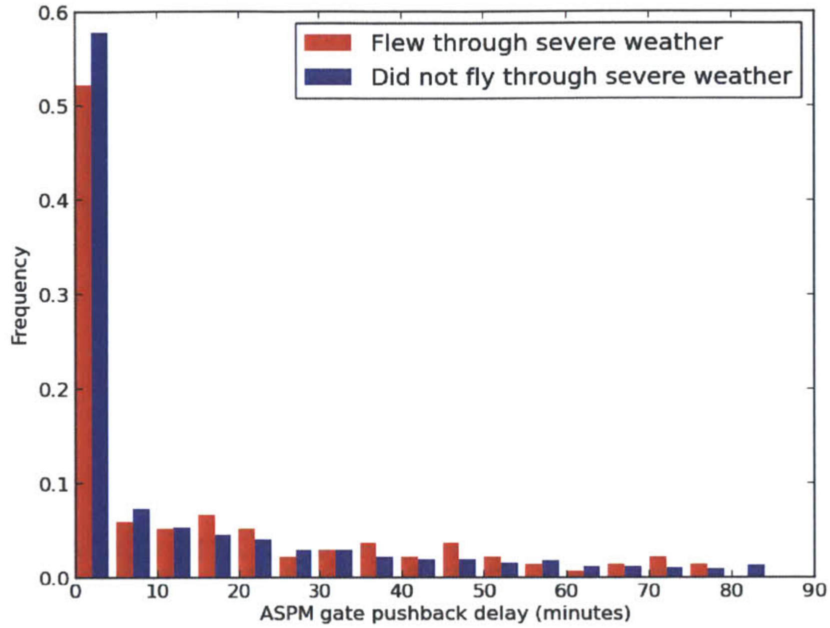

Figure 4-4: Normalized histogram of severe weather penetration by ASPM pushback delay.

schedule. Typically, a schedule is set by the airline several months in advance, which may be modified by a flight plan filed up to one day before the flight's departure. Because the flight plan is more recently updated, it is likely that the pilot is more aware of this time rather than the originally scheduled time. Over all flights departing in June, July, or August 2008, the delay based on flight plan differs from the delay based on schedule 13.17% of the time (13,401 out of 101,766 flights), with the scheduled delay greater 97% of the time (13,012 of 13,401 flights).

Figure 4-4 plots the normalized histograms for flights that did and did not penetrate severe weather as a function of the ASPM pushback delay. It is immediately apparent that even during weather events, the majority of flights do in fact push back from the gate on time. For flights that are delayed by more than 15 minutes, there is a slight bias towards weather penetration.

4.2.2 Wheels-off delay

We also consider the wheels-off delay as a feature in our model. This refers to the delay in taking off from the origin airport. Wheels-off delay is directly given by the ASPM database as `DLAFPOFF`. A long wheels-off delay could simply indicate a late pushback; however, a

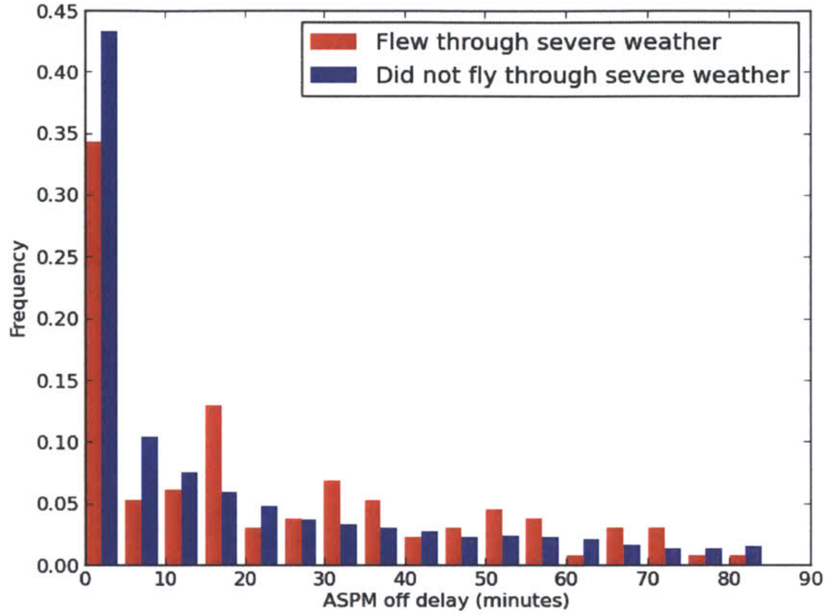

Figure 4-5: Normalized histogram of severe weather penetration by ASPM wheels-off delay.

long wheels-off delay with a short pushback delay would indicate a long tarmac delay at the origin airport. If pilots are more aware of flight time than block time, the wheels-off delay could be significant.

A normalized histogram of severe weather penetration by wheel-off delay is shown in Figure 4-5. Relative to Figure 4-4, there is an even more noticeable bias towards weather penetration in flights that take off late.

4.2.3 Airborne delay

Finally, we would like to consider the airborne delay at the time that a flight enters Chicago's terminal area. A flight is said to be delayed upon entering the terminal area if it took longer than usual to fly from its origin airport to within 200 km of its destination airport.

In order to estimate the airborne delay, it is necessary to compare each trajectory in the flightset with fair weather trajectories between the origin airport and ORD. Eight days from summer 2008 were chosen with minimal weather and no unusual events to serve as a baseline. While it is virtually impossible to find a day with absolutely no weather anywhere in the continental United States, there were no large-scale systems or widespread thunderstorms on the days chosen. These days are listed in Table 4.3.

For each of the 211 origin airports in the ETMS-ASPM flightset, we consider all flights

Table 4.3: Days with minimal weather used as a baseline for computing average flight times.

Fair weather days
2008-06-01
2008-06-03
2008-06-05
2008-06-07
2008-06-09
2008-06-11
2008-07-31
2008-08-05

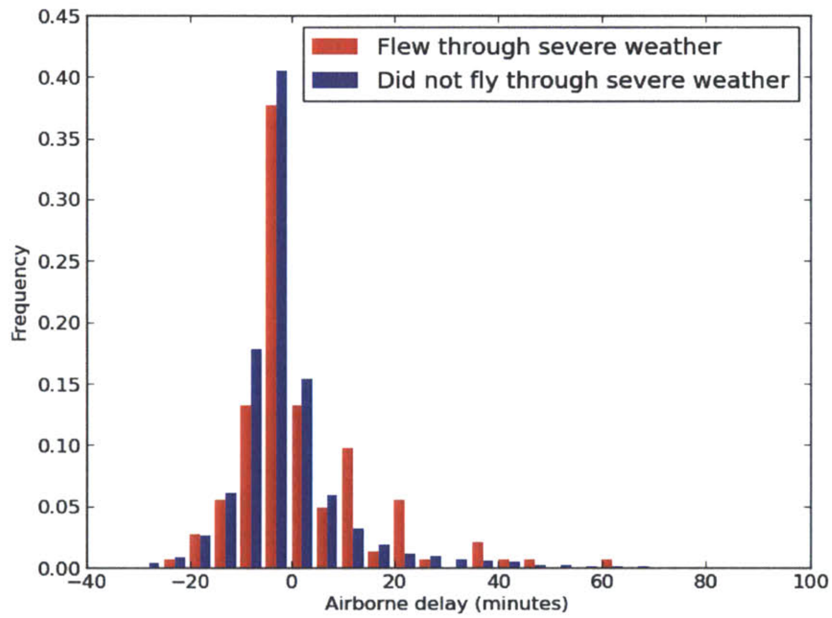

Figure 4-6: Normalized histogram of severe weather penetration by airborne delay at the time the flight reaches the Chicago terminal area.

from the origin airport to ORD during the eight fair weather days. For each flight, the ETMS flight time from wheels-off to entering the terminal area (200 km) around ORD is determined. Although there is naturally some variation even on fair weather days, the mean of these flight-to-terminal-area times is computed, stored in a separate database, and used as a point of comparison for computing airborne delay.

Once the baseline flight times have been estimated, the airborne time of each flight in the ETMS-ASPM flightset at the time it enters the terminal area is computed and compared to the stored fair weather estimate. The delay is the difference between these times. A flight that arrives earlier than the baseline is assigned a negative delay.

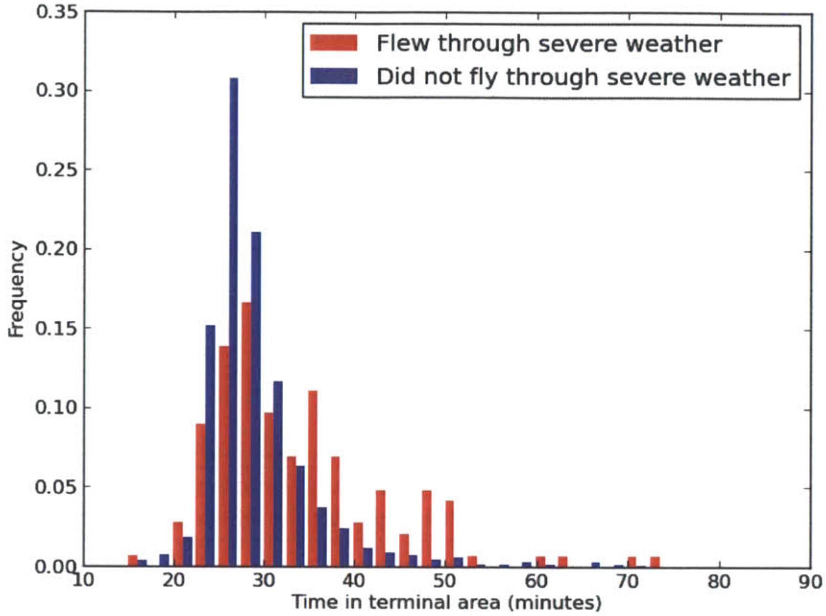

Figure 4-7: Normalized histogram of severe weather penetration by time spent in the terminal area.

Figure 4-6 contains the normalized histograms for airborne delay. As with both of the previous delay indicators, there is a bias towards severe weather penetrations with positive delay. There is an analogous negative bias when the flight has arrived at the terminal area earlier than usual.

With all of these delay indicators, it is should be noted that a positive correlation does not necessarily imply a choice on the part of the pilot. It may simply be the case that the presence of more severe weather makes any delay more likely. Nevertheless, it is clear that all the delay indicators have at least weakly positive skill in predicting pilot willingness to penetrate weather.

4.2.4 Time spent in the terminal area

Another feature considered is the total time each flight spends in the terminal area. This is a potentially problematic feature because it cannot be predicted in advance; the time can only be known after the flight lands, at which point a prediction would be unnecessary. Nevertheless, it turns out to correlate reasonably well with weather penetration behavior, so we include it here (Figure 4-7).

One might naively expect that longer terminal times would be correlated with lack of weather penetration since it might signal pilots deviating to avoid weather. However, our case studies indicate that in many cases where pilots deviate significantly to avoid weather, the blockage is usually pretty severe and avoiding it completely is impossible. In some cases this may be because there is weather very close to the airport; in these cases runway requirements prevent pilots from deviating.

4.2.5 Number of pilots in terminal area

For each flight, we consider the number of pilots in the terminal area at the time the flight in question first enters the terminal area. It is not apriori obvious whether this indicator would correlate positively or negatively with weather penetration. On the one hand, this number approximates the congestion level in the terminal. High levels of congestion would prevent pilots from deviating too much, which may result in increased likelihood of penetrating severe weather. On the other hand, the presence of severe weather in the terminal could lead to decreased numbers of flights entering the terminal in the first place.

As a proxy for congestion, the measure is problematic for several reasons. First, flights may not be equally distributed in all directions. For example, a flight entering from the northeast when most flights are clustered around the southwest would experience almost no congestion, but this would not necessarily be reflected in the raw flight count. Second, whether a high number of flights causes congestion depends on how the flights are sequenced. If the controllers are able to densely sequence many flights on a single route, the flow can continue uninterrupted even with a large number of flights. In contrast, if the flight paths are chaotic, a much smaller number of flights could be enough to cause delays.

As shown in Figure 4-8, the indicator is negatively correlated with penetration of severe weather, likely for the reasons outlined above: the presence of severe weather in the terminal area would cut down on the total number of flights. It may also indicate that when severe weather exists in the terminal area, higher flight counts occur only when the controller has found a way to efficiently route flights around weather. More features dealing with this type of stream behavior will be discussed in Chapter 5.

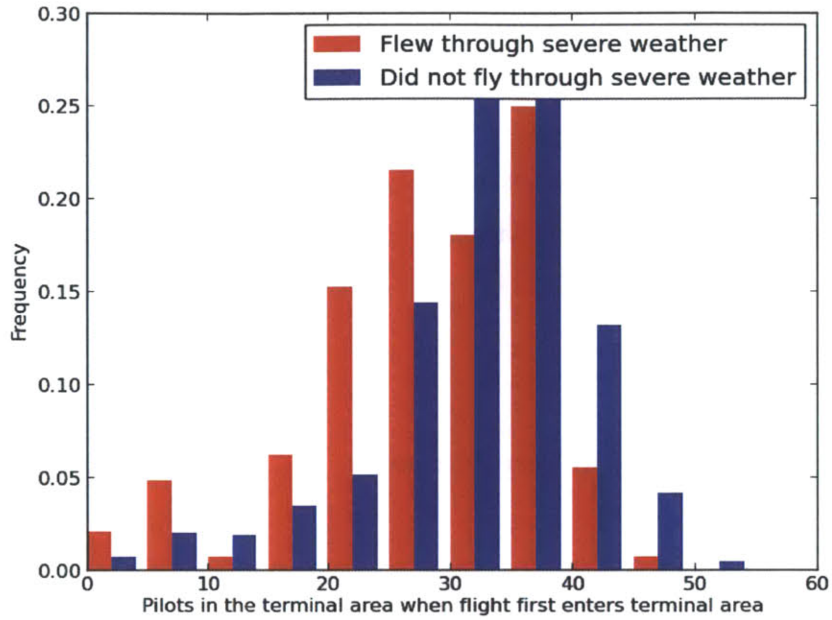

Figure 4-8: Normalized histogram of severe weather penetration by number of pilots in the terminal area when the flight first entered the terminal area.

4.3 Weather-based features

While this thesis is focused on operational factors that affect pilot penetration of severe weather, we found it useful to incorporate several weather-based features.

4.3.1 Conditions near the airport

Case studies indicated that weather near the terminal area was particularly difficult to avoid. A feature that captures whether weather is near the airport is therefore a reasonable one to use. To quantify this, we count the number of WAF pixels within 50 km of the airport that are 80% or above. This matches our threshold for weather penetration. 50 km was chosen because that is slightly greater than the longest commonly observed "trombone" paths taken when flights line up for landing on particular runways. While this feature produces numbers that are somewhat difficult to understand, it directly corresponds to the weather coverage near the airport.

Figure 4-9 indicates that this feature is strongly correlated with severe weather penetrations. When there are at least 500 WAF pixels of 80 or above within 50 km of the terminal area, most flights will penetrate severe weather. (For reference, there are about 8000 pixels

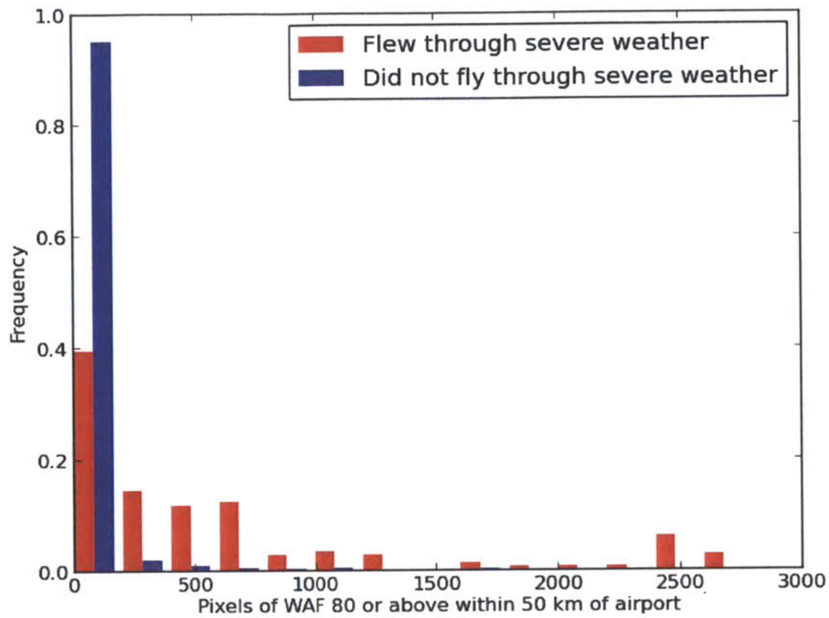

Figure 4-9: Normalized histogram of severe weather penetration by WAF coverage within 50 km of the airport.

within 50 km of the airport.) Although 6% coverage may seem low, it is enough to make it extremely difficult for flights to maneuver around when they need to land at the airport.

4.3.2 Conditions farther from the airport

To balance the previous feature, we introduce another feature that counts the number of pixels between 50 and 200 km away from the airport. In other words, this feature quantifies the weather within the terminal area but away from the airport.

Unlike the near-airport weather, in the annulus around the airport we see far more flights avoiding weather despite the presence of severe weather. (For reference, there about 118000 pixels in the annulus.) This is to be expected, since flights have far greater flexibility to avoid weather in the outer terminal.

4.3.3 Total amount of weather in the terminal area

Finally, we include a feature indicating the total amount of weather in the terminal area. Following the previous features, this is simply the total number of WAF pixels over 80 within 200 km of the airport.

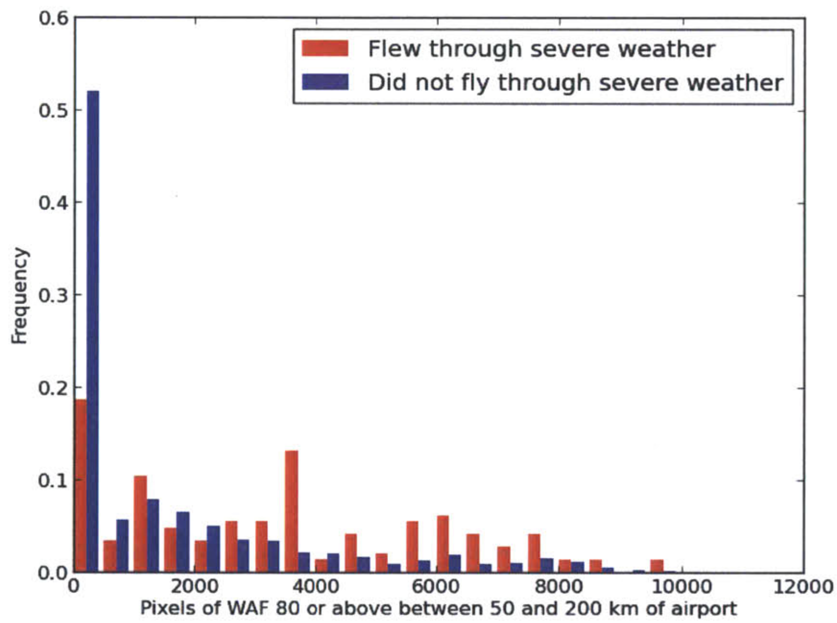

Figure 4-10: Normalized histogram of severe weather penetration by WAF coverage between 50 and 200 km of the airport.

Chapter 5

Traffic Flows

In our discussions with experts in the field and in our own case studies, it was frequently observed that pilot behavior was greatly influenced and indeed controlled by the behavior of those around them. Air traffic controllers refer to flows or streams of flights; it is important to understand how these streams are set up and how they change in order to understand the airspace dynamics. We therefore consider several features which help to identify whether a pilot is part of such a stream and how this affects their behavior. In order to do this, it is necessary to dynamically identify where the streams are and whether a trajectory is assigned to a particular stream.

The terminal area is where pilots coming from all directions are merged into one of several arrival streams in order to be properly spaced and sequenced for landing. Chicago O'Hare has four arrival fixes or cornerposts, each approximately 50 km from the airport. Arriving aircraft typically pass over one of these arrival fixes before being routed into the appropriate landing trajectory. These are listed in Table 5.1 and mapped in Figure 5-1. The four departure directions are interspersed with these four cornerposts. We implicitly divide the terminal area into two regions: the outer region, where flights are merged intoa single stream passing through one of four cornerposts; and the inner region, where flights follow a fairly well-specified route from the cornerpost to their assigned runways. Identifying the streams through each cornerpost is the heart of this analysis. From there, we define several more features which can be used to predict severe weather penetration.

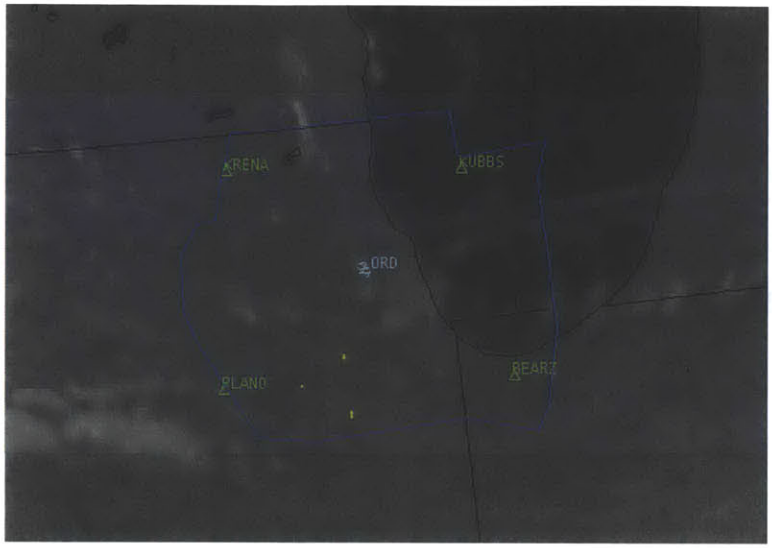

Figure 5-1: Map of Chicago O'Hare arrival fixes. O'Hare's TRACON is outlined in blue.

5.1 Defining streams and cornerposts

Although arriving aircraft are typically required to pass over one of these four cornerposts and have fairly specific flight paths within the near-terminal area, the route within the terminal area to the assigned arrival fix can vary significantly. Arrival and departure trajectories on July 8, 2008, at 1630Z, a typical fair weather afternoon, are plotted in Figure 5-2. Trajectories are plotted for departures within the last 40 minutes and for all arrivals airborne within 200 km of the airport. The triangles indicate each aircraft's current position; the aircraft's trajectory up to its current position is shown as a solid line while the future trajectory points are dotted.

This plot demonstrates some of the difficulty in identifying arrival streams. While all aircraft must be lined up for landing, they may be entering the terminal area from many different directions. Arrival trajectories into the different cornerposts vary in how closely they follow the same path. At the time in question, flights arriving to the northwest cornerpost tend to take more varied paths, while the flights at the southeast cornerpost all follow roughly the same path. At the southwest cornerpost, most flights follow nearly the same path, but two deviate significantly. It is therefore difficult to clearly define whether a given flight is truly part of a stream or if it is simply taking a similar route from its origin to the arrival cornerpost. This is significant, since a pilot in a tightly controlled stream could be influenced by the behavior of the preceding pilot in the same stream, whereas a

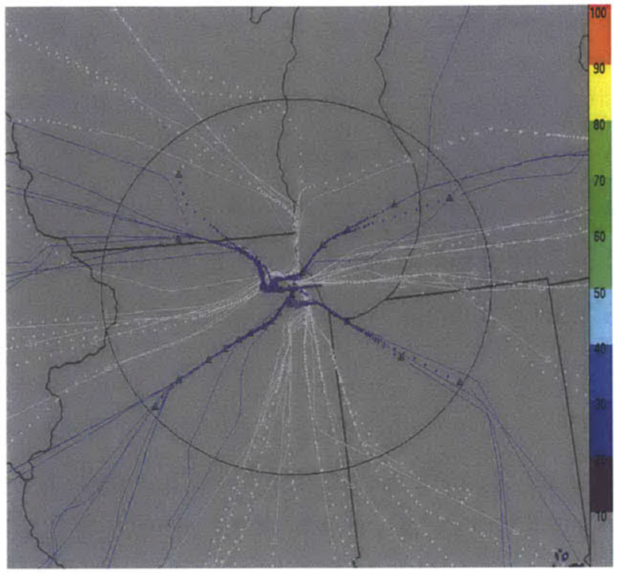

Figure 5-2: Terminal-area flights on July 9, 2008, at 1630Z. Departures are grey and arrivals are purple.

pilot who simply happened to be following a similar route would be more likely to make independent decisions.

Because of this difficulty, we do not attempt to quantify absolutely whether a flight is part of a stream at any given time. Instead we primarily use the flight's observed cornerpost as the basis for our analysis, identifying the number of flights at each cornerpost and how spread out they are. Guided by case studies, we then identify various features that could potentially influence pilot behavior.

5.2 Detecting active cornerposts

The first step in the process is to identify active cornerposts and how closely clustered the flows are around each cornerpost. Although exact distances vary, each cornerpost is approximately 50 km from the airport; this distance is used for all cornerposts for consistency. We examine the location at which each flight first comes within 50 km of the airport, and consider the angle of the line between this location and the airport. Approaching the airport from due east heading due west is considered 0 degrees. Note that this analysis is done using the 1-km grid contained in the weather data; as such, there will be some distortion due to the projection.

5.2.1 K-medians clustering

The problem of detecting clusters of incoming flight angles is analogous to the 1-dimensional k-medians clustering problem over the integers modulo 360. Given a set of points, we wish to find k centroids and assign each point to its nearest centroid such that the median error between each centroid and its set of assigned points is minimized. Although this problem is theoretically computationally difficult, we use a common heuristic which gives reasonable results given our application.

The heuristic begins with k randomly chosen centroids, assigns each point to the nearest centroid, and then readjusts the centroids to minimize the median error amongst all points assigned to it. Then the points are reassigned to the nearest centroid, which in some cases may not be the same one. The process is repeated until every point assignment remains unchanged or until a set number of iterations is reached. In most cases the heuristic converges after only a few iterations. However, depending on the input, the result can sometimes depend on the initial centroid assignments. Because of this, the algorithm is run 10 times with random starting configuration; the result with minimum average error across all centroids is used. Different numbers of repetitions were tried; no significant improvement was observed for more than 8 repetitions.

Finally, since the heuristic requires k to be specified, and it is not known in advance how many flight clusters there will be, the heuristic is run for $k = 1, ..., 8$. Although there are typically only 4 clusters, this range allows for spontaneous clusters which may form as a result of weather blocking regular routes. When k is too small, typically the average errors will be relatively large since streams will be far apart. When k is too large, there are two common outcomes. First, single flights can be assigned to their own cluster, especially when they are far from any other cluster. Second, two clusters may be very close to each other (within 1 or 2 degrees). Fortunately, these two are simple to check for; once these are eliminated, the result with lowest average error is chosen.

5.2.2 Results

For each 2.5-minute interval in our case set, we consider the set of flights with destination ORD within 200 km of the terminal area. The 2.5-minute temporal resolution matches the temporal resolution of the weather data. For each flight in this set, we determine the angle

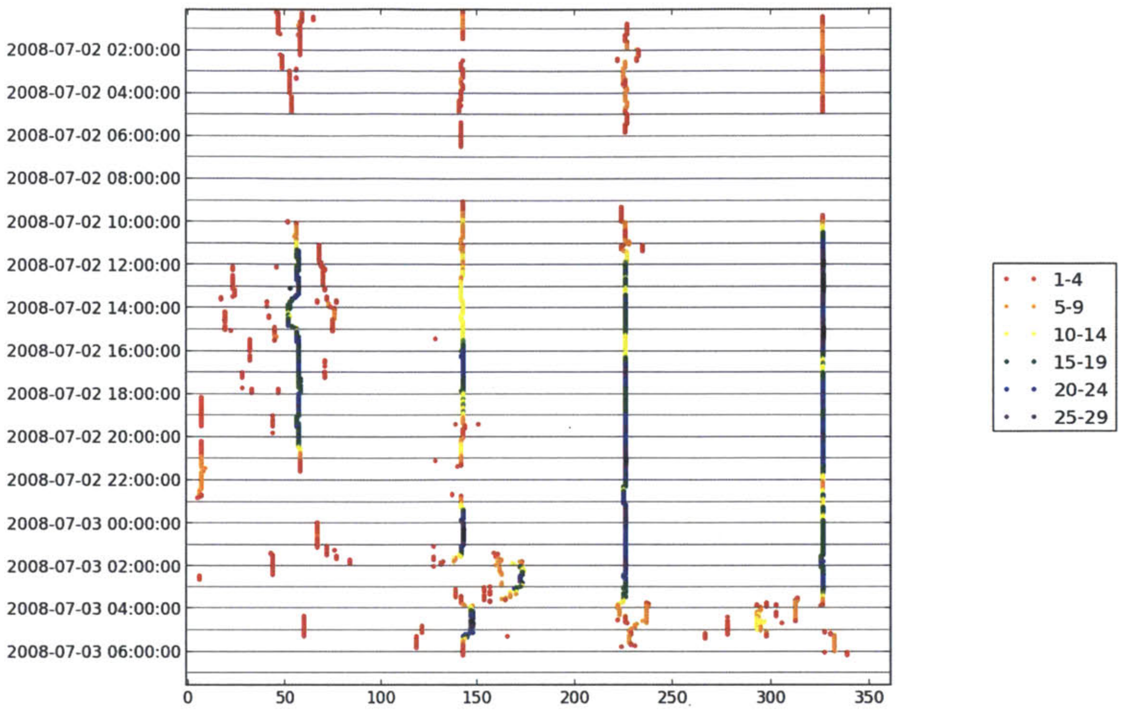

Figure 5-3: Arrival streams on July 2, 2008, at a radius of 50 km from the airport.

at which it crosses the circle of radius 50 km centered at the airport. These angles are fed into the k-medians clustering algorithm to detect where the currently-used arrival fixes are.

Once this is done, we can plot the locations of these clusters to observe how they change over the course of the day. This is shown in Figure 5-3 for one of the case days in this thesis, color-coded by the number of flights in each stream. Each stream has at least two flights assigned to it. As before, an angle of 0 represents a flight approaching from due east heading due west towards the airport. For comparison, arriving flight trajectories at 1400Z on the day in question are plotted in Figure 5-4. In this plot, the inner circle has radius 50 km and the outer circle has radius 100 km.

As can be seen from Figure 5-4, at 14Z on this day the flights arriving from the northeast tended to be more spread out, with at least three distinctive streams. This matches what we see in Figure 5-3, where at 14Z we observe one main stream with 15-19 flights at just over 50 degrees, with two smaller streams to either side. By contrast, the northwest and southeast cornerposts have two tightly clustered streams, evident in both plots. Finally, the arrival trajectories are somewhat more spread out in the southwest quadrant as seen in the map (Figure 5-3), but the clustering algorithm identified this as a single cluster with a

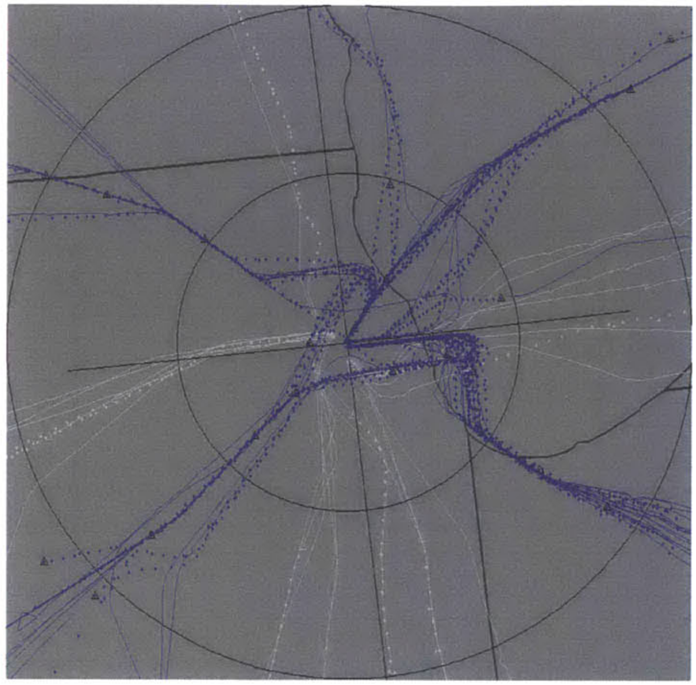

Figure 5-4: Arrival trajectories on July 2, 2008, 1400Z.

greater spread.

Similar charts can be made for each case day and share many similarities. Generally speaking, the four established cornerposts are stable, varying by only a few degrees. The arrival fixes and their approximate angles are listed in Table 5.1. Note that while these correspond closely to the actual fixes, the positions were not input into the clustering algorithm; rather, these were the locations most commonly detected by the algorithm. When an arrival fix is affected by weather, the fix is typically closed and flights are rerouted to other fixes. Although streams can sometimes shift position, analysis of our case days indicates that this is more likely to be the result of low traffic counts allowing greater controller flexibility. An example of the former can be seen in the two southern cornerposts at the end of the July 2 case in Figure 5-3. Multiple streams are most commonly seen in the northeast quadrant, and are observed during both fair weather events and convective weather events.

Table 5.1: Chicago O'Hare arrival fixes.

Fix	Cornerpost	Angle
KUBBS	Northeast	57
KRENA	Northwest	143
PLANO	Southwest	226
BEARZ	Southeast	327

5.3 Stream features

Once we have identified the flight streams entering the near-terminal area at any given time, we can extract several more features for each flight regarding its trajectory relative to the existing streams. In general, flights adhere closely to the prescribed flight path within the near-terminal area (within 50 km of the airport). This flight path depends on the assigned cornerpost and runway; we use these two pieces of data to identify each flight's near-terminal stream. However, as mentioned previously, outside 50 km, flights tend to be much more spread out, and it does not generally make sense to assign flights to specific streams. The features described below reflect these two observations.

5.3.1 Cornerposts

The first and simplest feature to examine is the assigned cornerpost itself. We consider both the cornerpost used and the current location of the actual stream.

Arrival fix

There are four standard ORD arrival fixes (Table 5.1). As we have seen, most flights pass over one of these. If the flight enters the near-terminal area within 5 degrees of a particular arrival fix, that flight is assigned to that fix. Otherwise, the flight's arrival fix is assigned a null value.

Arrival angle

Another feature considered is whether the flight enters the near-terminal area as part of a stream. These are indicated by the angle relative to the airport as described in the previous section. For each flight, the existing streams at the time the flight first enters the near-terminal area are retrieved. If the flight's distance from the nearest stream's centroid at the entry point is at most twice the average error for that stream, the flight is considered to be in the stream. This allows outliers while still taking into account the width of each stream.

5.3.2 Number of streams

The total number of streams at the time each flight enters the terminal area, excluding outliers, is included as another feature. Typically there would be four arriving streams, though during low-demand periods there are often fewer. Furthermore, under weather conditions it is possible that one or more would be closed as the result of weather. It is also possible that under weather conditions different aircraft would be rerouted into different streams in order to avoid flying through severe weather.

5.3.3 Previous pilot's behavior

A commonly heard conjecture is that a pilot is likely to follow the preceding pilot in a stream, and that this has an undue influence on whether a pilot choose to fly through severe weather. This may be especially problematic during times when the weather is worsening, but each individual pilot chooses to follow the previous pilot's path through weather. This feature examines this behavior in the near-terminal area, where flight streams are strictest and most clearly defined.

Each flight is assigned to a final flight path according to its arrival stream and runway. It is important to consider the runway because flights landing on different runways will have significantly different landing paths even if they are assigned to the same arrival fix. We then retrieve the weather penetration behavior (the maximum WAF penetrated by the pilot in the terminal area) of the preceding pilot on the same stream and runway.

5.3.4 Number of pilots in the same stream

If pilots do tend to follow each other, examining how flows are created and why they cease could help explain weather penetration behavior. Following the hypothetical situation described in the previous section, eventually the weather would worsen to such a degree that some pilot would advise the following pilot not to fly through. If this is the case, it would imply that pilots who are last in their stream are more likely to fly through severe weather.

We capture these features by counting the number of flights in the same stream within the preceding and following 30 minutes. This also gives a sense of how densely spaced the stream is, which may in itself be a significant factor. Thus, a low flight count in the 30 minutes before indicates that the flow is being established, while a low flight count in the following 30 minutes indicates that the flow is being shut down or rerouted. A low flight count before and after would indicate a low-density route; a flight is considered an outlier if there are no flights within a 30 minute window.

5.3.5 Stream width

If our primary hypothesis is that pilots are influenced the flows around, it is logical to consider how tightly controlled each flow is. The width of a flow (the spatial variance of trajectories in each stream) is a possible proxy indicator for how strictly pilots must follow the prescribed path.

It is not a priori clear whether this indicator would be positively or negatively correlated with weather penetration, assuming it has any skill. It is possible that a narrow stream would be positively correlated with weather penetration since pilots are less likely to request a deviation. On the other hand, streams typically become more spread out in weather events as each pilot tries to avoid particularly strong convective weather cells.

5.3.6 Runway

We previously used each flight's assigned runway to identify its near-terminal stream. The runway used can also be considered an independent feature. The landing runway is determined from the ETMS trajectory data; the runway used is clear from the final heading of each aircraft as determined from the last five trajectory points. In most cases, these indicate the aircraft's final heading unambiguously; the handful of ambiguous cases (usually due to a tightly curving arrival path) were manually assigned to runways.

Chicago O'Hare has three pairs of parallel runways (9/27, 4/22, 14/32) and a single runway (10/28) that is less often used. Since the single runway is quite close to one of the larger parallel runways, it is sometimes difficult to determine whether an aircraft is landing on 9/27 or 10/28. In cases of ambiguity, we assume that the aircraft is landing on one of the 9/27 runways. While it is possible to determine more accurately precisely which runway each aircraft is using, this precision is unnecessary since we are mostly concerned with the flight path used on the approach and not the runway itself.

Chapter 6

Predictive Modeling of Pilot Behavior

Using the features defined in the previous two chapters, we apply two predictive models to estimate the relative skill of the various features in predicting severe weather penetration.

6.1 Decision Trees

We first used a decision tree classification algorithm to model the data. A decision tree recursively partitions the data into two sets, finding a partition at each step that maximally differentiates the two sets. In our case, each step divides the flights into a set that did not penetrate severe weather and a set that did, while minimizing the misclassifications. A node is considered terminal when the number of flights at the node has reached some minimum threshold or when all partitions have high error rates.

Decision trees were chosen for several reasons. First, the model can be used with a relatively small sample size. Since the number of case days was so limited, this was an important consideration. Second, the decision tree model is a white-box model: the predicted outcome for each flight clearly follows from the flight's characteristics. Since the primary reason for using a predictive model is to understand the relative importance of various features, it was illustrative to use a white box model. Finally, it is able to handle both numerical and categorical data; our features include both types.

6.1.1 Balancing the dataset

Oversampling is a method frequently used in predicting rare events. This is used to balance the dataset when many more samples of one outcome exists in the dataset. In our case, we have many more examples of flights that did not penetrate severe weather (over 9000) relative to the number of flights that did (144). In these situations, a predictive model could be reasonably accurate by simply choosing the more likely event in all cases. This is in fact what we find: a decision tree model trained on the unmodified data produces a model with a high missed detection rate.

To re-balance the dataset, the flights that penetrated severe weather are duplicated in the training dataset to approximately match the number of flights that did not penetrate severe weather. The model is then trained and verified on this balanced dataset, which artifically boosts the weight of the smaller set.

It is also possible to balance the dataset by randomly discarding flights from the larger set. Due to the exceedingly small number of weather-penetrating flights, the majority of the dataset would need to be discarded; this was deemed undesirable.

6.1.2 Training the model

Each flight in the flight dataset was randomly assigned to one of two sets, a training set and a test set. Flights where some of the features could not be computed due to missing data or other reasons are rejected; this applies to approximately 9% of the flights. Weather-penetrating flights were duplicated to the degree necessary to balance each dataset. The training set was used to train a decision tree to predict severe weather penetrations; the tree was then validated using the test set. This procedure was repeated 8 times with different random partitions of the original data set. Two sample trees are shown in Figure 6-1. Each branch indicates the criterion for the left-hand daughter node; each node is labeled with the probability of severe weather penetration and the number of flights in the training set assigned to that node. For categorical values, indices are given rather than the full list of acceptable values for space reasons. The short names for the features are listed in Table 6.1.

Table 6.1: List of features used in the predictive model.

Variable name	Feature description
allwx	Number of WAF pixels ≥ 80 within 200 km of the airport
nearwx	Number of WAF pixels ≥ 80 within 50 km of the airport
farwx	Number of WAF pixels ≥ 80 between 50 and 200 km of the airport
flightrange	Origin-destination distance
termtime	Minutes spent in the terminal area
pilots	Number of pilots in the terminal area
outdelay	Pushback delay
offdelay	Wheels-off delay
airdelay	Airborne delay
wtc	Aircraft weight class
body	Aircraft size (Narrow or Wide)
wtc	Aircraft weight class
airline	Airline (XX if fewer than 100 flights)
runway	Landing runway
runwayp30	Number of pilots on the same stream in previous 30 min
runwayn30	Number of pilots on the same stream in next 30 min
cornerpost	Cornerpost (NW, NE, SE, SW)
stream	Nearest major incoming stream or X if outlier

6.1.3 Results

Although the eight trees are not identical, they share certain characteristics. First, the feature at the root of the tree is consistently a weather feature indicating the presence of severe weather very close to the airport. Thus, the most significant determinant is the weather itself, and not any operational factors. Second, in all eight trees, the right-hand side of the tree is significantly less complex than the left-hand side, and will generally yield a prediction of severe weather penetration. Since there is little to no deviation flexibility in this range, pilots who wish to land have little choice but to fly through severe weather, regardless of the operational characteristics of the flight. By contrast, on the left-hand side of the tree, the choice of whether to fly through severe weather is more complicated. Other than these two characteristics, the trees differed in which variables were used and at which level. This indicates that we are dealing with many weak predictors, which decision trees are not well-equipped to handle. Two features that are somewhat correlated may not both be used, despite their similarity.

Each tree was validated on the test set. The results are shown in Table 6.2. The decision trees trained in this method are accurate approximately 75% of the time. However, since this is validated on an oversampled dataset, the results should be taken more as an indication

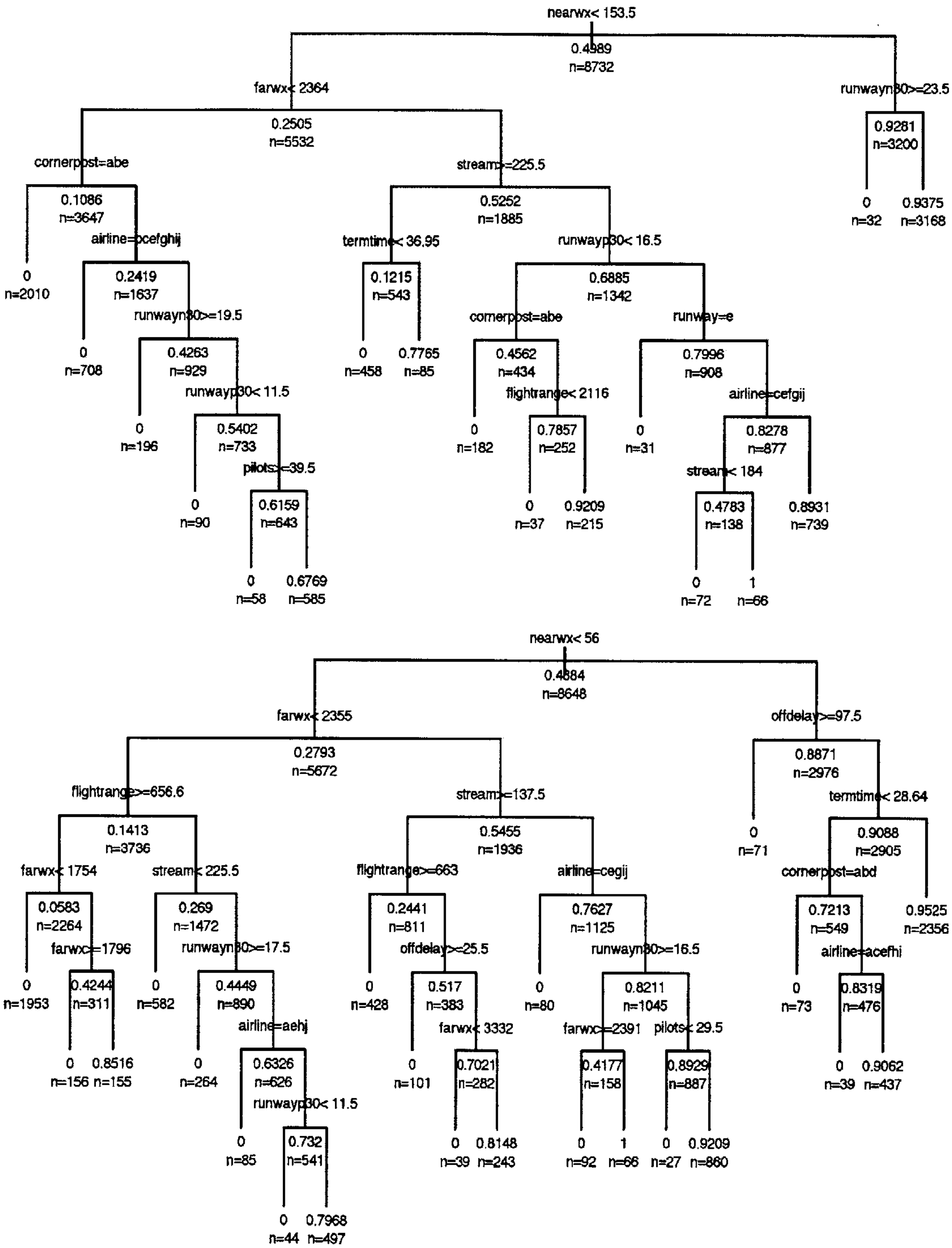

Figure 6-1: Two sample decision trees.

Table 6.2: Validation of the decision tree prediction model. The mean and standard deviation for the 8 trials are shown.

	Actual Yes	Actual No
Predicted Yes	2549 (443)	458 (18)
Predicted No	1578 (431)	3892 (106)

of the predictive power of different variables rather than as a predictive model.

6.2 Random forests

Random forests are a natural extension of decision trees, and are useful for predictions based on many weak predictors. Random forests were first described by Leo Breiman in 2001 [1]; this thesis utilizes the software package written by Breiman and others for our analysis. A random forest works by creating many decision trees using a random subset of the features and data in each one. Our model uses 500 trees in each forest, with each tree drawing 4 predictors at random. Each tree then votes on the outcome of the prediction variable.

Random forests are extremely robust. They are able to deal with many correlated variables without having one eclipse the other, as is possible with decision trees. The diversity of trees also solves the overfitting problem sometimes encountered with decision trees.

6.2.1 Dealing with unbalanced datasets

Instead of oversampling, random forests deal with unbalanced datasets in a more natural way. The vote threshold for severe weather penetration can be set explicitly and does not need to be 50%; this biases the classification algorithm in favor of the rarer event. This is especially useful when the penalty for missed detections is greater than the penalty for false alarms.

It is also possible to explicitly take equal numbers of samples from the two sets; however, it was found that specifying the sample sizes in this manner did not improve results.

6.2.2 Results

The random forest model was trained and tested using eight random partitions of the dataset. A summary of results is shown in Table 6.3 for several vote thresholds. Because the

Table 6.3: Summary of results from random forests with several different vote thresholds.

Threshold for yes vote	Predicted	Actual Yes	Actual No
.7	Yes	34 (3)	145 (28)
	No	33 (6)	4235 (59)
.5	Yes	51 (4)	556 (56)
	No	15 (4)	3823 (70)
.3	Yes	63 (5)	1728 (186)
	No	4 (4)	2651 (206)
.1	Yes	66 (7)	3662 (300)
	No	0 (0)	717 (293)

random forest model is tested using data that has not been oversampled, the results cannot be directly compared to the decision tree. Nevertheless, the random forest model does reasonably well on the test dataset, achieving accuracy rates up to 90%. Unfortunately, the overall accuracy is highly skewed towards the negative prediction. The lack of oversampling explains the high false alarm rate.

6.2.3 Sensitivity analysis

The results are highly dependent on the vote threshold, which controls the balance between false alarms and missed detections. The false alarm rate increases enormously as the vote threshold is lowered. At the same time, decreasing the vote threshold lowers the number of missed detections. A reasonable balance point seems to be somewhere between .3 and .5, depending on the relative costs of missed detections and false alarms. The sensitivity curve summarizing this tradeoff is plotted in Figure 6-2.

6.2.4 Variable importance

The random forest model allows us to rank the features by skill. To do this, we randomly permute the values of each variable and measure the decrease in accuracy of the resulting tree using a Gini index. This process is repeated for all trees in the forest containing the variable in question; the resulting average is the variable importance. A higher value indicates greater skill. These were computed for an arbitrary run of the random forest model with vote threshold .5 and are summarized in Table 6.4.

The model indicates that the most importance features remain weather features; these have greater significance than any operational feature except one. The time spent in the

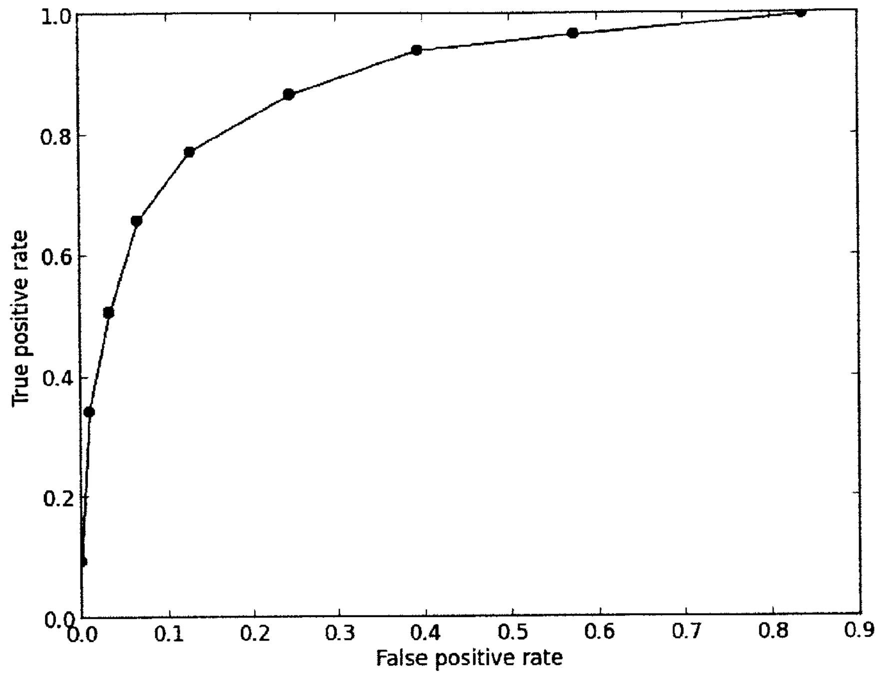

Figure 6-2: Sensitivity curve summarizing tradeoff between false alarms and missed detections, parametrized by the vote threshold.

Table 6.4: Variable importance as given by the random forest model.

Variable name	Feature importance
nearwx	22.7
termtime	16.3
farwx	12.3
allwx	11.6
flightrange	10.6
runwayp30	10.6
pilots	7.6
stream	7.4
airline	6.9
runwayn30	6.7
offdelay	5.4
airdelay	5.5
outdelay	3.9
runway	3.2
cornerpost	2.8
body	0.3

terminal area is significant, with longer terminal times correlated to higher likelihood of weather penetration; however, this feature is somewhat problematic since it can only be known after the flight has landed. A longer distance flight is also associated with higher likelihood of penetrating severe weather. There are several reasons this might be true. First, longer distances are correlated with larger aircraft, which may be better able to penetrate severe weather. Second, flights arriving from farther away are less likely to be impacted by ground delay programs or ground stops; there may simply be a larger proportion of long-distance flights during severe weather periods. Comparable in skill to the flight range is the number of preceding pilots on the same approach path, with higher numbers corresponding to increased likelihood of severe weather penetration. This supports our hypothesis that pilots tend to follow established paths, perhaps through worsening weather.

Thus, while most of the variables are only weakly correlated with severe weather penetration, we are able to develop a random forest prediction algorithm that correctly classifies

Chapter 7

Conclusions and Future Work

7.1 Summary and conclusions

Using a combination of case studies and predictive modeling, we have identified and tested features that correlate with pilot penetration of severe weather in the Chicago O'Hare terminal area. Case studies were used to identify and extract the relevant features, and their relative importance was measured using a random forest predictive algorithm. While our initial hypothesis was that operational factors were significant in determining which flights penetrated severe weather, our study shows that the primary indicators continue to be weather features, particularly the presence of weather very close to the terminal area.

Nevertheless, we found a number of operational features that weakly correlate with severe weather penetration. Despite having less importance, these features help shed light on the dynamics of the terminal area. In particular, the importance of several of the stream-based features may help us understand how pilots and air traffic controllers deal with weather in the terminal area. The most important conclusion was that pilots are more likely to penetrate severe weather when they are part of a stream that crosses through weather and less likely when they are "pathfinders" leading a stream; this implies that rerouting around weather is still often done on an ad hoc basis once a pilot has flown through severe weather and reported the event to an air traffic controller. Understanding these dynamics may lead to the development of more optimal weather mitigation strategies in the terminal area.

7.2 Future work

There is still a great deal of work to be done in understanding the impact of severe weather in the terminal area and when and why pilots choose to penetrate such weather. Our work plan for the next few months addresses some of the shortcomings of this thesis and considers implications for planning.

7.2.1 Expand databases

The predictive power of our model is severely limited by the relatively small number of weather-penetrating flights in our dataset. Ideally, we would acquire data from more recent years to expand our case set and verify that the same patterns hold year over year.

7.2.2 Adding more features

An expanded case set would likely suggest additional features that could be added to the model. We may also explore adding advisory features, though the strong correlation with weather events could make such features confusing.

7.2.3 Implications for weather forecasting

Finally, we would like to explore the implications of forecasting on observed pilot behavior. We would like to compare the severe weather penetration events to the forecasts several hours earlier to understand whether the weather penetration event was a surprise. If most weather penetration events were a result of unforecasted weather, for example, pilots who flew through severe weather may have had no other choice. On the other hand, if the forecasts largely match the actual weather, this would imply that such weather could have been avoided, and such weather penetration events may be a calculated decision on the part of the pilot or air traffic management.

Furthermore, our study may have implications for forecasting. It is often difficult in forecasting to ascertain the precise location of weather cells, particularly several hours in the future. This is generally considered particularly problematic when the weather may or may not be right over the airport. However, we have seen several cases where pilots are willing to fly through weather when it is very close to the airport, but avoid the same weather when it is farther away and they have more room to maneuver. If this is in fact the case, the precise proximity of weather to the airport may not be so crucial, at least for arriving airborne aircraft.

Bibliography

[1] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[2] R. DeLaura and J. Evans. Exploratory study of modeling enroute pilot convective storm flight deviation behavior. In *12th American Meteorological Society Conference on Aviation, Range and Aerospace Meteorology.*, Atlanta, GA, 2006.

[3] R. DeLaura, M. Robinson, M. Pawlak, and J. Evans. Modeling convective weather avoidance in enroute airspace. In *13th American Meteorological Society Conference on Aviation, Range, and Aerospace Meteorology*, New Orleans, LA, 2008.

[4] M. Matthews and R. DeLaura. Evaluation of enroute convective weather avoidance models based on planned and observed flights. In *14th American Meteorological Society Conference on Aviation, Range, and Aerospace Meteorology*, Atlanta, GA, 2010.

[5] M. Robinson, J. Evans, and B. Crowe. En route weather depiction benefits of the nexrad vertically integrated liquid water product utilized by the corridor integrated weather system. In *10th American Meteorological Society Conference on Aviation, Range and Aerospace Meteorology.*, Portland, OR, 2002.