2.4 COMMERCIAL AIRCRAFT ENCOUNTERS WITH THUNDERSTORMS IN THE MEMPHIS TERMINAL AIRSPACE *

Dale A. Rhoda, Benjamin G. Boorman, Elizabeth A. Bouchard, Mark A. Isaminger, Margo L. Pawlak
Massachusetts Institute of Technology
Lincoln Laboratory
Lexington, Massachusetts 02420-9185

1. INTRODUCTION

Thunderstorms are dynamic obstacles to the flow of air traffic. Aircraft routing in the presence of thunderstorms is as dynamic as the position and intensity of the storms. The question of where pilots will and will not fly is relevant to the decisions made by human air traffic managers as well as to the development of automated decision aid tools. In order to accurately anticipate which routes will be useable one needs to be able to 1) forecast the relevant weather variables, and 2) convert those weather variables into a quantitative probability that pilots will request deviations from the nominal route. The Convective Weather Integrated Product Team at the FAA is improving the accuracy and lead time of forecasts of thunderstorm products (Theriault, et al., 2000). This paper provides an update on our examination of the issue of probability of deviation.

In our recent examination of 63 hours of weather and flight track data from the DFW airspace (Rhoda and Pawlak, 1999a,b) we combined several weather variables (measurements, not forecasts) to correctly predict pilot deviation and penetration behavior for 70-85% of the encounters between thunderstorms and aircraft arriving at DFW and Dallas Love (DAL) airports. We also found that pilots were more likely to penetrate strong precipitation when they: 1) were near the arrival airport, 2) were following another aircraft, 3) were flying after dark, 4) had been delayed in the air by 15+ minutes upstream of the DFW airspace. We did not find any statistically significant difference between the percentages of thunderstorm penetrations by various airlines.

We also found that persistent penetration of storms near the airport is sometimes abruptly interrupted—presumably by wind shear alerts from air traffic controllers or cautionary pilot reports from the penetrating aircraft. When the arrivals cease, aircraft on the final approach course may turn suddenly to the left or right to avoid the weather that caused the interruption. Aircraft that abort the approach sometimes fly through very intense precipitation—sometimes through downdrafts that are causing microburst outflows at the surface.

The work described in this paper applies the methodology from the DFW study to data collected in the Memphis Terminal Radar Approach Control (TRACON). The methodology is described briefly here and in more detail in (Rhoda and Pawlak, 1999b). We developed several probability of deviation classifiers using a portion of the Memphis data and tested them on the remaining data to determine if it is possible to predict whether pilots will penetrate or deviate around the storms. We also tested the classifiers that were developed in the DFW study on the MEM data and vice versa. We repeated the DFW hypothesis tests for various dichotomies of encounters: near/far, leading/following, light/dark, delayed/undelayed.

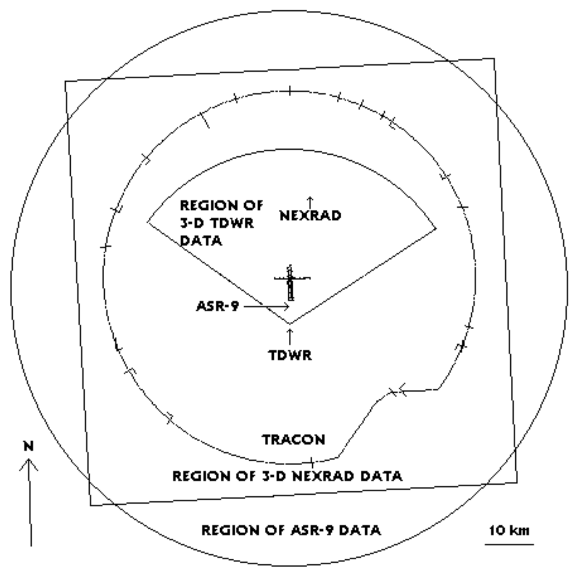

Figure 1. Memphis TRACON and sensor coverage regions for the purposes of this study.

2. APPROACH

Aircraft position data and weather data were collected in the Memphis TRACON for 14 thunderstorm days. The weather data were extracted from parts of the storms that the planes penetrated and from the parts of the storms that clearly caused aircraft to deviate. A pattern classification software package was employed to determine which combinations of weather variables best explain the pilots' penetration and deviation behavior. Several statistical classifiers were trained and tested to assess their suitability for generating a map of pilots' probability of deviating around storms. Finally, several hypotheses were tested regarding the correlation of non-weather variables to the penetration and deviation behavior.

---

* This work was sponsored by the Federal Aviation Administration under Air Force Contract No. F19628-95-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the U.S. Government. Corresponding author address: Dale Rhoda, Massachusetts Institute of Technology, Lincoln Laboratory, 244 Wood Street, Lexington, MA 02420-9185; e-mail: daler@ll.mit.edu

American Meteorological Society, Ninth Conference on Aviation, Range, and Aerospace Meteorology
Orlando, FL, September 11-15, 2000

3. WEATHER SENSORS AND VARIABLES

Weather data were obtained from the Memphis prototype of the Integrated Terminal Weather System (ITWS) (Evans, 1994). Specifically, data were collected from one fan-beam Airport Surveillance Radar (ASR-9), one pencil-beam Terminal Doppler Weather Radar (TDWR), and one pencil-beam Next Generation Weather Radar (WSR-88D or NEXRAD) as well as the National Lightning Detection Network (NLDN). Figure 1 shows the outline of the Memphis TRACON along with the radar locations and coverage regions. The meteorological products derived from the sensor data were:

3.1 Airport Surveillance Radar (ASR-9)

- Precipitation with false echoes due to Anomalous Propagation (AP) removed (NWS six-level VIP scale)
- Percent of each quadrant of the ASR coverage covered in level 2 or higher precipitation
- Percent of each quadrant of the ASR coverage covered in level 4 or higher precipitation

3.2 Next Generation Weather Radar (NEXRAD)

- 3-D radar reflectivity (dBZ)

The following 2-D products were computed for each 1km x 1km column of 3-D reflectivity:

- Maximum reflectivity
- Height of the center of mass of reflectivity
- Highest altitude of significant radar return (echo top)
- Lowest altitude of significant radar return (echo bottom)
- Vertically integrated liquid water (VIL)

3.3 Terminal Doppler Weather Radar (TDWR)

- Microburst strength
- Number of microburst alerts generated for all runways in the (1, 5 & 10) minutes preceding the encounter with the thunderstorm
- Number of windshear alerts generated for all runways in the (1, 5 & 10) minutes preceding the encounter with the thunderstorm
- 3-D radar reflectivity

The following 2-D products were computed for each 1km x 1km column of 3-D reflectivity:

- Maximum reflectivity
- Height of the center of mass of reflectivity
- Highest altitude of significant radar return (echo top)
- Lowest altitude of significant radar return (echo bottom)
- Vertically integrated liquid water (VIL)

3.4 National Lightning Detection Network (NLDN)

- Flashrate of cloud-to-ground lightning strikes

4. FLIGHT TRACK AND DELAY DATA

4.1 Flight Track Data

Flight track data were recorded from the Automated Radar Terminal System (ARTS) in the Memphis TRACON. The data included both flight plans and aircraft positions within 60 nautical miles of the airport. Aircraft positions are updated every five seconds and were collected with the ASR-9 radar. Data were collected for both arriving and departing flights. This paper only discusses the flights arriving at Memphis International Airport. At the time of this writing, the departure data are being analyzed in a manner similar to the work described here. The data for each flight were post-processed to compute:

- Arrival fix
- Runway
- Flight path length inside the TRACON
- Arrival time
- Range from the airport at the time of the storm cell encounter

4.2 Airline Service Quality Performance (ASQP) Delay Data

Major US airlines each submit a monthly report to the U.S. Department of Transportation listing every scheduled domestic flight, its scheduled departure and arrival times, actual push-back and gate-arrival times, as well as the takeoff and landing times. These data were combined with the flight plan and flight track information to compute whether or not the flights were delayed in the air before they arrived in the Memphis airspace.

5. DATA PROCESSING AND REDUCTION

The data for storm cell penetrations were processed in a different way than the data for storm cell deviations.

5.1 Penetration Processing

An algorithm examined each flight and searched for instances where the aircraft entered weather that exceeded the penetration threshold for one or more variables. To avoid biasing the study results toward one sensor, there was a penetration threshold for one variable from each of the weather radars: ASR precipitation (threshold of 2 on the NWS 6-level VIP scale); TDWR VIL and NEXRAD VIL (threshold of 2 kg/m^2 for VIL). A penetration was defined as the sequence of observations of the aircraft for which one or more of the penetration variables exceeded its threshold. (Note that penetrations of NWS VIP level 0 and 1 were ignored in this study UNLESS those regions had VIL values that exceeded the threshold of 2 kg/m^2. There were several penetrations with VIL greater than 2 and VIP of 0 or 1 in the cone-of-silence directly above the ASR-9.)

Some aircraft penetrated multiple storms. Each penetration of a storm cell consisted of multiple aircraft observations. Each penetration was reduced to a single “encounter observation” that consisted of one representative value for each of the weather variables. For all variables except echo bottom and center of mass, the encounter observation is the maximum value from each of the penetration observations. The echo bottom value is the minimum value and the center of mass value is the median value from the penetration observations.

5.2 Deviation Processing

It is difficult to have software automatically identify aircraft that deviate from their intended flightpath due to weather. In this study, a human analyst reviewed sequences of animated images of weather and flight track data and judged which aircraft deviated around weather. The analyst used software to draw a box around the weather which was believed to cause the deviation.

Subsequent analysis software extracted all of the weather variables from all of the x,y,z locations in the box at the time of the encounter and computed "encounter observations" by taking the minimum echo bottom value, the median center of mass value, and the maximum value for each of the other variables in the entire 3-D region of airspace enclosed by the box.

There were multiple instances of aircraft penetrating weather while deviating around other weather. The two categories are not mutually exclusive. These instances were treated as two separate encounters. The weather variables for the penetration were extracted from the Cartesian bins that were occupied by the aircraft and the variables for the deviation were extracted from the bins that the plane avoided—those enclosed by the human analyst's box.

6. DATA CASES

The data set consisted of 77 hours of weather and aircraft data from 14 different days between January and July of 1999. Approximately 1,935 aircraft entered the MEM airspace with the intention of landing at MEM. During that period 943 of those aircraft had a total of 1,936 encounters with storm cells—1,028 were deviations and 908 were penetrations.

Figure 2 shows a histogram of the number of penetrations and deviations as a function of ASR VIP level for all 1,936 encounters. Keep in mind that level 1 precipitation was essentially ignored in the study unless

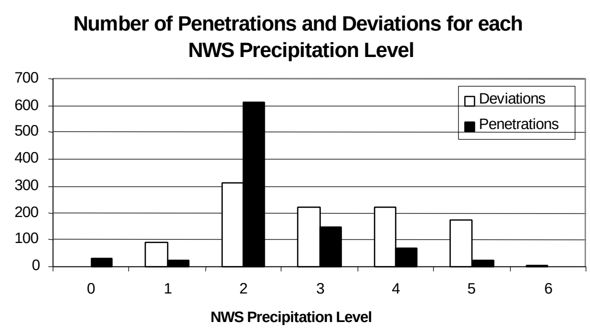

Figure 2. Penetrations and Deviations vs. NWS Precipitation Level.

the VIP values were significant in those regions. There were many penetrations of level 2 weather. The number of deviations was larger than the number of penetrations for level 3 and higher weather. This corresponds well to the findings of the DFW study and to the air traffic controllers' oft-quoted rule-of-thumb that pilots begin to deviate when the weather reaches level 3 or greater. Figure 3 shows the number of penetrations and deviations as a function of range from the airport. Figure 4 shows the intensity of the weather that was penetrated as a function of range from the airport.

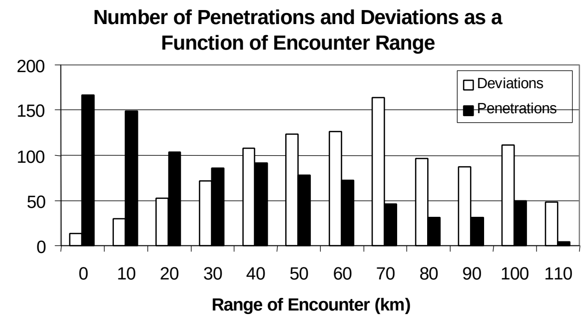

Figure 3. Penetrations and Deviations vs. Range.

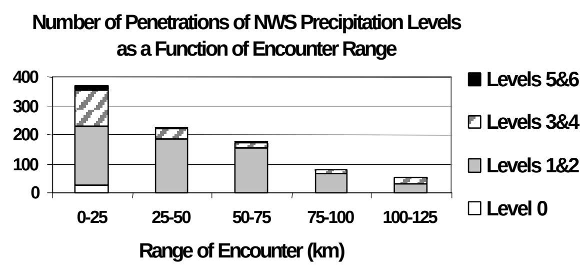

Figure 4. NWS Precipitation Level vs. Range for Penetrations

7. STATISTICAL ANALYSIS

This section of the paper considers which weather variables are correlated with the pilots penetration and deviation behavior. We do not consider flight-related variables because it would be highly desirable to design a classifier that could reliably compute the probability of deviation without needing to know what types of aircraft are flying around, which airlines they belong to, or what order they are flying in. It turns out that weather variables alone describe most of the variation in pilot behavior. The forthcoming project report describing this work will discuss the pros and cons of adding non-weather variables to probability of deviation classifiers.

The statistical analysis of the weather variables in this study was done in two parts. The data were randomly split into two groups: two-thirds of the encounters were assigned to the "feature selection and classifier training" group; one-third of the encounters were assigned to the "classifier testing" group. First, the training data were analyzed to find the combination of variables that best explain the variation in penetration and deviation behavior. Second, several statistical classifiers were trained and then tested on data that was not used in the classifier training exercise.

7.1 Feature Selection

A software package named LNKnet (Lippmann, 1993) was used to perform a forward and backward search through the weather variables employing a k-nearest-neighbors, leave-one-out technique to assess the explanatory power of various combinations of input variables or "features." The Memphis results corroborated the DFW findings that the penetration and deviation behavior is well correlated with three types of variables: storm intensity, range from the airport, and the horizontal extent of precipitation coverage.

Table 1 lists three storm intensity variables and the percentage of variation in the training data set that each variable explains by itself. Both range and weather coverage variables (percent of ASR quadrant covered in level 2+ and level 4+ precipitation) also added explanatory power to the analysis. Table 1 also shows the total percentage of variation explained with the addition of those variables.

Table 1. Percentage of variation in the training data set that each storm intensity variable explains individually and with range and weather coverage variables.

	DZ	ASR	VIL
Storm intensity	71	65	62
Add range	77	70	70
Add weather coverage	82	77	77

The 3-D radar reflectivity variable has the most explanatory power. It would not be practical, however, to base a classifier on 3-D reflectivity. In order to be helpful to air traffic planners and automation systems, a probability-of-deviation classifier would need to run on a forecast weather product; the system would need to forecast probability-of-deviation out 20-30 minutes into the future. The technology to accurately forecast 3-D storm structure 30 minutes into the future simply does not exist at this time. For the forseeable future a probability-of-deviation classifier will need to use a 2-D representation of storm intensity.

7.2 Statistical Classifier Training and Testing

Several neural net classifiers were trained using a storm intensity variable, range, and the weather coverage variables. The classifiers were trained on two-thirds of the encounters and then tested on the remaining one-third. In the testing phase of the analysis, each classifier generated an estimate of the probability-of-deviation. If that value was greater than or equal to 50% then the encounter was predicted to be a deviation; if the value was less than 50% then the encounter was predicted to be a penetration. To score the classifiers, the predictions were compared with the actual pilot behavior. Table 2 lists the percent of encounters that were incorrectly classified for the classifiers with different storm intensity variables. Because the test data set was independent of the feature selection and training phases of the analysis, the results of the two-thirds training and one-third testing analysis yield our best estimates of the error rates that would be achieved by using the classifier on data from additional days.

Table 2. Percentage of encounters incorrectly classified for various combinations of input variables. Train on 2/3; test on 1/3

Using a storm intensity variable, percent of the ASR quadrant covered in level 2+ weather, percent of the ASR quadrant covered in level 4+ weather, and range.

	DZ	ASR	VIL
DEV	17	20	22
PEN	11	17	14
Total	14	19	18

The data were also split into fourteen non-random subsets—one for each day in the study. A classifier was trained on thirteen days of data and tested on the remaining day for each of the fourteen combinations of thirteen days/one day. Table 3 lists the average percentage of test data points that were incorrectly classified. Finally, to assess the applicability of classifiers from one airport to data at another airport, classifiers were trained on the entire DFW data set and tested on the entire MEM data set. Then the process was reversed. Table 4 lists the percentages of test data which was incorrectly classified. The MEM classifiers perform very well on the DFW data.

Table 3. Average percentage of MEM encounters incorrectly classified for various combinations of input variables. 14 permutations - train on 13 days; test on 1 day.

Using a storm intensity variable, percent of the ASR quadrant covered in level 2+ weather, percent of the ASR quadrant covered in level 4+ weather, and range.

	DZ	ASR	VIL
DEV	13	18	19
PEN	20	28	28
Total	15	20	21

Table 4. Percentage of encounters incorrectly classified using DFW and MEM classifiers.

Using a storm intensity variable, percent of the ASR quadrant covered in level 2+ weather, percent of the ASR quadrant covered in level 4+ weather, and range.

	Train on DFW / Testa on MEM			Train on MEM / Test on DFW		
	DZ	ASR	VIL	DZ	ASR	VIL
DEV	28	28	43	6	19	10
PEN	7	15	6	12	16	16
Total	18	22	25	10	17	14

8. CONTROLLERS' RULE OF THUMB

Air traffic controllers often say that pilots generally fly through weather that shows up as level 1 or 2 on the ASR-9 and they begin to ask for deviations when the storms reach level 3 (41 dBZ). A simple binary tree classifier was trained to determine whether our data set validates that heuristic rule of thumb. LNKnet determined that if ASR precipitation is the only variable available, the lowest error rate is indeed achieved by classifying all encounters with levels 1 and 2 weather as penetrations and all encounters with level 3+ weather as deviations. The resulting classifier incorrectly classifies 33% of the 1,936 encounters in the data set. All of the classifiers listed in Tables 2-4 perform better than the simply ASR classifier by taking advantage of information about range, weather coverage, and/or pencil-beam estimates of storm intensity.

9. HYPOTHESIS TESTS WITH NON-WEATHER VARIABLES

In the DFW study, several hypothesis tests on non-weather variables yielded results that were statistically significant. Those hypothesis tests were performed on the MEM data, too.

9.1 Range from Airport

Pilots in this data set were more likely to penetrate strong precipitation as they drew near the destination airport. Near the airport (<25 km), 75% (138/183) of the encounters with level 3 and higher weather resulted in penetrations. Farther from the airport (25+ km) only 15% (104/684) of the encounters with level 3+ weather resulted in penetrations. This result is statistically significant at the 99% confidence level and corroborates the result from DFW.

Upon hearing the DFW results, several pilots have offered explanations for this result. Workload in the cockpit is very high as the aircraft approaches the airport. There is less lateral leeway in the flight path as the plane approaches the runway. Airborne radar is subject to ground clutter as the plane approaches the ground. Pilots report that they put a great deal of emphasis on weather information from air traffic controllers (i.e., windshear and microburst alerts) and from the pilots of preceding planes.

One difficulty with the practice of penetrating the storms until there is windshear or an unpleasant pilot report (PIREP) is that as the weather on the runways becomes more severe, the weather <u>near</u> the runways is likely to become more severe as well. If normal approaches are suddenly interrupted by a windshear alert or an alarming PIREP then the pilots that are lined up on final approach usually abort their landings and turn left or right. This sharply increases the both the pilot and the air traffic controllers' workloads. Pilots must "clean up" their aircraft configurations so the planes will fly rather than land and they are busy flying the vectors being issued by the air traffic controllers. The controllers focus on the task of separating aircraft and establishing an approach route to other runways or moving the aircraft to airspace that is free of thunderstorms. Air traffic controllers are not equipped to provide warnings about severe weather to the left or right of the nominal approach path. The pilots who break off the approaches are not likely to have time to manipulate the controls on their airborne radars for optimal assessment of precipitation intensity. Finally, there are rarely any recent PIREPs from the airspace beside the nominal approach path. Pilots who break off their approaches sometimes fly into regions where the weather is as intense or more intense than the weather that caused them to abort the landing. This potential danger highlights the importance of anticipating the need for storm-free escape routes on final approach. In the MEM data set, there were several instances of missed approaches and aborted approaches flying into strong precipitation. The forthcoming project report describing this work will contain detailed descriptions of those scenarios.

9.2 Leaders vs. Followers

Each aircraft encounter was classifed by a human analyst as either a "leader" or a "follower." Leaders were defined to be aircraft who flow along a route that had not been used by a preceding aircraft for at least ten minutes. Followers were aircraft that flew along a route that had been used by another aircraft within the preceding ten minutes. One can imagine that followers might have access to a PIREP from a preceding aircraft. If ten minutes had passed since the last plane flew through the area then the likelihood of a relevant PIREP would be diminished. There were no statistically significant differences between the penetration behavior of leaders and followers in the MEM data set. Forty-seven percent (385/825) of the leaders penetrated storm cells. Forty-seven percent (523/1111) of the followers penetrated storm cells. Twenty-six percent of the leaders that encountered level 3+ storms (89/343) penetrated them. Twenty-nine percent of the followers that encountered level 3+ storms penetrated them.

9.3 Delayed Upstream vs. Not Delayed

A subset of the aircraft in this data set belonged to airlines that are required to report their on-time performance statistics to the Department of Transportation every month. Those statistics include scheduled flying time, and actual take-off time. Those numbers were used, along with the time that the aircraft were observed to first enter the ASR-9 field of view, to estimate whether the aircraft had sustained airborne delays upstream of the Memphis airspace. Aircraft that sustained more than 15 minutes of airborne delay were compared to those that had sustained fewer than 15 minutes of delay.

Delay statistics were available for 483 of the 1,936 encounters. Forty percent of the aircraft that were delayed 15+ minutes upstream of the Memphis airspace penetrated the storms that they encountered (17/43). Thirty-eight percent of the aircraft that were less delayed upstream penetrated the storms that they encountered (167/440). Seventeen percent of the planes that had been delayed 15+ minutes and encountered level 3+ precipitation penetrated the precipitation (3/18). Twenty-one percent of the planes that were not significantly delayed and encountered level 3+ precipitation, penetrated that precipitation (45/211). None of these differences were statistically significant.

9.4 Daylight vs. Dark

The encounters were split into those that occurred during daylight hours (1 hour before AM twilight to one hour after PM twilight) and those that occurred when it was dark (all other hours). Encounters at night were more likely to result in penetrations than encounters during daylight hours. Forty-four percent of the encounters during the daylight were penetrations (655/1486). Fifty-six percent of the encounters at night were penetrations (253/450). Encounters with level 3+ storms at night were more likely to result in penetrations than those in the daytime. Twenty-four percent of the daytime encounters with level 3+ storms were penetrations (161/673). Forty-two percent of the nighttime encounters with level 3+ storms resulted in penetrations (81/194). These differences were statistically significant at the 99% confidence level. The differences might indicate that pilots use the visual appearance of the storms during daylight hours to help make the deviation decision.

9.5 Airline by Airline

The flight plan information from the ARTS system included airline and flight identification information. Three airlines had 100+ encounters in the data set; one major passenger-carrying airline, one commuter or "feeder" airline, and a major package-carrying airline. Their percentages of penetrations of all storms were 39%, 49%, and 55% respectively (160/406, 195/400, 127/233). Their percentages of penetrations of level 3+ storms were 21%, 26%, and 35% respectively (39/188, 47/178, 40/113). The commuter airline and the package-carrier were more likely to penetrate storms than the major passenger-carrier; the confidence level on those differences is 99%. The package carrier was more likely to penetrate storms than the commuter airline but the difference was NOT statistically significant at the 99% confidence level.

With regard to the propensity to penetrate level 3+ storms, the difference between the passenger-carrier and the commuter airline was not statistically significant at the 99% level. Nor was the difference between the commuter and the package-carrier. The difference between the package-carrier and the passenger-carrier, however, was statistically significant at the 99% level. It is impossible, however, to separate this difference from the daytime and nighttime difference because the vast majority of the package-carrier encounters occurred after dark and the vast majority of the passenger-carrier encounters occurred during daylight.

10. CONCLUSION

The data in this study once again affirm the rule-of-thumb that, far from the airport, aircraft generally deviate around level 3+ storms in the terminal area. To predict whether pilots will penetrate storms or deviate around them it is helpful to incorporate information about range from the airport, weather coverage, and a storm intensity estimate from pencil-beam radars. The classifiers that were trained on data from Memphis performed well on data from Dallas and vice versa.

The Memphis data corroborate the Dallas study in the findings that pilots are more likely to penetrate strong precipitation near the airport and after dark. The consistent pattern of apparently uneventful penetration of strong precipitation near the airport indicates that 1) systems that detect and predict wind shear near the airport are very important for pilots' situational awareness, and 2) although many aviation hazards are associated with regions of strong precipitation, radar reflectivity alone is not an adequate indication of hazard level.

The Memphis data did not corroborate the Dallas findings that pilots are more likely to penetrate if following another aircraft or if they were delayed upstream during the current leg of flight.

Opportunities for future work include: analysis of departure behavior (the Memphis departure data are currently being examined); analysis of storm penetration and deviation behavior in the en-route regime; analysis of storm cell encounters elsewhere in the country; interaction with airborne radar manufacturers to ensure the best possible information for pilots on final approach; comparison of recorded airborne weather radar data with data collected from ground-based sensors; examination of the DFW and MEM data where the unit of observation is not the individual storm cell encounter but rather the status of traffic flow along individual TRACON arrival routes—is the traffic flowing normally or are there deviations from the route; interaction with the pilot community to understand the penetration / deviation decision on final approach; examination of aborted approach scenarios in the presence of weather.

REFERENCES

Evans, James E. and E.R. Ducot, 1994: "The Integrated Terminal Weather System (ITWS)" The Lincoln Laboratory Journal, vol 7, number 2, 1994, pp. 449-474.

Lippman, Richard P., 1993: L. Kukolich, and E. Singer, "LNKnet Neural Network, Machine-Learning, and Statistical Software for Pattern Classification," The Lincoln Laboratory Journal, vol 6, number 2, 1993, pp. 249-268.

Rhoda, Dale A. and M.L. Pawlak, 1999a: "The Thunderstorm Penetration/Deviation Decision in the Terminal Area," American Meteorological Society, Eighth Conference on Aviation, Range, and Aerospace Meteorology, Dallas, TX, 10-15 January 1999, pp. 308-312.

Rhoda, Dale A. and M.L. Pawlak, 1999b: "An Assessment of Thunderstorm Penetrations and Deviations by Commercial Aircraft in the Terminal Area," MIT Lincoln Laboratory Technical Report, NASA/A-2, 1999.

Theriault, K.E., M.M. Wolfson, B.E. Forman, R.G. Hallowell, M.P. Moore, R.J. Johnson, Jr., 2000: "FAA Terminal Convective Weather Forecast Algorithm Assessment," *AMS Ninth Conference on Aviation, Range, and Aerospace Meteorology* (this volume), Orlando, FL, 11-15 September 2000.