THE THUNDERSTORM PENETRATION / DEVIATION DECISION
IN THE TERMINAL AREA*‡

Dale A. Rhoda and Margo L. Pawlak
MIT/Lincoln Laboratory
Lexington, Massachusetts

1.0   INTRODUCTION

During thunderstorm periods, terminal air traffic planners make a number of key decisions. They decide when to close and re–open arrival fixes, departure fixes, and runways; they anticipate and execute changes in runway configuration; they negotiate routing and flow rate decisions with Air Route Traffic Control Center (ARTCC) traffic managers; and they set the airport acceptance rate. In making each of these decisions, the traffic planner looks at a weather radar display and makes an educated guess at answering the two following questions:

- What will the weather be like in the airspace and time period in question?
- Will the pilots be able and willing to fly through that airspace during that time?

The same two questions will be important for advanced terminal automation systems. One key element of air traffic automation systems such as the Center–TRACON Automation System (CTAS) is the calculation of candidate trajectories for each aircraft for the time period of automation control. To make this calculation, the automation software must know which routes will be usable during the control period.

The first of the two fundamental questions is being addressed by the convective weather Product Development Team (PDT) of the FAA's Aviation Weather Research program. (Wolfson, 1997; Wolfson, 1999; Hallowell, 1999; Forman, 1999; Evans, 1997) The second fundamental question is the subject of the work reported here.

The state of the art answer to the second question is a widely quoted air traffic control rule–of–thumb which says that pilots generally do not penetrate precipitation that is NWS VIP level 3 (i.e. 41 dBZ) or higher. That is not to say that air traffic controllers always vector aircraft around lev-

Corresponding author address:
Dale A. Rhoda,
MIT Lincoln Laboratory
244 Wood Street, Lexington, MA 02173
email:daler@ll.mit.edu
phone: (617) 981–0848 fax: (617) 981–0632

el 3+ cells but rather that they begin to anticipate pilot requests for deviations when the weather approaches level 3. A suite of new weather sensors have become available that provide much more comprehensive information on convective weather features than was available in the past. Additionally, flight–related data such as preceding pilot behavior and whether a flight is running late are easier to obtain than in the past. In this study we develop an objective quantitative assessment of which weather and flight–related variables best explain pilot deviation decision–making.

2.0   APPROACH

The study's approach is as follows: Aircraft position data and weather data were collected in the Dallas–Fort Worth Terminal Radar Approach Control (TRACON) for several 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 / deviation behavior. Several statistical classifiers were trained and tested to assess their suitability for generating a map of pilots' probability of deviation around weather. Finally, several hypotheses were tested regarding the correlation of flight–related variables to the penetration / deviation behavior.

3.0   WEATHER SENSORS AND VARIABLES

Weather data were obtained from three fan beam Airport Surveillance Radars (ASRs), two pencil beam Terminal Doppler Weather Radars (TDWRs), and one pencil beam Next Generation Weather Radar (NEXRAD) in the Dallas–Fort Worth TRACON as well as the National Lightning Detection Network (NLDN). The meteorological sensors and the products derived from them were as follows:

Airport Surveillance Radar (ASR–9)
- Precipitation (NWS six–level VIP scale)
- Percent of each quadrant of the TRACON covered in level 2 or higher precipitation
- Percent of each quadrant of the TRACON covered in level 4 or higher precipitation

Next Generation Weather Radar (NEXRAD)
- Probability of severe hail
- Mesocyclone detection
- Tornado detection
- 3–D Radar reflectivity (DZ)

Additionally, the following products were calculated for each vertical column of reflectivity:

– Maximum reflectivity
– Height of the maximum reflectivity
– Height of the center of mass of the reflectivity
– Highest altitude of significant radar returns (echo top)
– Lowest altitude of significant returns (echo bottom)
– Vertical extent of region with significant radar returns
– Vertically integrated liquid water content (VIL)

Terminal Doppler Weather Radar (TDWR)
– Microburst detection
– Gust Front detection
– 3–D Radar reflectivity (DZ)

Additionally, the following products were calculated for each vertical column of reflectivity:

– Maximum reflectivity value
– Height of the maximum reflectivity
– Height of the center of mass of the reflectivity
– Highest altitude of significant radar returns (echo top)
– Lowest altitude of significant returns (echo bottom)
– Vertical extent of region with significant radar returns
– Vertically integrated liquid water content

National Lightning Detection Network (NLDN)
– Flashrate of cloud–to–ground lightning strikes

4.0 FLIGHT TRACK AND DELAY DATA

ARTS Flight Track Data

Flight track data were recorded in real–time from an ARTS at the Dallas–Fort Worth (DFW) airport. The data included both flight plans and aircraft positions within the terminal area. Aircraft positions updated every five seconds. Data were collected for both arriving and departing flights. This study only considered the flights arriving at DFW or Dallas Love (DAL) airports. The data were post–processed to automatically compute:
– Arrival fix
– Runway
– Flight pathlength inside the TRACON
– Arrival time
– Range of the encounter from the airport

Airline Service Quality Performance (ASQP) Delay Data

The major airlines each submit a monthly report to the U.S. Department of Transportation listing every scheduled domestic flight and its scheduled departure time, scheduled arrival time, actual departure time, wheels–off time, wheels–on time, and arrival time. These data were used in a hypothesis test described in section 10.2.

5.0 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 NWS six–level precip scale), TDWR VIL, and NEXRAD VIL (threshold of 2 kg/m^2). A penetration was defined as the sequence of observations for which one or more of the penetration variables exceeds its threshold.

Some aircraft penetrated multiple storms. Each penetration of a storm cell consisted of multiple weather 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 value is the maximum value from 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 are deviating from their intended flightpath. 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. Figure 1 shows the flight track of an aircraft deviating around a storm cell with level 4 ASR precipitation.

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 the minimum value for the echo bottom variable, the median value for the center of mass variable, and the maximum value for all other variables in the entire 3–D region of airspace that was avoided.

6.0 DATA CASES

The dataset consists of 63 hours of weather and aircraft data from nine different days during the spring and summer of 1997. Approximately 4300 aircraft landed at the DFW and DAL airports during that period and 1279 of those aircraft had a total of 1952 encounters with storm cells. Of the 1952 aircraft encounters with storm cells, there were 642 deviations and 1310 penetrations.

Figure 2 shows a histogram of the number of penetrations and deviations vs. ASR precipitation level for all 1952 encounters. Level 1 precipitation was essentially ignored in this study. There are many penetrations of level 2 and the number of deviations becomes larger than the number of penetrations for level 4 and higher weather. This corresponds well to the controllers' rule–of–thumb that pilots begin to deviate when the weather reaches level 3 or greater.
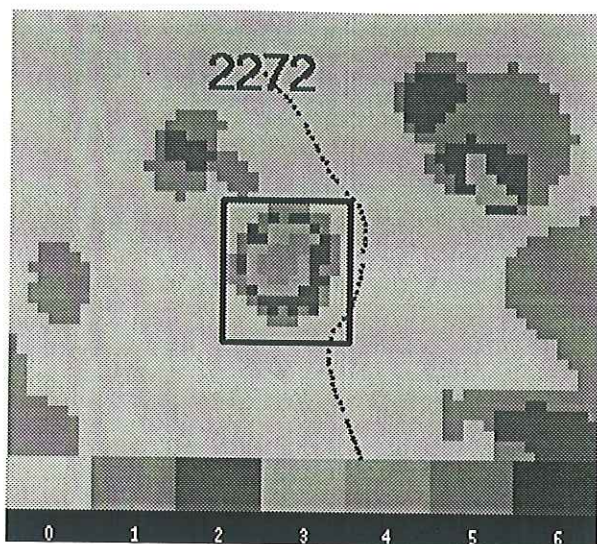
Figure 1. An aircraft deviating around a storm cell. An analyst has drawn a box around the weather that is believed to be responsible for the deviation.
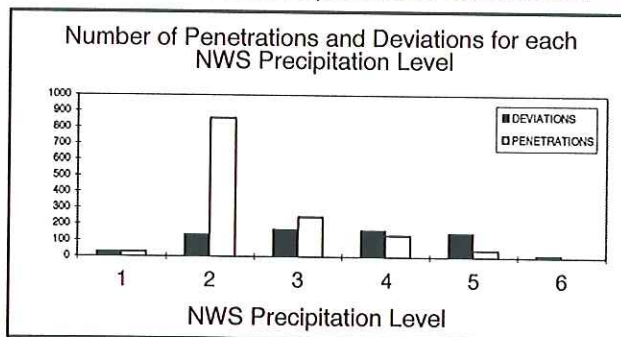

Figure 2. Penetrations and deviations vs. NWS VIP level

7.0   STATISTICAL ANALYSIS OF WX VARIABLES

This section of the paper considers which weather related variables are correlated with the pilots' penetration/deviation behavior. We do not consider any flight–related variables because it would be highly desirable to design a classifier that could reliably compute the probability–of–deviation without knowing 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 the dataset.

The statistical analysis of the weather variables in this dataset was done in two parts. First, the data were analyzed to find the combination of variables that best explained the variation in the penetration/deviation data. Second, several statistical classifiers were trained and then tested on data that was not used in the classifier training phase.

7.1.   Feature Selection

A software package named LNKnet (Lippmann, 1993) was used to perform a forward and backward search through the input variables employing a k–nearest–neighbors, leave–one–out technique to assess the ex-

planatory power of various combinations of input variables or "features." The best combination explained 94% of the variation in the penetration/deviation decisions using just five variables: range from the airport, pencil–beam radar reflectivity (DZ), ASR precipitation level (ASR), percent of the TRACON quadrant covered in at least low–intensity (level 2+) precipitation (QUAD_CVG_LO), and percent of the TRACON quadrant covered in high–intensity (level 4+) precipitation (QUAD_CVG_HI).

The radar reflectivity variable alone explained about 80% of the variation in the dataset. The magnitude of this variable's explanatory power is not surprising because most commercial aircraft are equipped with weather radar that indicates the presence of light, medium, and heavy rain. The ground–based radar reflectivity variable should be strongly correlated with the radar information that was available to the pilots at the time that they made the penetration/deviation decision.

It would not be practical, however, to design a probability–of–deviation classifier based on a 3–D radar reflectivity product. 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 predict the 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 foreseeable future, a probability–of–deviation classifier will need to use a 2–D representation of storm intensity. Therefore the LNKnet feature selection process was run a second time without the radar reflectivity variable.

Without considering the 3–D radar reflectivity variable, LNKnet found that 89% of the variation in the data can be explained by four variables: range from the airport, vertically integrated liquid water (VIL), QUAD_CVG_LO, and QUAD_CVG_HI. Seventy–four percent of the variation is explained by VIL alone. Again, it is not surprising that VIL has a great deal of explanatory power because VIL is computed solely from the 3–D radar reflectivity variable.

Broadly speaking, this analysis indicated that there are three categories of variables that are strongly correlated with penetration and deviation behavior: storm intensity, weather coverage in the surrounding region, and range from the destination airport. The dataset contains a number of variables that are estimates of storm intensity. Statistical classifiers were trained and tested using all of those variables. (Rhoda, 1998) This paper will discuss the performance of the classifiers that used DZ, VIL, and ASR. These variables correspond to the variable with the most explanatory power (DZ), the variable with the second–most explanatory power (VIL), and the only weather variable that controllers see on their operational displays (ASR).

7.2.   Statistical Classifier Training and Testing

LNKnet is capable of training neural networks, likelihood classifiers, nearest neighbor classifiers, rule–based classifiers, and committee classifiers. In the exploratory

phase of this project, nearest neighbor, rule–based, and neural network classifiers were trained and tested. The neural network classification technique out–performed the other two and was used throughout the remainder of the study. All of the neural net classifiers constructed in this study employed the range variable, the two weather coverage variables, and one storm intensity variable. Three separate storm intensity variables — DZ, ASR, and VIL — were used to train and test three separate classifiers.

The data were split randomly into thirds and DZ, VIL, and ASR classifiers were trained and tested on each permutation of two–thirds training and one–third testing. Finally the data were separated into nine subsets where each subset corresponded to a storm day in the dataset. For each day in the dataset, DZ, VIL, and ASR classifiers were trained on eight days of data and tested on the ninth. Table 1 lists the average error rates for the two–thirds/one–third splits. Table 2 lists the average error rates for the eight days/one day splits. The DZ classifier has the lowest errors, followed by the VIL and ASR classifiers. The low error rates imply that it may be possible to generate reliable maps of probability–of–deviation for use in air traffic management decision aid tools.

	DZ	VIL	ASR
DEV	12	19	29
PEN	3	5	7
TOTAL	6	10	15

Table 1. Average Percentage of Test Data Incorrectly Classified by Neural Net Classifiers. Train on 2/3; test on 1/3; three different permutations.

	DZ	VIL	ASR
DEV	11	20	28
PEN	5	7	14
TOTAL	7	11	18

Table 2. Average percentage of test data incorrectly classified by neural net classifiers. Train on eight days; test on one day; nine permutations.

8.0   CONTROLLERS' RULE–OF–THUMB

A simple binary tree classifier was trained to replicate the controllers' rule–of–thumb. LNKnet determined that if ASR precipitation is the only variable available, the lowest error rate is achieved by classifying all encounters with levels 1 and 2 weather as penetrations and all encounters with level 3+ precipitation as deviations — this finding corresponds exactly to the controllers' rule–of–thumb. Unfortunately the resulting classifier incorrectly classified 30% of the 1952 observations in this dataset. All of the classifiers described in section 7.2 that took advantage of information about the range of the encounter and weather coverage performed better than the rule–of–thumb classifier.

9.0   EXPLANATORY WEATHER VARIABLES

Again, this analysis indicated that there are three categories of variables that are strongly correlated with penetration and deviation behavior: storm intensity, weather coverage in the surrounding region, and range from the destination airport.

9.1.   Storm Intensity

The dependency on storm intensity is intuitive; pilots are trained to avoid intense convective activity.

9.2.   Weather Coverage

The dependence on weather coverage is also somewhat intuitive. If the storm being encountered is the only cloud in the sky then pilots often have plenty of latitude to deviate around the cell. If the area is covered in widespread precipitation then the pilot will likely have to fly through some precipitation to reach his/her destination so penetration becomes more likely. Finally, if the region is covered with strong convective activity, pilots tend to avoid the region altogether.

9.3.   Range From the Airport

Pilots in this dataset were more likely to penetrate intense thunderstorms as they drew nearer the destination airport. Near the airport (<20km), 91% of the encounters with heavy weather (NWS level 3+) resulted in penetrations (211/231). Farther from the airport (>20km), only 31% of the heavy weather encounters resulted in penetrations (212/678).

There are several possible explanations for this behavior and more research will be required to determine whether the finding is generally true and if so, why. The explanations fall into two broad categories: First, it is possible that when the pilots are near the airport at low altitude with a high workload level, they are less able to assess the severity of the weather and therefore they inadvertently penetrate significant weather. Second it is possible that the pilots are aware of the intensity of the weather and yet they are more willing to penetrate heavy weather near the airport. Both categories of explanation are discussed at some length in the technical report that describes this research. (Rhoda, 1998) Whatever the reason, the fact that pilots penetrate thunderstorms with level 3, level 4, and even level 5 precipitation near the airport, highlights the importance of wind shear detection and prediction systems that alert controllers and pilots to hazardous conditions near the runway.

10.0   FLIGHT–RELATED VARIABLES

There were several flight–related variables in the study that were well–suited for hypothesis tests that yielded interesting results.

10.1.   Leaders and Followers

The data in this study indicate a correlation with the behavior of preceding pilots. Aircraft that follow closely behind a preceding aircraft are more likely to penetrate heavy weather than aircraft that do not. In this study "lead-

ers" were defined to be aircraft who flew 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. Twenty–six percent of the leaders that encountered heavy weather in this dataset (79/298) penetrated the storms. Fifty–six percent of the followers that encountered heavy weather (344/610) penetrated the weather. When the analysis is restricted to encounters with heavy weather within 20 km of the airport, the percentages increase. Fifty–eight percent of the leaders penetrated the storms (21/36) and ninety–eight percent of the followers did so (190/194). The difference between the leaders' and followers' propensity to penetrate is statistically significant at the .01 level.

10.2. Aircraft Behind Schedule

Aircraft in this study first appear in the flight track data when they are 60 nautical miles from the DFW airport. From that point it typically takes 20 minutes to fly to the DFW or DAL airports. In this study, aircraft that arrived at the radar boundary having already flown 15 minutes longer than the scheduled flying time to that point in the trip (i.e. within five minutes of the scheduled flying time for the entire trip) were more likely to penetrate heavy weather than those that arrived earlier. Fifty–one percent of the encounters with heavy weather made by "late" planes (39/77) resulted in penetrations. Only 15% of the heavy weather encounters made by aircraft that were not "late" (79/531) resulted in penetrations. The difference between the early and late aircraft's propensity to penetrate is significant at the .01 level. [Note. The scheduled flying times were not available for all of the aircraft that encountered heavy weather.]

10.3. Various Airlines

There were seven airlines that had more than 20 encounters with light weather (level 1 or 2) in the dataset. There were no statistically significant differences in the airlines' propensity to penetrate or deviate around light weather.

There were six airlines that had more than 20 encounters with heavy (level 3+) weather in the dataset. There were no statistically significant differences in the airlines' propensity to penetrate or deviate around the heavy weather.

11.0 CONCLUSION

The data in this study affirm the air traffic controllers' rule–of–thumb. If the only variable available is the NWS six–level precipitation intensity then it is best to say that pilots tend to deviate around weather that is level 3 or higher. Range and weather coverage information improve the performance of the rule–of–thumb classifier. Even more accurate classifiers may be constructed with products from pencil–beam radars; both 3–D reflectivity (DZ) and vertically integrated liquid water (VIL) result in classifiers with lower error rates than the ASR classifiers. Other factors that seem to be correlated with penetration/

deviation behavior include leader/follower status and on–time status.

Initial analysis indicates that the storm intensity, range, and weather coverage variables yield promising results in candidate statistical classifiers. Given a reliable forecast of storm intensity, it should be possible to generate a forecast of the pilots' probability–of–deviation for use in traffic management decision–aid tools.

Pilots in this study penetrated some surprisingly intense weather near the airport. The reasons for this are not clear. The penetration/deviation decision near the airport warrants further analysis and research.

Opportunities for future work include: analysis of departing aircraft in the TRACON; analysis of aircraft in the en–route environment; analysis of storm cell encounters in other parts of the country; interaction with the pilot community to understand the penetration/deviation decision as the plane nears the airport; implementation of a candidate statistical classifier to generate probability–of–deviation maps; and interaction with the convective weather forecasting community to generate and evaluate forecast probability–of–deviation maps.

12.0 REFERENCES

Evans, James E.(1997), "Operational Problem of Convective Weather in the National Airspace System," AMS Convective Weather Forecasting Workshop; (Invited Presentation) Seventh Conference on Aviation, Range, and Aerospace Meteorology, 2–7 February 1997, Long Beach, CA.

Forman, Barbara E. (1999), R.G. Hallowell, M.P. Moore, M.M. Wolfson, J.E. Evans, "User Needs for Convective Weather Forecasts," *Present conference.*

Hallowell, Robert G. (1999), M.M. Wolfson, B.E. Forman, M.P. Moore, B.A. Crowe, T.M. Rotz, D.W. Miller, "The Terminal Convective Weather Forecast Demonstration," *Present conference.*

Lippmann, Richard P. (1993) , L. Kukolich, 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. (1998), M.L.Pawlak, "An Assessment of Thunderstorm Penetrations and Deviations By Commercial Aircraft in the Terminal Area," MIT Lincoln Laboratory Technical Report, NASA/ A–2, 1998.

Wolfson, Marilyn M. (1999), B.E. Forman, R.G. Hallowell, M.P. Moore, "The Growth and Decay Storm Tracker", *Present conference.*

Wolfson, Marilyn M. (1997), C.K. Mueller, M. Eilts, "Convective Weather Forecasting for FAA Applications", http://www.faa.gov/aua/awr/convwxpa.htm