6.1

THE TERMINAL CONVECTIVE WEATHER FORECAST DEMONSTRATION AT THE DFW INTERNATIONAL AIRPORT *‡

R.G. Hallowell[1], M.M. Wolfson[1], B.E. Forman[1], M.P. Moore[1], B.A. Crowe[1], T.M. Rotz[1], D.W. Miller[1], T.C. Carty[2], and S.F. McGettigan[3]

1 – MIT Lincoln Laboratory; Lexington, MA
2 – William J. Hughes Technical Center; Federal Aviation Administration; Atlantic City, NJ
3 – Raytheon Systems Company; Atlantic City, NJ

1. INTRODUCTION

The FAA Convective Weather Product Development Team (PDT) is tasked with developing products for convective weather forecasts for aviation users. The overall product development is a collaborative effort between scientists from MIT Lincoln Laboratory (MIT/LL), the National Center for Atmospheric Research (NCAR), and the National Severe Storms Laboratory (NSSL). As part of the PDT, MIT/LL is being funded to develop algorithms for accurately forecasting the location of strong precipitation in and around airport terminal areas. We began by consulting with air traffic personnel and commercial airline dispatchers to determine the needs of aviation users. Users indicated that convective weather, particularly line storms, caused the most consistent problems for managing air traffic. These storms are by far the major cause of aircraft delays and diversions.

MIT/LL has already developed the Integrated Terminal Weather System (ITWS) which combines a variety of near–airport sensors to provide a wide range of current weather information to aviation users. Raytheon is currently building the production ITWS system which will be deployed at 45 major airports by 2003. The initial capability ITWS already provides some convective weather predictive capabilities in the form of storm motion vectors and "Storm Extrapolated Positions" (SEP; leading edge of storm at 10 and 20 minutes). But ITWS users indicated a desire for enhanced forecasts which showed the full spatial extent of the weather, how the weather would change (grow or decay) and extended forecast time periods to at least out one hour. Our approach is to develop an algorithm which may be added as a future product improvement to the ITWS system.

Previous attempts at producing forecasts have focused on convective initiation and building from short–term (20–30 min) cell forecasts. Our "reverse time" approach of attacking longer time scale (60 min) features first is an outgrowth of addressing user needs and the discovery of improved tracking techniques for large scale precipitation features. The "Growth and Decay Tracker" developed by MIT/LL (Wolfson et.al., 1999) allows us to generate accurate short and long term forecasts of large scale precipitation features.

This paper details the Terminal Convective Weather Forecast (TCWF) demonstration ongoing at Dallas/Ft. Worth International Airport (DFW) and discusses the underlying algorithm being developed.

*This work was sponsored by the Federal Aviation Administration. The views expressed are those of the authors and do not reflect the official policy or position of the U.S Government.

‡Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Air Force.

†*Corresponding author address:* Robert G. Hallowell, MIT Lincoln Laboratory, 244 Wood Street, Lexington, MA 02420–9108; e–mail *bobh@ll.mit.edu*

2. ALGORITHM BACKGROUND

The design of the TCWF algorithm and the look and feel of the display were dictated by five main considerations gathered from user interviews before the experiment began (Forman et.al., 1999):

1. Most users expressed a desire for forecasts out to 60 minutes.
2. Forecasts should show the spatial extent of the weather. Users expressed a need to determine not only when weather would impact operations, but also when the weather would stop impacting operations.
3. The real weather needed to be clearly distinguished from the forecasted weather.
4. Users indicated a desire to see animated loops of the past and forecasted weather (as opposed to contoured regions or overlaid forecasts).
5. Users required more or less forecast accuracy depending on their role. The amount of accuracy needed in time and space increased as the airspace being controlled approached the runways.

3. ALGORITHM METHODOLOGY

The initial system uses NEXRAD wide–band radar data to create 2–D Cartesian files of 1km and 4km full resolution VIL (Vertically Integrated Liquid water) products using the NEXRAD algorithm. VIL images are converted into interest images representing the standard VIP level using an equation developed by Lincoln Laboratory (Troxel & Engholm, 1990). This conversion is done to facilitate the future incorporation of TDWR and ASR–9 weather channel data into the base precipitation images. ASR weather is only captured in 6–level values and, therefore a common mapping of value to interest is needed for each radar included in the analysis. Table 1 shows the relative values of VIL, VIP and the corresponding VIL "interest" value.

Table 1. Conversion values between VIL, VIP level and VIL interest values.

VIL (Kg/m^2)	NWS "VIP" level (ASR reflectivity)	VIL Interest Level
0.14	1 (18 dBZ)	42
0.76	2 (30 dBZ)	84
3.50	3 (41 dBZ)	128
6.90	4 (46 dBZ)	171
12.0	5 (50 dBZ)	213
32.0	6 (57 dBZ)	255

The next process generates "large scale" images utilizing an elliptical filter designed to give high scores to large elongated features within the image while diminishing smaller scales. The "large scale" filtered images are sent to the Growth & Decay tracker which generates pixel–by–pixel estimates of the motion field. As a necessary speed–up to processing, only the 4km images are used to generate vectors. The 1km vectors are created by super–sampling the 4km motion field. Eventually, we will run the 1km vectors and sub–sample to obtain the 4km field. Only one scale of vectors is used to create 1km and 4km forecasts because discrepancies between the two scales' vectors can lead to 1km and 4km forecasts being advected in different direction.

Finally, the terminal scale (1km grid out to 50 nm range) and regional scale (4km grid out to 200 nm range) vectors and full scale images are used to generate forecast images. A three–step process is used to obtain forecast images. First, the motion field is advected to the desired forecast time period by using the current vector estimates to "put" the vectors in time. In other words, the X and Y vector components are multiplied by the forecast time period to obtain the relative placement of the vector in the forecast image. Second, because the "put" of the vectors can leave holes in the data, the resultant vector field is smoothed using a 3x3 Cressman filter. Finally, the forecasted X and Y components of the vector field are used to "get" the VIL interest value which belongs in the forecasted pixel. The "get" estimates the offset to the desired interest location by reversing the X and Y components and multiplying by the time period desired.

As an aid to users, the forecast and VIL interest images are used to automatically score the 30 and 60 minute forecasts in realtime (scoring is detailed below). The forecast images and scoring information are then sent to the Web Page Display utility which prepares the images for the display.

4. FORECAST DISPLAY

The TCWF display is delivered to users through a web (HTML) page. The web interface allows a great deal of flexibility for presenting information, but in a form which is already familiar to many users (GIF images, hypertext links, etc). The main page for the TCWF web site is shown with annotations in Figure 1. The left panel allows users to select the type of forecast they would like to view in the display window. There are two ranges: 50 nm for examining weather in the terminal TRACON and, 200 nm for viewing weather outside the TRACON. Users may select animated movie loops of the 30–60 minute forecasts with or without past weather, or fixed forecast images of the 10, 20, 30, 40, 50, or 60 minute forecast.

The forecast images have two key components: the display window, and forecast accuracy (scoring). The past weather is shown using standard 6–level NWS colors, while the forecast is shown as two shades of yellow. The solid yellow (light grey in Figure 1), represents regions where there is a high probability that level 3+ weather will exist. The hatched light yellow (grey stippled) region indicates regions where there is a moderate probability that level 3+ weather will exist at the forecast time period. Users felt that two levels of probability gave them the flexibility for long–term planning (where even moderate probability regions could be avoided) or short–term operations (using the more restrictive high probability). The realtime scoring numbers are designed to give users an estimate of the current accuracy of the algorithm (details follow).

In addition, there are buttons which can give users on–line help, ftp (file transfer) capabilities, and links to the Lincoln Laboratory and FAA Aviation Weather Research web sites.

5. DFW DEMONSTRATION

The Dallas/Ft. Worth International Airport (DFW) was chosen as the site for the first Terminal Convective Weather Forecast experimental site for two main reasons. First, DFW is the busiest airport in the US (FAA,1998) and, as such, can benefit greatly from increased planning operations. Second, the MIT/LL already operates an ITWS prototype at DFW, making data access and analysis much more efficient than operating an independent site.

The experimental system was deployed for three main purposes:

1. Demonstrate the utility of up to 60 minute precipitation forecasts of long–lived convective linestorms. If the product is valuable it will be considered as a possible upgrade to the ITWS.
2. Demonstrate and improve the forecast display concepts to air traffic and airline users.
3. Develop an open–architecture algorithm to permit incremental development of terminal growth and decay forecasts.

Figure 2 illustrates the realtime connections to users for the DFW demonstration. The DFW web server is used to serve ATC users over the ITWS prototype intranet, while the MIT/LL web server (at Lexington, MA) serves airline and scientific users over the internet. The TCWF web page is a private web site, however background information on TCWF and other programs is available at *http://www.ll.mit.edu/AviationWeather.*

6. SCORING

Users requested a realtime indication of how well the forecast is doing, indicating it would help them decide when to use or not use the forecast. While it is difficult to accurately measure how well your current forecast is doing, it is easy to determine how well the forecast has done in the past hour. To obtain this measure, forecast images are scored, in realtime, against the

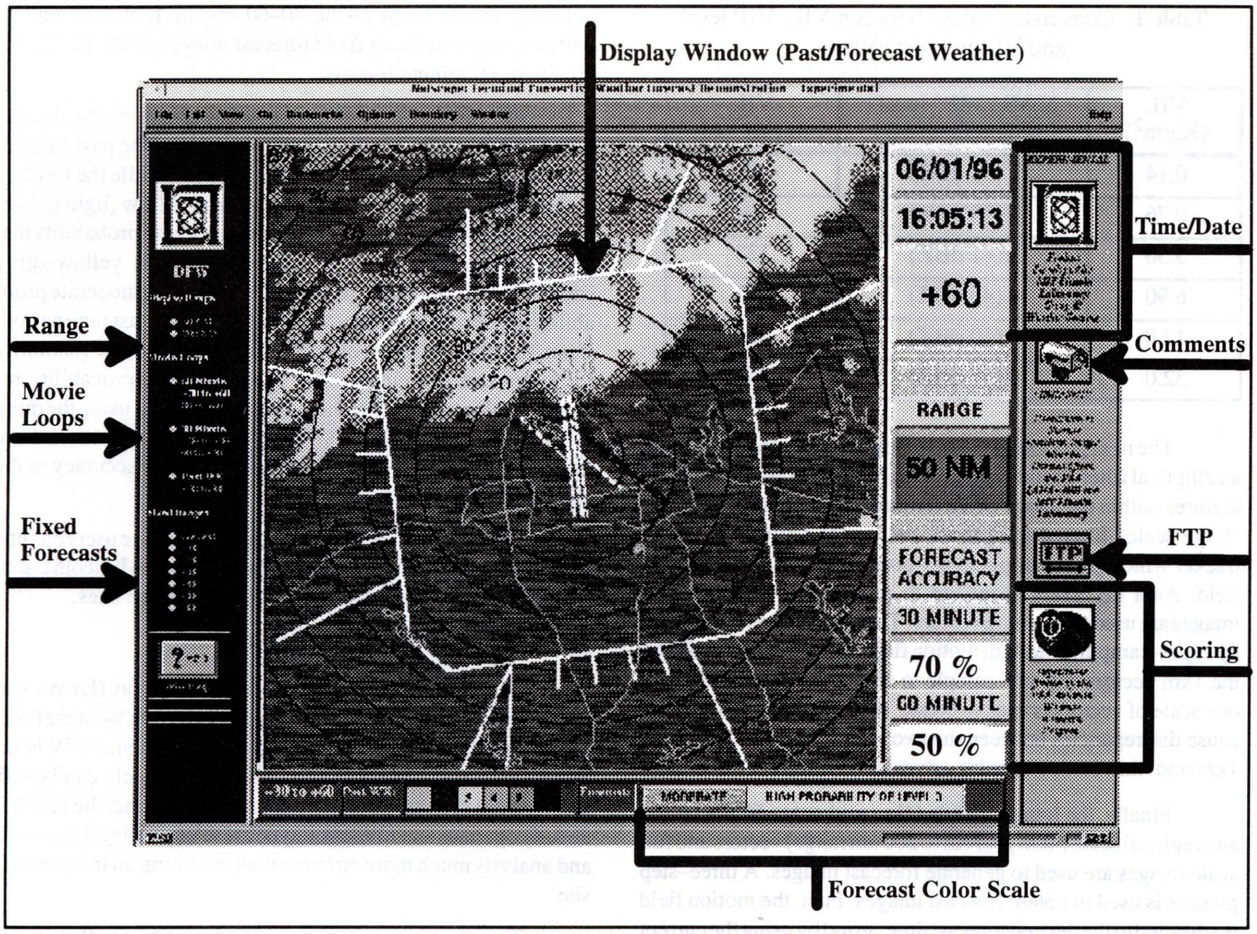


Figure 1. Annotated view of TCWF Web site. Example shows the 50 NM range (Terminal) 60 minute forecast.

"true" level 3+ weather image at the forecast time period. "Truth" for our algorithm is the actual area of level 3+ weather as depicted by the NEXRAD VIL (as shown in Table 1).

There are many ways to present scoring numbers, the simplest being binary or pixel–by–pixel scoring across the image. For example, Figure 3 shows a pixel map of of the binary score for a forecast region, and uses letters to indicate "hits" (correct forecasts), "misses" (failure to forecast), and "false alarms" (incorrect forecasts).

Binary scoring produces numbers that are easy to understand, but they do not reflect the *value* of the forecast for air traffic purposes. We can see from Figure 3, that the regions of misses and false alarms surround the region of hits. This could be due to a minor error in the storm motion estimate, or slight changes in the shape of the storm complex. Extensive discussions with users with ATC and airline personnel indicates there is some room for error around the margins of predicted cells. In general, users indicated that a 1–hour forecast which was within 5 nm should be considered a valid forecast.

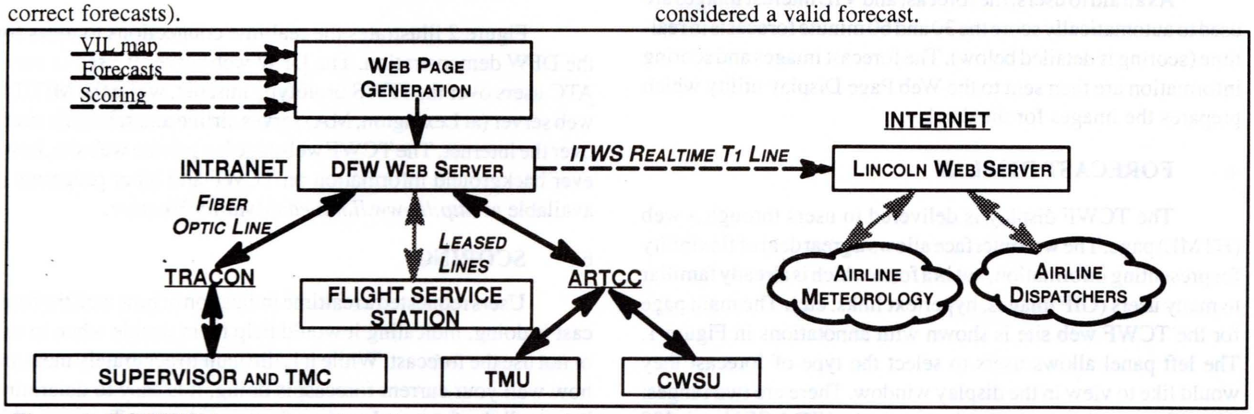


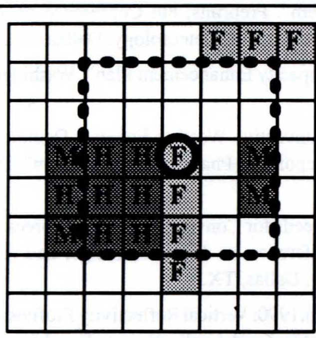
Figure 2. TCWF Web–page creation, display and dissemination.

Therefore, we implemented a "fuzzy" scoring methodology that looks in a 5x5 kernel (for 4km pixels, giving a range of 10 km or 5.4 nm in each direction) centered on the pixel being scored. For example, if a "high probability" level 3+ forecast is made in a pixel and no true level 3+ is present at that pixel, the forecast will not be counted as a false alarm if there is level true 3+ in any of the surrounding 5x5 pixels. Conversely, if a pixel does not contain a "high probability" forecast, but truth indicates there is level 3+ weather (a miss), the pixel will be registered as a "hit" if any forecast pixel within the 5x5 scoring kernel contains a level 3+ forecast. The 5x5 scoring kernel, as shown in Figure 3, omits the corners, so the quasi–circular patch consists of 21 total pixels. The 5x5 scoring column in Figure 3 shows the POD, PFA and CSI numbers for the example shown. These numbers, we feel, more accurately represent the value of the forecast to air traffic users.

The scoring numbers shown on the demonstration web page for the 30 and 60 minute forecasts, respectively, are average CSI scores. The Critical Success Index was chosen because its' value balances hits and false alarms, yielding a better measure of the value of the overall forecast. There is a time–lag equal to the forecast time period being presented, because forecasts cannot be scored until a truth image exists for the forecasted time period.

Minute–by–minute changes in scores are noisy and annoying to users. Therefore, we chose to present users with a 30 minute running average of the 30 and 60 minute CSI, quantized to 5% increments. While past performance does not necessarily indicate current performance, it is the best measure we have at the present time.

Users indicated that our scoring numbers were sometimes too conservative. The current scoring only scores the high probability forecast regions against "truth" regions of VIP level 3 or higher. Many of the users commented that our scores were unrepresentatively low during events where there was very little level 3, but the moderate probability forecast accurately portrayed the region of level 2 weather. In addition, air traffic controllers have indicated that convective weather forecasts within 5 nm **and** 10 minutes should be considered valid (Browne,1999). The introduction of temporal verification, in addition to the current



	BIN	5x5
POD	**63.6**	**86.7**
PFA	**50.0**	**18.7**
CSI	**38.9**	**72.2**

$$\mathrm{POD} = \frac{H}{H+M} \qquad \mathrm{PFA} = \frac{F}{F+H} \qquad \mathrm{CSI} = \frac{H}{H+M+F}$$

Figure 3. Example illustrating fuzzy scoring technique (H=Hit, F=False, M=Miss)

spatial verification will be introduced in the near future. We are looking into techniques for better matching our scoring numbers to the *value* perceived by the user.

7. RESULTS

The Terminal Convective Weather Forecast system has been operating at DFW since March 23rd, 1998. Unfortunately, the Dallas weather pattern was particularly dry during the 1998 summer convective season. As such, operational experience with large line storms (the type of weather the initial system was focused on) was limited. From March through early August there were only 6 days of large scale convective weather which were significant for aviation. There were sixteen other days where air–mass storms were present in the 200 nm range but either never impacted the 50 nm range or occurred during periods where traffic was not impacted. Using the scoring technique outlined above, the average POD, PFA and CSI scores for the 6 operationally significant days are shown in Table 2.

Table 2. 30 Minute Scores for DFW Convective Weather Days.

DATE	TYPE	POD	PFA	CSI
3/30/98	Line	81 %	8 %	76 %
4/26/98	Line	77 %	22 %	62 %
5/27/98	Line	78 %	24 %	62 %
6/04/98	Line	84 %	17 %	72 %
8/04/98	Air Mass	68 %	33 %	52 %
8/05/98	Air Mass	69 %	32 %	52 %

Table 3. 60 Minute Scores for DFW Convective Weather Days.

DATE	TYPE	POD	PFA	CSI
3/30/98	Line	62 %	17 %	57 %
4/26/98	Line	59 %	43 %	43 %
5/27/98	Line	61 %	44 %	42 %
6/04/98	Line	67 %	35 %	52 %
8/04/98	Air Mass	46 %	55 %	30 %
8/05/98	Air Mass	43 %	56 %	28 %

While the average POD for a day is one measure of performance, we note that the POD changes as the storm evolves. As an example, Figure 4 shows a time series of the 30–minute POD and PFA for the March 30th, 1998 linestorm. The POD at the start of the storm is low, but once the line storm has formed the POD stays relatively high.

8. USER FEEDBACK

The two main forms of user feedback were those relayed through the MIT/LL DFW ITWS prototype staff and, more formally, the evaluation done by the Willam J. Hughes Technical Center (FAA–TC,1998). Fortunately, we were able to begin op-

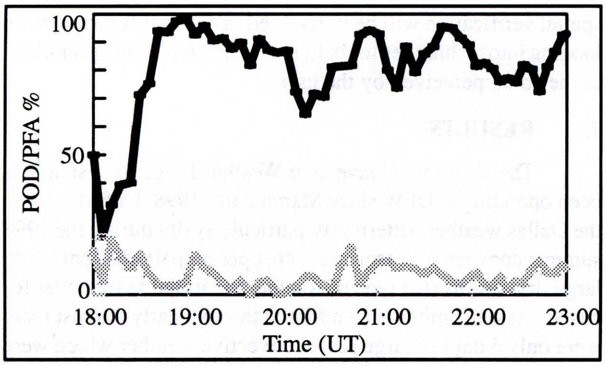


Figure 4. Time series plot of 30–minute POD (black) and PFA (grey)for 3/30/98 storms.

eration in the early spring when several line storms impacted DFW. Airline users, however, were not connected until after the early storms had occurred. Therefore, the user feedback is exclusively from the air traffic community. ITWS site staff were able to capture feedback on a number of occasions on the morning after weather events, while the Technical Center utilized a structured form and interview to analyze usage. Both forms of input are used below to broadly summarize user feedback.

Most users found the TCWF product useful and fairly easy to use. ATC users indicated that they frequently used the product during line storms to estimate when gates/runways would close or re–open. Due to the dry 1998 weather conditions, no quantitative benefits analysis could be done. However, ITWS site staff reported two specific days of delay benefit:

March 30th

A long thin line of heavy weather was oriented north–south and moved rapidly thru the TRACON throughout the afternoon. After the mission, the TRACON TMC said that the TCWF was indicating that the airport would be impacted about 30 minutes earlier than expected. The product was used throughout the mission, but the largest benefit was derived from knowing when the airport would be impacted and using that information to plan when arrivals would stop and departures would be held.

April 26th

A broad line of convective weather moved over the airport. At various points all the gates and runways were impacted by weather. The TRACON TMU personnel said the algorithm indicated that the storms would slow down as they neared the airport. Using this information, the TMU delayed closing each cornerpost a little longer than normal. Only the last two aircraft allowed to enter the TRACON prior to shutting off the gates were unable to land due to weather on the airport. The TMC claimed "We're able to make a more educated guess [as to when to shut off the cornerposts]. It put the heavy weather right at the right spot."

The TMC later estimated they would have had 7 aircraft unable to land without TCWF.

While users indicated they relied on the 30 minute forecast more than the 60 minute forecast, they still found the 60–minute forecast useful (particularly ARTCC users). The majority of users indicated that the forecast accuracy score matched their impression of the actual forecast accuracy. Most of the users who said the score did not match their impression, indicated the actual accuracy was better than the score.

Most of the users thought the 60 minute forecast was optimal, although ARTCC users expressed a desire for 90–120 minute forecasts as well. The 50 nm range was more frequently viewed than the 200 nm range. Users thought there was no driving need for a display range past 200 nm, but an intermediate 100 nm range at 1–2 km resolution would be useful.

Further studies are needed to determine specific benefits of the TCWF. The lack of even the normal volume of convective weather, and a limited number of ATC users (20), makes the comments made here useful but not conclusive.

9. CONTINUED DEVELOPMENT

The TCWF demonstration continues to run at DFW and past cases are being analyzed to increase the product usefulness and reliability. The 60 minute forecast is currently on schedule for a demonstrational validation test in 2000 (Table 4). For example, optimization of the tracking techniques have increased our 60–minute CSI scores by 7 percentage points on April 26th and June 4th over those shown in realtime (Table 3). Continued improvements are expected when a feature detector to predict large scale storm decay is added to the system this coming year.

Upgrades to the display are also planned to increase the overall usability of the system. In particular users indicated that the color contrast between moderate and high probability forecasts could be enhanced.

Table 4. Demonstration plan for TCWF system.

Year	Location	Prediction Time
1998	Dallas/Ft.Worth	30–60 min
1999	Dallas/Ft.Worth Orlando	30–60 min
2000	Dallas/Ft.Worth Orlando New York	60 min Dem/Val
Additional sites may be added if funding permits.		

REFERENCES

Browne, K.P., 1999: "Aviation Weather Requirements in the Air Traffic Management System". Preprints, 8th Conference on Aviation, Range, and Aerospace Meteorology, Dallas, TX.

FAA, 1998: "1998 Aviation Capacity Enhancement Plan", Washington, D.C., 1998.

FAA–TC, 1998: "Terminal Convective Weather Forecast Demonstration Quick Look Report For Phase I Data Collection", Draft August 6, 1998;

Forman, B. E.., 1999: "User Needs for Convective Weather Forecasts". Preprints, 8th Conference on Aviation, Range, and Aerospace Meteorology, Dallas, TX.

Troxel, S.W. and C.D. Engholm, 1990: Vertical Reflectivity Profiles: Averaged Storm Structures and Applications to Fan–beam Radar Weather Detection in the U.S., Preprints 16th Conference on Severe Local Storms, Kananaskis Park, Alta.,Canada, 213–218.

Wolfson, M.M., 1999: "The Growth and Decay Storm Tracker". Preprints, 8th Conference on Aviation, Range, and Aerospace Meteorology, Dallas, TX.