An Evaluation of WSR-88D Severe Hail Algorithms along the Northeastern Gulf Coast

ERIC LENNING* AND HENRY E. FUELBERG

The Florida State University, Tallahassee, Florida

ANDREW I. WATSON

NOAA/National Weather Service, Tallahassee, Florida

(Manuscript received 24 October 1997, in final form 3 August 1998)

ABSTRACT

Software build 9.0 for the Weather Surveillance Radar-1988 Doppler (WSR-88D) contains several new or improved algorithms for detecting severe thunderstorms. The WSR-88D Operational Support Facility supports testing and optimization of these algorithms by local National Weather Service offices. This paper presents a new methodology for using *Storm Data* in these local evaluations. The methodology defines specific conditions a storm cell must meet to be included in the evaluation. These conditions include cell intensity and duration, population density along the cell track, and any previous severe reports in the county where the storm is located. These requirements avoid including storm cells that may have produced severe weather where reports would be very unlikely. The technique provides a more accurate picture of algorithm performance than if *Storm Data* is used with no special considerations.

This study utilizes the new methodology with data currently available for the Tallahassee, Florida, county warning area (TLH CWA). It describes the performance of two algorithms used for detecting severe hail. The first is the Probability of Severe Hail (POSH), a component of the build 9.0 Hail Detection Algorithm. The second is the algorithm that calculates vertically integrated liquid (VIL).

Early results show that the recommended POSH threshold of 50% appears appropriate for the TLH CWA. This suggests that the height of the freezing level provides a reasonably good estimate of the best severe hail index (SHI). However, early results also indicate that the average wet-bulb temperature from 1000 to 700 mb (low-level wet-bulb temperature) might produce an even better indication of the SHI threshold. Similarly, the threshold for VIL is highly correlated to the low-level wet-bulb temperature. Finally, the VIL algorithm is found to perform as well as the POSH parameter if the best VIL threshold can be determined in advance. Since the database used in these evaluations was relatively small, these findings should be considered tentative.

1. Introduction

The National Weather Service (NWS) is undergoing a major modernization program that will improve the quality and reliability of its products and services. A vital component of this program is the Weather Surveillance Radar-1988 Doppler (WSR-88D). The WSR-88D continuously scans the atmosphere, while its specialized algorithms search for various storm features. These algorithms alert the user to potentially hazardous weather events and can be optimized for any WSR-88D site. This optimization is very important, especially for statistical algorithms developed from a relatively small database (Witt et al. 1998a) or for locations with a climate different from where the algorithms were developed (Winston and Ruthi 1986).

The build 9.0 release of WSR-88D software contains an entire suite of storm detection algorithms. Those that are replacements for algorithms in software build 8.0 are expected to yield considerably improved results over their predecessors. The WSR-88D Operational Support Facility (OSF) has encouraged individual NWS offices to perform local evaluations of these new algorithms and to determine optimum threshold values for their particular county warning areas (CWA; Crum 1995a; Witt et al. 1998a). To facilitate this process, the OSF has released the WSR-88D Algorithm Testing and Display System (WATADS) and has made level II archived radar data available from the National Climatic Data Center (Crum 1995b).

The task of obtaining a proper verification database for these local evaluations is best accomplished by spe-

* Current affiliation: PRC Inc., McLean, Virginia.

Corresponding author address: Eric Lenning, 2W1, 1500 PRC Drive, McLean, VA 22102.
E-mail: Lenning_Eric@prc.com

cial field projects (e.g., Kessinger and Brandes 1995; Rasmussen et al. 1994). However, such projects usually are not feasible for individual NWS offices. Instead, most NWS offices build a verification database using reports gathered through their standard warning verification efforts. These reports are published monthly in *Storm Data.* Unfortunately, the imprecise and incomplete nature of many reports in *Storm Data* makes it difficult to use this publication with confidence (Witt et al. 1998b; Hales 1993).

The primary objective of this paper is to present a new methodology that will allow NWS offices to use the information in *Storm Data* more effectively when performing local algorithm studies. The secondary objective is to report early results of a study that used the new methodology to evaluate the performance of two build 9.0 algorithms in the Tallahassee, Florida, county warning area (TLH CWA). The first is the probability of severe hail (POSH), one of three component algorithms in the new hail detection algorithm (HDA) (Witt et al. 1993). The second is the algorithm that calculates vertically integrated liquid water content (VIL; Greene and Clark 1972).

2. The WSR-88D build 9.0 VIL and POSH algorithms

Both the VIL and POSH algorithms are used to infer severe hail (diameter $\geq$ 19 mm). Although the VIL algorithm was not originally designed for this purpose, it has proven widely successful for hail prediction. The VIL algorithm is retained from build 8.0 with only minor changes. In contrast, the POSH parameter of the HDA is new since build 8.0. The original hail algorithm only indicated whether a cell was conducive to producing hail, whereas the HDA contains three separate algorithms that estimate the probability of hail (any size, POH), POSH, and the maximum expected hail size (MEHS). The performance of the new HDA has not been formally evaluated in many areas of the country, including the northeastern Gulf Coast.

The practice of using VIL and POSH for severe hail prediction is based on our current knowledge of hail formation. While the exact process of hail formation is not fully understood, nearly all basic theories (e.g., Foster and Bates 1956; Miller 1972; Browning and Foote 1976; Moore and Pino 1990) agree that hail forms above the freezing level and that a strong updraft must be present to support the hailstones during their development. Users of VIL and POSH incorporate these theories to help determine whether a cell is likely to produce hail. Thus, both algorithms are indirect methods of hail detection. Forecasters must rely on these indirect methods since the WSR-88D currently is not equipped to detect hailstones directly.

While the VIL and POSH algorithms serve similar functions, they provide different information. The VIL algorithm converts the maximum reflectivities at each level of a cell into liquid water content using relationships derived by Marshall and Palmer (1948). It then integrates these values through the depth of the cell to determine the liquid water content for the entire cell (Greene and Clark 1972). Large values of VIL correspond to strong radar reflectivities, thereby suggesting strong updrafts and the potential for heavy rain and possibly hail.

The POSH algorithm, in contrast, examines the area of the cell where hail formation occurs, that is, the portion above the freezing level. It calculates a severe hail index (SHI) based on reflectivities of at least 40 dBZ at altitudes above the freezing level. It ignores the smaller reflectivities that usually are associated with rain. More specifically, SHI is related to the hailfall kinetic energy ($\dot{E}$) of a cell (Waldvogel et al. 1978a; Waldvogel et al. 1978b; Waldvogel and Schmid 1982). Like VIL, hailfall kinetic energy is calculated using the maximum reflectivity at each level of a cell. It is defined as

$$\dot{E} = 5 \times 10^{-6} W(Z) 10^{0.084Z}, \tag{1}$$

where $W(Z)$, the reflectivity weighting factor, defines the transition from rain to hail. Here, $W(Z)$ is defined as

$$W(Z) = \begin{cases} 0 & \text{for } Z \leq 40 \text{ dB}Z \\ 0.1(Z - 40) & \text{for } 40 \text{ dB}Z < Z < 50 \text{ dB}Z \\ 1 & \text{for } Z \geq 50 \text{ dB}Z \end{cases} \tag{2}$$

with Z in dBZ and $\dot{E}$ (J m^{-2} s^{-1}). Reflectivity values less than 40 dBZ are considered liquid, while values greater than 50 dBZ are considered hail.

A temperature-based weighting function, WT(H), is included in the SHI equation to reduce the influence of seasonal and latitudinal variations in lapse rate. It is calculated as

$$\text{WT}(H) = \begin{cases} 0 & \text{for } H \leq H_0 \\ \dfrac{H - H_0}{H_{-20} - H_0} & \text{for } H_0 < H < H_{-20} \\ 1 & \text{for } H \geq H_{-20}, \end{cases} \tag{3}$$

where H is the height above ground level, H_0 is height of the freezing level, and H_{-20} is height of the -20°C level.

Finally, SHI is calculated by combining the temperature based weighting function with the hailfall kinetic energy to yield

$$\text{SHI} = 0.1 \int_{H_0}^{H_T} \text{WT}(H) \dot{E}\, dH, \tag{4}$$

in (J m^{-1} s^{-1}), where H_T is the storm-top height of each cell. Values of SHI can vary widely and rapidly during the lifetime of an individual cell.

VIL and SHI are useful for determining whether hail may be present in a cell. However, algorithms also should assess whether the hail that forms will melt be-

fore reaching the surface. Hail that falls through a warm, humid layer will melt significantly. According to Miller (1972), a wet-bulb-zero altitude near 3000 m supports the possibility of large hail at the surface. Rasmussen and Heymsfield (1987) also demonstrated that environmental temperature and humidity significantly affect the rate at which hail melts.

The POSH algorithm incorporates a warning threshold (WT) to estimate hail melt. The WT is the SHI value that is associated with severe hail on the given day. The Warning Threshold Selection Model (WTSM) uses the height of the freezing level (H_0) in kilometers to calculate the WT based on

$$\text{WT} = 57.5H_0 - 121. \quad (5)$$

The POSH algorithm (6) is designed such that a POSH value of 50% occurs when the SHI for a cell equals the WT for that day:

$$\text{POSH} = 29 \ln(\text{SHI}/\text{WT}) + 50. \quad (6)$$

For example, if the freezing level on one day is 3 km, (5) indicates that the warning threshold is 51.5. This means a SHI value of 51.5 produces a POSH of 50%. Using (6) in this scenario, SHI values of 288 or larger produce a POSH of 100%. In contrast, when the freezing level is 4 km, a SHI of 109 yields a POSH of 50%, and SHI values of 611 or larger produce a POSH of 100%.

The two coefficients in the WTSM equation (5) are defaults for every radar on which build 9.0 software is installed. They were derived from only 8 days of data, 5 of which were from Oklahoma (Witt 1993). Since these coefficients are adaptable parameters on the WSR-88D, TLH and other sites can adjust them to optimize local performance of the POSH algorithm. This paper evaluates the performance of the WTSM in the TLH CWA and explores other possibilities for a better WTSM equation.

The VIL algorithm does not include a daily threshold for severe hail. A VIL that produces severe hail at the surface on one day may not do so on another. For this reason, NWS meteorologists use several techniques for determining the ''VIL of the day,'' that is, the VIL associated with severe hail on that day. One technique uses the maximum VIL and reported hail size of the day's first known hail cell to predict hail sizes for later cells. Unfortunately, this means that no warning is issued if the first cell produces severe hail. Furthermore, since the first hail cell usually displays multiple values of VIL, it is not always clear which value to use as the VIL of the day. To address this problem, Paxton and Shepherd (1993) developed an empirical equation to forecast VIL of the day (VOD) for Florida thunderstorms:

$$\text{VOD} = -1500(T_{500} + T_{400})^{-1}, \quad (7)$$

where T_{500} and T_{400} are the temperatures at 500 and 400 mb, respectively.

This paper evaluates these two techniques for determining VIL of the day in the TLH CWA and explores other possibilities that might provide a better VIL of the day equation. It also seeks to determine whether SHI or VIL is more reliable for hail detection.

3. Data and methodology

The majority of severe hail events in the TLH CWA occur from March to June (Lenning et al. 1996). Therefore, many years may pass before TLH compiles a statistically significant number of events. This problem is compounded by the need to have level II data for each hail event. Nevertheless, it is important to develop a methodology to use when such a database does become available and to study the performance of these algorithms based on the data that currently are available. That was the objective of the current study.

The TLH NWS office collected 29 reports of severe hail between July 1995 and June 1996. Unfortunately, the level II data recorder at TLH was offline during nearly a third of these events, leaving only 20 for our evaluation. Furthermore, 4 of the 20 reports fell within the TLH radar cone of silence, defined here as a circle of 45-km diameter around the radar. Since cells within this region were too close to the radar to be sampled properly, they were unsuitable for this study. The 16 remaining severe hail reports provide a database that is suitable for performing an early evaluation of these algorithms using the new methodology that is proposed.

a. Rules for using Storm Data

We used reports from *Storm Data* (NOAA 1995, 1996) to verify algorithm predictions. Such reports are likely to be used in studies by other NWS offices, but they are subject to several limitations. Most notably, they limit the number of radar-derived hail predictions that can be evaluated with great confidence. For this reason, we established four rules to govern which volume scans would be evaluated. The first three rules apply to volume scans of cells that were not associated with a severe hail report, while the last rule applies to volume scans of cells that did produce severe hail reports. These rules helped ensure that *Storm Data* was used appropriately in this evaluation. These rules compose our proposed methodology for future evaluations.

The first rule addresses the lack of null reports in *Storm Data*. Since *Storm Data* seldom provides information for counties that did not receive a warning (Hales 1993), one often must infer that if no reports were received, then no severe weather occurred. This scenario is most common with short-lived hailstorms. Such storms may not attract the attention of anyone on the ground who would make a report, or anyone at the radar screen who could issue a warning and seek a report. To avoid including cells from such storms in our database, our first rule required a cell to maintain a significant intensity for at least two radar volume scans (~12 min).

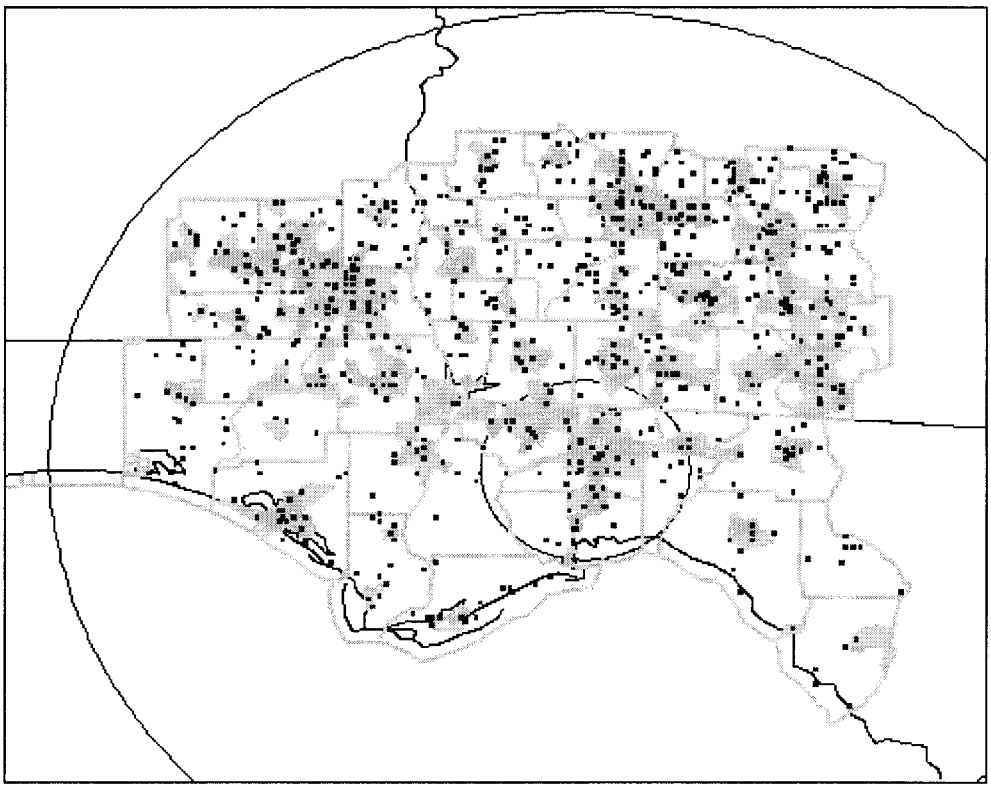

FIG. 1. Severe hail and wind reports for the Tallahassee county warning area during the period from 1955 to 1995. A total of 1619 reports is plotted. Shaded areas are census block groups with a population density of at least 15 people km^{-2}. County lines also are drawn, as well as range rings 45 and 230 km from the TLH radar.

This minimum intensity threshold was based on the POH parameter of the HDA (Witt et al. 1998a). POH measures strong reflectivity aloft to determine the probability of hail, yet it is independent of the VIL and POSH algorithms that are being evaluated. A cell satisfied the first rule if two consecutive volume scans maintained a POH of at least 70%.

The second rule addresses the problem of population density. Since most severe hail reports in *Storm Data* are generated through NWS warning verification efforts, a hailstorm that develops over water or a sparsely populated area probably will not be reported. To address this problem and reduce the potential for dubious false alarms, Witt et al. (1998b) suggested that algorithm evaluations be performed only over metropolitan areas, where the population density is relatively high. The city of Tallahassee is the largest such area in the TLH CWA, but it is a limited domain. Furthermore, Tallahassee is located within the radar cone of silence, an unsuitable location for these evaluations.

We approached the issue of population density by using TIGER/Line™ U.S. Census Data on CD-ROM (Bureau of the Census 1992). The historical distribution of severe storm reports in the TLH CWA was compared to several population densities. Ultimately, a threshold of 15 people per square kilometer was chosen, since it correlated well to the historical distribution of reports (Fig. 1). Specifically, reports tend to cluster in areas where the population density is at least 15 people km^{-2}. Few areas exceeding this threshold have no reports at all. Furthermore, this threshold is small enough to provide a reasonably large geographical area for the evaluation. To summarize, our second rule required a cell to pass over an area of relatively large population density ($\geq$15 people km^{-2}) to be considered for scoring.

The third rule addresses the lack of detailed information in *Storm Data*. NWS severe thunderstorm and tornado warnings need only one report per county to be verified. This verification does not require a complete description of the storm that produced the report, nor does it require information about other storms in the same county. Since warnings typically last 30–60 min, we established a rule for avoiding situations where the NWS had verified a warning and was no longer seeking reports from that county. Specifically, if a report was already received for a county, no other cells in that county were included in the database for 45 min after the time of the last report.

In summary, a cell that did not produce a severe hail report must satisfy three rather stringent requirements to be included in the evaluation.

1) At least two consecutive volume scans must meet a minimum intensity threshold (POH $\geq$ 70%).
2) The cell must be located over a relatively large population density ($\geq$15 people km^{-2}) while meeting the intensity requirement.
3) There must be no other severe reports of any kind in that county for at least 45 min prior to the time that the cell meets the two previous requirements.

If these three steps are taken to avoid including cells that may have produced severe hail where a report was unlikely, we believe that *Storm Data* can be a useful source of information for building a database to score algorithm predictions. Although these rules produce a significantly smaller database than if every volume scan were included, we believe they provide a more accurate picture of algorithm performance.

The fourth and final rule pertains to cells that did produce a severe hail report. This rule also addresses the lack of detailed information in *Storm Data.* As mentioned above, NWS verification procedures do not require a complete description of the storm that produced a report. As a result, *Storm Data* usually provides only enough data to verify algorithm predictions for one volume scan of one cell in each county. For this reason, algorithm predictions for cells that produced a severe hail report were scored according to a time-windows methodology (Witt et al. 1998b; Burgess 1993). Specifically, several volume scans of a cell were associated with a severe hail report: the volume scan(s) nearest in time to the report, two scans before this time, and one scan after. Thus, the fourth rule asks the algorithm to give approximately 12 min of lead time for each event and to continue to warn throughout the event. While these may not be realistic goals, they are desirable skills for a useful algorithm.

Uncertainties or even mistakes in the recorded time and/or location of a severe weather report are other limitations of *Storm Data* that must be overcome. In the current study, the Radar Analysis and Display System (Sanger 1994) was used to identify storm cells, while WATADS provided the VIL and SHI for each cell. Each severe hail report in the database was matched with the cell most likely responsible for that report. The position of the cell at the time of the report frequently did not match the location of the report. However, it usually was possible to adjust the report time slightly to correspond to when that cell did pass over the location of the report. In the two cases where this was not possible, the reports were discarded, leaving a total of 14 reports out of the original 29.

b. An example of applying these four rules

An application of the scoring methodology is shown in Fig. 2, a summary of 58 volume scans from seven selected storm tracks for the period 0120–0329 UTC 20 February 1996. These tracks were determined by the Storm Cell Identification and Tracking (SCIT) algorithm (Witt and Johnson 1993). SCIT has been shown to successfully identify storm centroids in approximately 97% of the cases (Burgess 1993; Witt and Johnson 1993).

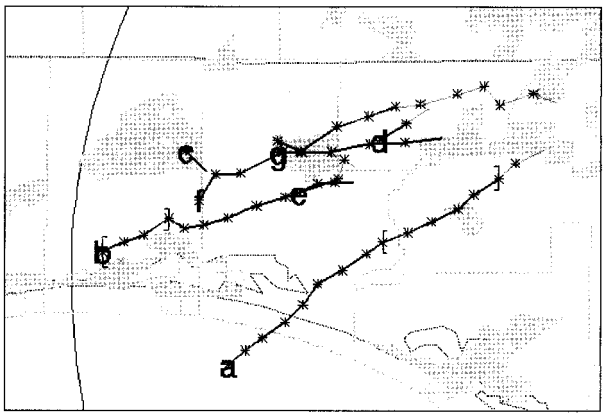

FIG. 2. Seven selected storm tracks for the period 0120–0329 UTC 20 February 1996. Shaded areas are census block groups with a population density of at least 15 people km^{-2}. County lines also are drawn, as well as the range ring 230 km from the TLH radar. Bold segments indicate a POH of at least 70. Brackets enclose volume scans that were associated with a severe-hail report using time windows.

Storm tracks in Fig. 2 are labeled with the letters centered over the first volume scan of each track. Each asterisk represents the completion of one volume scan, while the last asterisk for each cell is not plotted. The radar data from this and all other periods were gathered using Volume Coverage Pattern 21, which scans nine different elevation angles over a 6-min interval. Therefore, each segment in Fig. 2 represents a period of about 6 min. The initial segment of a cell is not plotted since there is no information about its starting point. Bold segments join volume scans that satisfy the first rule of this study by meeting the minimum intensity threshold (POH $\geq$ 70%) in succession. Brackets enclose volume scans that are associated with a severe hail report using time windows. County outlines are shown in gray, along with features such as streets, highways, and property lines in areas where the population density is at least 15 people km^{-2}. The range ring 230 km from TLH also is shown. Since the SCIT algorithm does not identify cells farther than 230 km from the radar, cells beyond this range were not used in the study.

The seven cells in Fig. 2 are examined individually to illustrate the scoring procedure. Cells a and b produced severe hail reports. Severe hail was reported near cell a over a period of 16 min from 0216 to 0232 UTC. Volume scans 11, 12, and 13 of this cell correspond to 0218, 0224, and 0230, respectively. According to the time-windows methodology and the rules of this study, algorithm predictions from volume scans 9–14 are associated with those severe hail reports and are scored, while the remaining predictions for this cell are ignored. Volume scans from cell b are included because hail was

TABLE 1. Summary of volume scans that were scored on 20 February 1996.

County	Location	Cell	Time (UTC)	Scans	Reason
Okaloosa	Crestview	c	0149–0155	1–2	Meets requirements
	Eglin AFB	b	0149–0206	1–4	Hail: 0200 UTC
Walton	DeFuniak Springs	f	0143–0206	3–7	Meets requirements
		g	0206–0224	1–4	Meets requirements
Washington	Vernon	a	0206–0236	9–14	Hail: 0216–0232 UTC

reported near Eglin Air Force Base (not shown) at 0200 UTC. Since scan 3 corresponds to 0201 UTC, algorithm predictions from scans 1–4 are scored according to the time-windows methodology, while the remaining predictions for this cell are ignored.

The other five cells in Fig. 2 did not produce severe reports. Therefore, they are examined using the three, more stringent, requirements to determine which volume scans to score and which to discard. For cell c, both volume scans are scored since this (bold) segment occurs over a relatively populated (shaded) area. For cells d and e, no volume scans are scored because the cells never pass over a densely populated area. Finally, for cells f and g, only volume scans corresponding to the bold segments over population areas are scored. These are scans 3–7 for cell f (which moves toward the northeast) and scans 1–4 for cell g. While the last segment of cell g also meets the intensity and population requirements (from 0230 to 0236 UTC), these scans are not scored because of our third rule: there were previous reports in this county associated with cell a from 0216 to 0232 UTC.

In summary, we included 21 volume scans from five cells in the dataset for 20 February 1996 (Table 1). Other days were scored similarly.

c. The database

For the six days of this evaluation, the time-windows methodology associated 60 volume scans of radar data with the 14 severe hail reports. An additional 140 volume scans were not associated with severe hail reports but did meet each of the three rules for inclusion into the database. These 200 volume scans correspond to 200 individual algorithm predictions. Although only six days and 14 severe hail reports are evaluated, the results are derived from a large number of algorithm predictions, each of which has been verified with a large degree of confidence.

For each study day, the observed TLH sounding that was nearest in time to the reports (0000 or 1200 UTC) was used to calculate thermodynamic parameters and to provide wind data for the algorithms. Although this was a future sounding in some cases, it was chosen because it was thought to be more representative of the actual storm environment than the sounding taken earlier.

d. Performance statistics

The algorithms were evaluated using the probability of detection (POD), false alarm ratio (FAR), and critical success index (CSI) statistics (Donaldson et al. 1975). These are defined as

$$\mathrm{POD} = \mathrm{YY}/(\mathrm{YY} + \mathrm{YN}), \tag{8}$$

$$\mathrm{FAR} = \mathrm{NY}/(\mathrm{YY} + \mathrm{NY}), \quad \text{and} \tag{9}$$

$$\mathrm{CSI} = \mathrm{YY}/(\mathrm{YY} + \mathrm{NY} + \mathrm{YN}), \tag{10}$$

where YY (a hit) means the event did occur and was predicted, NY (a false alarm) means the event did not occur but was predicted, and YN (a miss) means the event did occur but was not predicted. Here, YY is an accurate positive prediction of severe hail, while NY is an inaccurate positive prediction. Similarly, YN is an inaccurate negative prediction. An additional parameter, NN, represents an accurate negative prediction, although it is not part of the above equations.

e. An alternative approach

There are numerous possible variations to the methodology demonstrated above for 20 February (Fig. 2). Although many are less time consuming, their results are less reliable. One alternative is to score all 58 predictions from the seven cells, only excluding those portions of cell tracks over water. This approach still requires each cell to maintain a POH $\geq$ 70% for two volume scans at some point during its lifetime. A cell would never have to pass over a densely populated area to be scored. The number of predictions associated with a severe hail report would be the same for both approaches. Thus, the alternative method would yield an identical number of hits and misses at each POSH threshold, but an increased number of false alarms. The result would be a smaller CSI at every threshold. Section 4a includes a comparison of performance statistics on 20 February from both approaches.

4. Results

The four rules described in the previous section were used with the data that were available for the TLH CWA from July 1995 through June 1996. The results provide preliminary information about the performance of the build 9.0 VIL and POSH algorithms.

TABLE 2. Overall POSH performance. Total is the total number of positive (Y) predictions, YY is the number of hits, YN is the number of misses, and NY is the number of false alarms at or above each threshold. CSI is the critical success index, POD the probability of detection, and FAR the false alarm ratio. Boldface text highlights the largest CSI value. Computed performance statistics are shown in italics.

	0	10	20	30	40	50	60	70	80	90	100
Total	200	152	128	101	83	57	37	25	17	10	5
YY	60	57	53	50	48	43	28	23	16	9	5
YN	0	3	7	10	12	17	32	37	44	51	55
NY	140	95	75	51	35	14	9	2	1	1	0
CSI	*0.30*	*0.38*	*0.39*	*0.45*	*0.51*	***0.58***	*0.41*	*0.37*	*0.26*	*0.15*	*0.08*
POD	*1.00*	*0.95*	*0.88*	*0.83*	*0.80*	*0.72*	*0.47*	*0.38*	*0.27*	*0.15*	*0.08*
FAR	*0.70*	*0.63*	*0.59*	*0.51*	*0.42*	*0.25*	*0.24*	*0.08*	*0.06*	*0.10*	*0.00*

a. Performance of the POSH algorithm

The POSH algorithm was designed so that a threshold of 50% should give the best overall performance (CSI) on any given day (Witt et al. 1998a). That is, if warnings are issued only for cells with POSH values of 50% or greater, a relatively high POD and low FAR should be achieved. Of course, forecasters may issue warnings at any threshold.

Table 2 describes the overall performance of POSH at various thresholds for the entire study. For example, if a POSH threshold of 10% had been used for issuing warnings on every day of this study, there would have been 152 positive (Y) predictions of severe hail by the algorithm. However, only 57 of these predictions would have been associated with a severe hail report (YY), while 95 of them would not have been associated with a report (NY). There also would be three cases in which severe hail was reported when POSH was less than 10% (YN). A threshold of 10%, therefore, produces a very high POD (0.95), but also a high FAR (0.63), resulting in a relatively low CSI (0.38).

Table 2 suggests that a POSH threshold of 50% is most appropriate for the TLH CWA. This 50% threshold produces a CSI of 0.58, the largest CSI of any threshold (Table 2). The POD is 0.72, while the FAR is 0.25. For this study, 43 of the 60 volume scans associated with severe hail reports have POSH values at or above 50%, compared to only 14 of the 140 volume scans not associated with severe hail reports. For POSH thresholds greater than 50%, the POD decreases more than the FAR, producing a smaller CSI.

TABLE 3. Regional performance results for the POSH algorithm. TLH is Tallahassee, MLB is Melbourne, SP is southern plains, MR is Mississippi River area, NUS is northern United States, and All includes MR, SP, NUS, and MLB.

Region	POD	FAR	CSI
Witt et al. (1998a)			
All	0.78	0.69	0.29
MR	0.80	0.83	0.16
SP	0.82	0.54	0.41
MLB	0.71	0.75	0.23
NUS	0.71	0.58	0.36
This study			
TLH	0.72	0.25	0.58

Current POD values are similar to those derived by Witt et al. (1998a) for four different regions of the country (Table 3). Since Witt et al. used a 20-min time window and different verification rules, results for FAR are much smaller in the current study. The CSI results in Witt et al. for Melbourne, Florida, are better than those for the Mississippi River area (composed of St. Louis and Memphis), but worse than those for the southern plains and the northern United States. The smallest FAR derived by Witt et al. is in the southern plains. These varying statistics suggest the possibility of a regional bias in the POSH algorithm.

The differences in FAR (Table 3) could occur if NWS offices in the southern plains are more effective at verifying warnings. However, the differences also may suggest that the height of the freezing level used in the WTSM does not necessarily indicate how rapidly falling hail will melt. Factors such as high humidity and warm temperatures in the layer through which hail falls may be equally important, especially in certain areas of the country.

POSH performance on each individual day (Table 4) helps determine whether the POSH threshold of 50% consistently produces the best CSI. On 20 February, the best CSI is produced by POSH thresholds between 60% and 70%, not the desired threshold of 50%. On 12 July, based on a single severe hail report, 40% appears to give the best threshold. On the other four days, however, the 50% threshold does produce the best CSI. The CSI at 50% on these four days ranges from 0.33 to 0.80. Additional data and further studies will be necessary to show how well the 50% threshold performs in the TLH CWA on a regular basis. However, these initial statistics suggest that 50% is not always the most appropriate threshold. Furthermore, on the days when a POSH of 50% is appropriate, the CSI at this threshold shows room for improvement.

The CSI for POSH would be very different if this study utilized the alternative approach mentioned earlier (section 3e), which ignores the effects of population density. POSH performance determined by the alternative approach on 20 February is described in Table 5. On this day, the relative performance of POSH at different thresholds is similar for both methods. That is, POSH values from 60% to 70% still produce the

TABLE 4. POSH performance for individual days. Boldface CSIs correspond to best POSH thresholds. FAR values of −1.0 are undefined (since YY = 0). Other notation as in Table 2. Boldface text highlights the largest CSI values. Computed performance statistics are shown in italics.

12 July 1995											
POSH	0	10	20	30	40	50	60	70	80	90	100
Total	21	4	2	2	1	0	0	0	0	0	0
YY	1	1	1	1	1	0	0	0	0	0	0
YN	0	0	0	0	0	1	1	1	1	1	1
NY	20	3	1	1	0	0	0	0	0	0	0
CSI	*0.05*	*0.25*	*0.50*	*0.50*	***1.00***	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
POD	*1.00*	*1.00*	*1.00*	*1.00*	*1.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
FAR	*0.95*	*0.75*	*0.50*	*0.50*	*0.00*	*−1.00*	*−1.00*	*−1.00*	*−1.00*	*−1.00*	*−1.00*
20 February 1996											
POSH	0	10	20	30	40	50	60	70	80	90	100
Total	21	21	21	21	19	14	7	7	6	4	1
YY	10	10	10	10	10	9	7	7	6	4	1
YN	0	0	0	0	0	1	3	3	4	6	9
NY	11	11	11	11	9	5	0	0	0	0	0
CSI	*0.48*	*0.48*	*0.48*	*0.48*	*0.53*	*0.60*	***0.70***	***0.70***	*0.60*	*0.40*	*0.10*
POD	*1.00*	*1.00*	*1.00*	*1.00*	*1.00*	*0.90*	*0.70*	*0.70*	*0.60*	*0.40*	*0.10*
FAR	*0.52*	*0.52*	*0.52*	*0.52*	*0.47*	*0.36*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
6 March 1996											
POSH	0	10	20	30	40	50	60	70	80	90	100
Total	69	49	39	29	26	21	13	6	2	0	0
YY	19	17	15	15	14	13	7	5	2	0	0
YN	0	2	4	4	5	6	12	14	17	19	19
NY	50	32	24	14	12	8	6	1	0	0	0
CSI	*0.28*	*0.33*	*0.35*	*0.45*	*0.45*	***0.48***	*0.28*	*0.25*	*0.11*	*0.00*	*0.00*
POD	*1.00*	*0.89*	*0.79*	*0.79*	*0.74*	*0.68*	*0.37*	*0.26*	*0.11*	*0.00*	*0.00*
FAR	*0.72*	*0.65*	*0.62*	*0.48*	*0.46*	*0.38*	*0.46*	*0.17*	*0.00*	*−1.00*	*−1.00*
15 April 1996											
POSH	0	10	20	30	40	50	60	70	80	90	100
Total	37	33	32	27	23	16	10	7	6	5	4
YY	19	19	18	17	16	14	9	7	6	5	4
YN	0	0	1	2	3	5	10	12	13	14	15
NY	18	14	14	10	7	2	1	0	0	0	0
CSI	*0.51*	*0.58*	*0.55*	*0.59*	*0.62*	***0.67***	*0.45*	*0.37*	*0.32*	*0.26*	*0.21*
POD	*1.00*	*1.00*	*0.95*	*0.89*	*0.84*	*0.74*	*0.47*	*0.37*	*0.32*	*0.26*	*0.21*
FAR	*0.49*	*0.42*	*0.44*	*0.37*	*0.30*	*0.13*	*0.10*	*0.00*	*0.00*	*0.00*	*0.00*
10 May 1996											
POSH	0	10	20	30	40	50	60	70	80	90	100
Total	25	21	18	11	8	5	2	2	2	1	0
YY	7	6	5	3	3	3	1	1	1	0	0
YN	0	1	2	4	4	4	6	6	6	7	7
NY	18	15	13	8	5	2	1	1	1	1	0
CSI	*0.28*	*0.27*	*0.25*	*0.20*	*0.25*	***0.33***	*0.13*	*0.13*	*0.13*	*0.00*	*0.00*
POD	*1.00*	*0.86*	*0.71*	*0.43*	*0.43*	*0.43*	*0.14*	*0.14*	*0.14*	*0.00*	*0.00*
FAR	*0.72*	*0.71*	*0.72*	*0.73*	*0.63*	*0.40*	*0.50*	*0.50*	*0.50*	*1.00*	*−1.00*
7 June 1996											
POSH	0	10	20	30	40	50	60	70	80	90	100
Total	27	24	16	11	6	5	5	3	1	0	0
YY	4	4	4	4	4	4	4	3	1	0	0
YN	0	0	0	0	0	0	0	1	3	4	4
NY	23	20	12	7	2	1	1	0	0	0	0
CSI	*0.15*	*0.17*	*0.25*	*0.36*	*0.67*	***0.80***	***0.80***	*0.75*	*0.25*	*0.00*	*0.00*
POD	*1.00*	*1.00*	*1.00*	*1.00*	*1.00*	*1.00*	*1.00*	*0.75*	*0.25*	*0.00*	*0.00*
FAR	*0.85*	*0.83*	*0.75*	*0.64*	*0.33*	*0.20*	*0.20*	*0.00*	*0.00*	*−1.00*	*−1.00*

largest CSI. It is difficult, however, to claim that these statistics are accurate in an absolute sense. Specifically, the largest CSI obtained by the alternative approach (0.41) is much lower than almost every CSI obtained by the preferred approach (Table 4). Furthermore, these alternative performance statistics, while simpler to calculate, are not directly comparable to those from other regions of the country where population effects might be different.

To determine how much improvement is possible on any of the six days, actual results for each day (Table 4) can be compared to the best possible results for that

TABLE 5. POSH performance for 20 February 1996 using the alternative approach described in section 3e. Other notation as in Table 4. Boldface text highlights the largest CSI values. Computed performance statistics are shown in italics.

POSH	0	10	20	30	40	50	60	70	80	90	100
Total	54	51	47	39	32	23	14	14	12	9	4
YY	10	10	10	10	10	9	7	7	6	4	1
YN	0	0	0	0	0	1	3	3	4	6	9
NY	44	41	37	29	22	14	7	7	6	5	3
CSI	*0.19*	*0.20*	*0.21*	*0.26*	*0.31*	*0.38*	***0.41***	***0.41***	*0.38*	*0.27*	*0.08*
POD	*1.00*	*1.00*	*1.00*	*1.00*	*1.00*	*0.90*	*0.70*	*0.70*	*0.60*	*0.40*	*0.10*
FAR	*0.81*	*0.80*	*0.79*	*0.74*	*0.69*	*0.61*	*0.50*	*0.50*	*0.50*	*0.56*	*0.75*

day (Fig. 3). Specifically, Fig. 3 contains the CSI for the entire range of SHI values on each day. The SHI threshold that produced the best CSI is considered the ''best'' SHI threshold for that day. For example, Fig. 3b shows results for the 21 volume scans on 20 February. On this day, the smallest SHI is 22, while the largest is 237. A SHI threshold of 50 maximizes the POD relative to the FAR, producing a CSI of 0.80. On most days, the best CSI is produced by a relatively narrow range of SHI values (Fig. 3). Only 20 February has a unique best value of SHI. On days where a range of SHI thresholds produces the same CSI, the median SHI threshold in this range is used as the best SHI.

Comparing the best CSI for each day (Table 4) with the best possible CSI (Fig. 3) shows that the POSH algorithm produces the best possible CSI on four of the six days (12 July, 6 March, 10 May, 7 June). On no day does it have a significantly lower CSI than it could have. This suggests that the default POSH algorithm performs relatively well in the TLH CWA using a WTSM based on the freezing level.

An improved WTSM would more accurately predict the optimum SHI threshold for each day. It would help ensure that the POSH threshold of 50% always corresponds to the largest possible CSI on each day, while also increasing the CSI at this threshold. Therefore, four

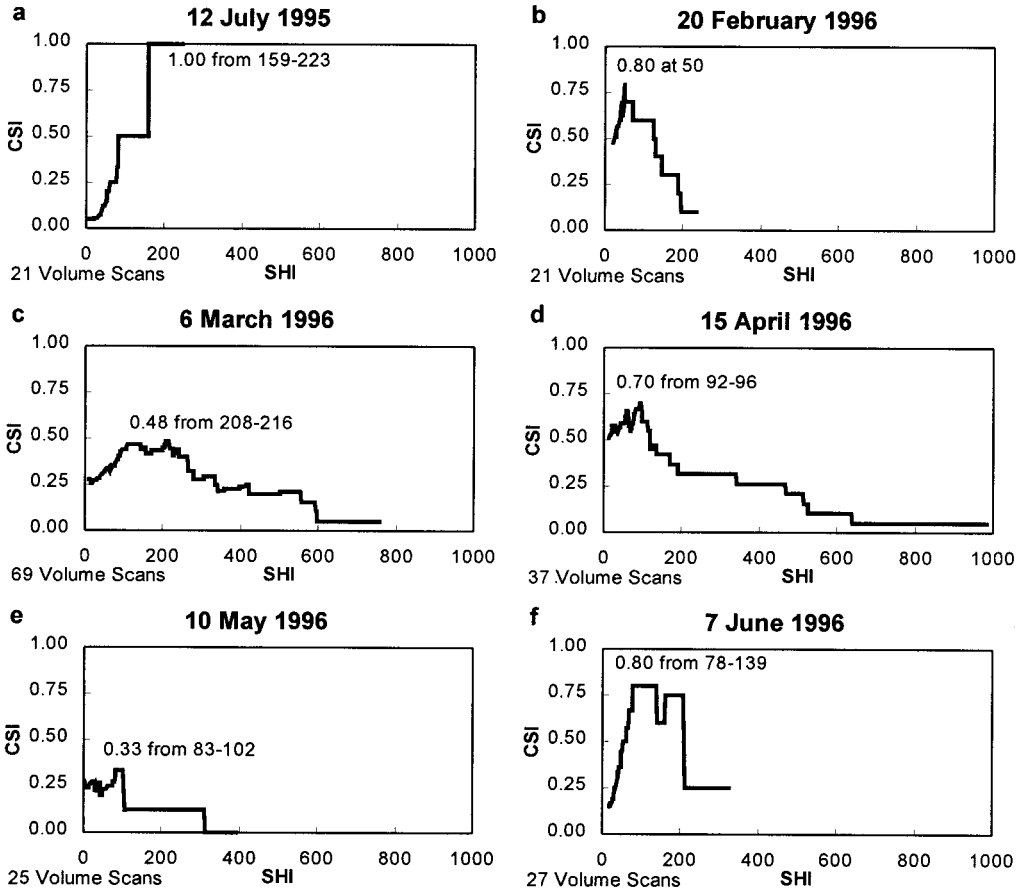

FIG. 3. Plots of CSI for the range of SHI values on each day of the study. The best SHI, or range of SHI, is indicated along with the best CSI. Flat segments correspond to areas where the ratio of hits, misses, and false alarms is constant over a range of SHI values.

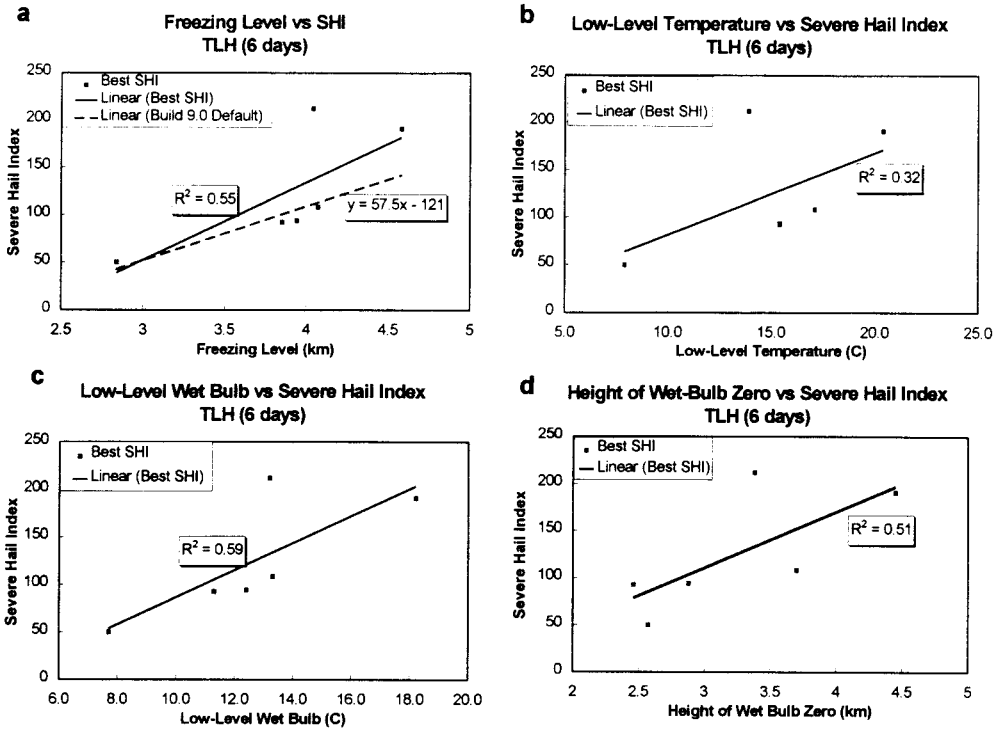

FIG. 4. Plots of best SHI vs (a) freezing level (km), (b) low-level temperature (°C), (c) low-level wet-bulb (°C), and (d) height of wet-bulb zero (km) for each day. The least squares fit line and associated variance are indicated for each parameter. Threshold from build 9.0 WTSM also is shown in (a).

different thermodynamic parameters were studied for possible inclusion in an improved WTSM equation for the TLH CWA. The first parameter is the freezing level, part of the current WTSM equation (5). The second is the average 1000–700-mb temperature (low-level temperature, LLT), while the third is the average 1000–700-mb wet-bulb temperature (low-level wet-bulb temperature, LLWB). The LLT and LLWB both influence the rate at which hail melts (Rasmussen and Heymsfield 1987). The fourth parameter, the height of wet-bulb zero (WBZ), traditionally has been used to determine whether large hail is likely on a given day (Miller 1972).

The LLT and LLWB were computed by interpolating temperature data from observed soundings into 25-mb intervals. Each value then was multiplied by the logarithm of the pressure at that level and averaged from 1000 to 700 mb. These relatively simple computations could easily be performed in an operational setting. The linear least squares fit between the four parameters and the best SHI (Fig. 3) is shown in Fig. 4 for each study day. The explained variance also is shown.

The solid line in Fig. 4a denotes the linear relation between the freezing level and the best SHI on each day. This line should be compared to the dashed line, which represents the default equation for the build 9.0 WTSM equation (5). Both equations show a similar relationship; however, the equation for TLH has a somewhat steeper slope. This suggests that the default WTSM in build 9.0 may increasingly underestimate the WT at TLH as the freezing level increases, which would lead to a higher FAR.

Differences between the default WTSM and the regression equation from this study (Fig. 4a) demonstrate the importance of evaluating build 9.0 algorithms for individual sites. In this case, the two coefficients in the WTSM equation (5) are the defaults for every radar on which build 9.0 software has been installed. However, they were derived from only 8 days of data, 5 of which were from Oklahoma (Witt 1993).

Although our relationship for TLH (Fig. 4a) is derived from only 6 days of data, 200 algorithm predictions were considered, each of which satisfied stringent requirements for inclusion. Even so, these early results may change as the level II database at TLH grows and more cases are investigated.

The remaining parameters in Fig. 4 yield variances that range from 0.32 for LLT (Fig. 4b) to 0.59 for LLWB (Fig. 4c). These preliminary results suggest that a WTSM based on the LLWB rather than the freezing level might improve POSH performance. Since the LLWB incorporates both low-level temperature and moisture, this conclusion is reasonable. The current build 9.0 WTSM does not utilize temperature or relative humidity (RH) in the lower layers of the troposphere. Although Witt et al. (1998a) explored the effect of a relative humidity adjustment factor in the WTSM, they

TABLE 6. Low-level (1000–700 mb) relative humidity and 700 mb relative humidity (%).

Date	Relative Humidity (%)	
	Low-Level	700 mb
12 July 1995	83	97
20 February 1996	81	80
6 March 1996	92	93
15 April 1996	64	21
10 May 1996	47	17
7 June 1996	68	90

found that it yielded only a 2% increase in the performance of POSH. Such a small improvement may result from their use of 700-mb dewpoint depression to represent the average RH in the lower troposphere. Data for TLH show that the RH at 700 mb differs considerably from the average RH between 1000 and 700 mb on four of the six study days (Table 6). The use of LLWB in the WTSM should be investigated further with an expanded dataset.

In summary, current results suggest that the POSH algorithm performs well in the TLH CWA. Use of the 50% threshold appears to give the best CSI when issuing warnings. The WTSM generally provides the POSH algorithm with a good estimate of the proper SHI threshold for each day. However, the WTSM may underestimate the proper SHI threshold for the higher freezing levels. Therefore, results might be improved by incorporating both the moisture and temperature in the lower troposphere, for example, the low-level wet-bulb temperature. More data are needed to confirm these early results.

b. Performance of the VIL algorithm

The VIL algorithm does not incorporate a threshold selection model to help determine when a cell is likely producing severe hail. Therefore, the VOD parameter (Paxton and Shepherd 1993) often is used as a starting point for determining the best daily VIL in the Florida region. Table 7 shows the widely varying performance of VOD for the TLH CWA. The VOD performs rather poorly on three of the days (12 July, 6 March, and 10 May), with CSI values of 0.20, 0.34, and 0.23, respectively. On the other three days, however, VOD performs relatively well.

The performance of best VIL also is shown in Table 7. The best VIL is the warning threshold that produces the best possible CSI on each day. Best VIL was determined from Fig. 5 in the same way best SHI was determined from Fig. 3. Although 15 April (Fig. 5d) has two values of best VIL, 37 is used for this study since 16.5 is considered unrealistically small for 15 April in the TLH area. Results in Table 7 show that the CSI produced by VIL of the day is considerably smaller than the best possible CSI on four of the six days. These preliminary data suggest that VOD is not appropriate for choosing the VIL threshold in the TLH CWA.

An alternative to VOD is to use the maximum VIL and associated hail size of the first hail cell on each day (1st VIL, Table 7). In the current study, this value is similar to the best VIL on only one day (12 July). On the other five days, the 1st VIL is much greater than the best VIL. This might be expected on two of these days. Specifically, since initial hail sizes on 20 February and 6 March are larger than 19 mm, cells with a smaller VIL might produce severe hail with a diameter closer to 19 mm. Therefore, the appropriate VIL threshold still is unclear after the first severe hail event. These results illustrate the limitations of using 1st VIL to detect severe hail.

The limitations of using either 1st VIL or the current VOD demonstrate the need for an improved VOD parameter. We examined several alternatives for inclusion in a better VIL of the day equation, specifically the freezing level, LLT, LLWB, and WBZ. The relationship between these four parameters and the best VIL of each day is shown in Fig. 6. The LLT and WBZ (Figs. 6b and 6d) demonstrate the poorest ability to predict the best VIL. Conversely, the LLWB and the freezing level (Figs. 6c and 6a) show the greatest ability. These preliminary findings are similar to those for SHI (Fig. 4), where LLWB and the freezing level also showed the greatest ability to predict an appropriate threshold.

These early results suggest that a useful VIL of the day equation could be based on the low-level wet-bulb temperature calculated from the nearest observed or predicted sounding. This conclusion is similar to that for the SHI parameter. If the best VIL threshold could be confidently predicted ahead of time, the VIL algorithm might be as convenient to use as the POSH algorithm for determining if a cell will produce severe hail. This

TABLE 7. VOD and best VIL, with CSI for both. Also, maximum VIL for the cell with the first hail report (1st VIL), with size of hail. Scans is the number of volume scans for each day. Computed performance statistics are shown in italics.

Date	Scans	VOD	CSI	Best VIL	CSI	1st VIL	Size (mm)
12 July 1995	21	45	*0.20*	66	*1.00*	68	19
20 February 1996	21	33	*0.63*	36	*0.82*	58	25
6 March 1996	69	44	*0.34*	54	*0.44*	61	38
15 April 1996	37	38	*0.57*	37	*0.59*	61	19
10 May 1996	25	44	*0.23*	52	*0.43*	66	19
7 June 1996	27	54	*1.00*	52	*1.00*	67	19

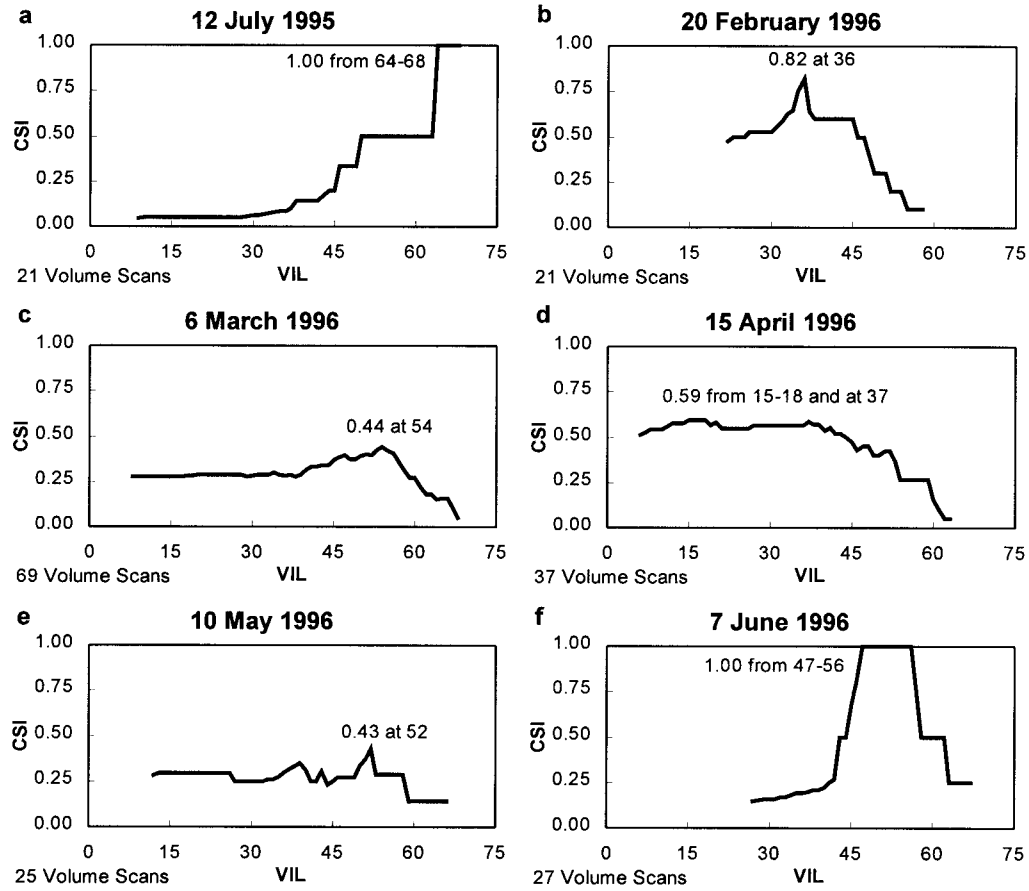

FIG. 5. Plots of CSI for the range of VIL values on each day of the study. The best VIL, or range of VIL giving the best CSI, is indicated. Other notation as in Fig. 3.

will require much additional study using an expanded dataset.

c. SHI, VIL, and hail size

It is useful to determine whether SHI or VIL is more reliable for hail detection. Table 8 lists the best SHI and VIL thresholds for each study day, that is, the values producing the best CSI. On three of the days, 12 July, 20 February, and 6 March, VIL and SHI yield similar CSIs. On 15 April, however, SHI gives a slightly better CSI, while on 10 May and 7 June, VIL provides a somewhat better CSI. These tentative results suggest that neither parameter is fundamentally more reliable. Instead, both show similarly good or poor performance depending on the conditions of each particular day.

Figure 7 shows time plots of VIL and SHI for four cells associated with severe hail reports (Figs. 7a–d) and four cells not producing severe hail reports (Figs. 7e–h) on 6 March 1996. Each cell with a severe hail report exhibits large values of both SHI and VIL. These values also show the rapid growth and decay that is characteristic of cells associated with severe weather over Florida (Paxton and Shepherd 1993).

One might question why the cell in Fig. 7c produces a severe hail report, since it exhibits relatively small values of VIL and SHI that do not undergo rapid fluctuation. In fact, values for the cell in Fig. 7h look more impressive, yet it does not produce a severe hail report. Since all cells in Fig. 7 are located between 120 and 180 km from the radar, distance from the radar is not a factor. Furthermore, each cell passes over a relatively dense population area ($\geq$15 people km^{-2}) near the time of greatest intensity. The important difference between the four severe hail-producing cells is the diameter of hail that was reported from each. The cells in Figs. 7a, 7b, and 7d produced hail with diameters of 38 mm, 64 mm, and 44 mm, respectively. The report for the cell in Fig. 7c, however, is only 19 mm, the minimum to be classified severe.

Witt et al. (1998a) concluded that hailstone diameter is proportional to the magnitude of SHI. That is, SHI is useful for distinguishing between significant and marginal hail events, or between marginal events and nonevents. The results in Fig. 7 support Witt's conclusion and his use of SHI in the build 9.0 algorithm for determining MEHS (Witt et al. 1998a). MEHS was not evaluated in the current study.

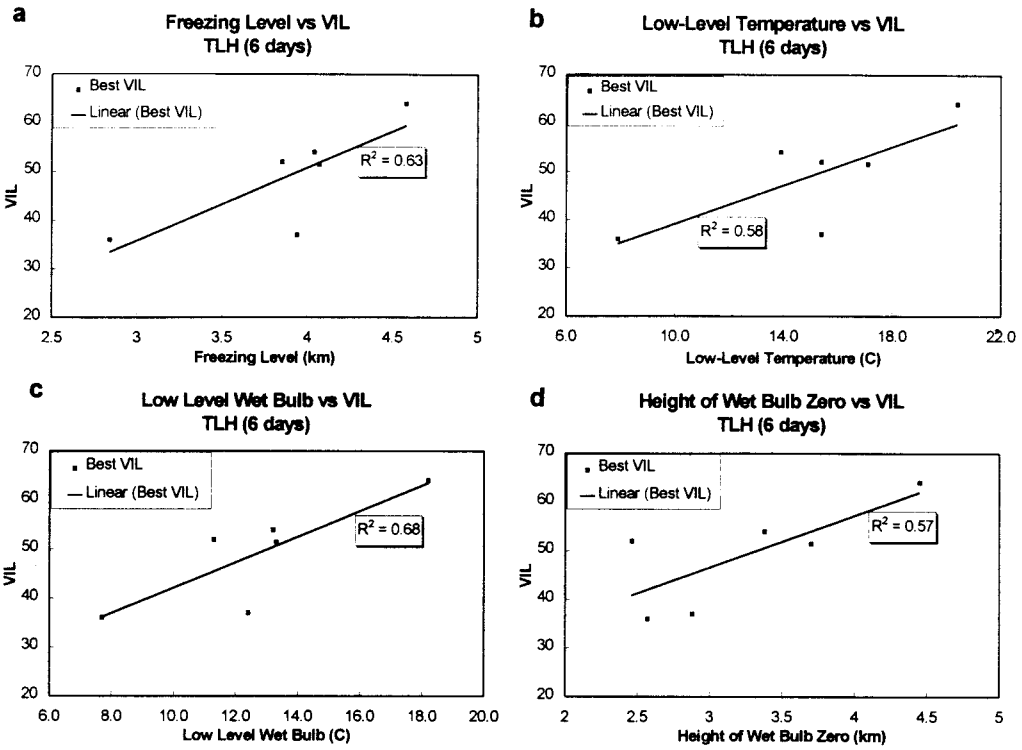

FIG. 6. Plots of best VIL vs (a) freezing level (km), (b) low-level temperature (°C), (c) low-level wet-bulb (°C), and (d) height of wet-bulb zero (km) for each day. The least squares fit line and associated variance are indicated for each parameter.

These early results suggest that neither SHI nor VIL is fundamentally more reliable in the TLH CWA. Both show similarly good or poor performance depending on the particular day. In addition, both appear useful for distinguishing between significant and marginal hail events, or between marginal events and nonevents. Finally, these results indicate that an appropriate and useful VIL of the day equation and an improved WTSM could be based on the low-level wet-bulb temperature calculated from the nearest observed or predicted sounding.

5. Conclusions

Software build 9.0 for the WSR-88D contains several new or improved algorithms for detecting severe thunderstorms. The WSR-88D Operational Support Facility recommends and supports testing and optimization of these algorithms by local NWS offices. This is especially important for statistical algorithms that were developed in a different climate. The task of obtaining a proper verification database for these evaluations is best accomplished by special field projects. Unfortunately, individual forecast offices, with their limited time and resources, usually cannot conduct such projects.

TABLE 8. Best SHI and best VIL for each day, with corresponding CSI values. Scans is the number of volume scans used in the analysis for each day.

Date	Scans	Best SHI	CSI	Best VIL	CSI
12 Jul 1995	21	191	1.00	66	1.00
20 Feb 1996	21	50	0.80	36	0.82
6 Mar 1996	69	212	0.48	54	0.44
15 Apr 1996	37	94	0.70	37	0.59
10 May 1996	25	93	0.33	52	0.43
7 Jun 1996	27	109	0.80	52	1.00

This paper has presented a new methodology for effectively using the information in *Storm Data* locally to evaluate algorithm predictions. The methodology defines specific conditions a storm cell must meet to be included in the evaluation. These conditions require a cell to maintain a POH of at least 70% for two consecutive volume scans while located over an area where the population density is at least 15 people km^{-2}. Cells are excluded if other severe weather was reported during the previous 45 min in the same county. These steps avoid including most storms for which a severe hail report was unlikely. This new methodology produces a smaller database than if every volume scan were included. However, based on our comparison of results from two approaches, we believe it provides a more accurate picture of how well an algorithm performs.

Utilizing this new technique and data currently available for the TLH CWA, the paper has described the performance of the build 9.0 VIL and POSH algorithms. Although definite conclusions should not be drawn at

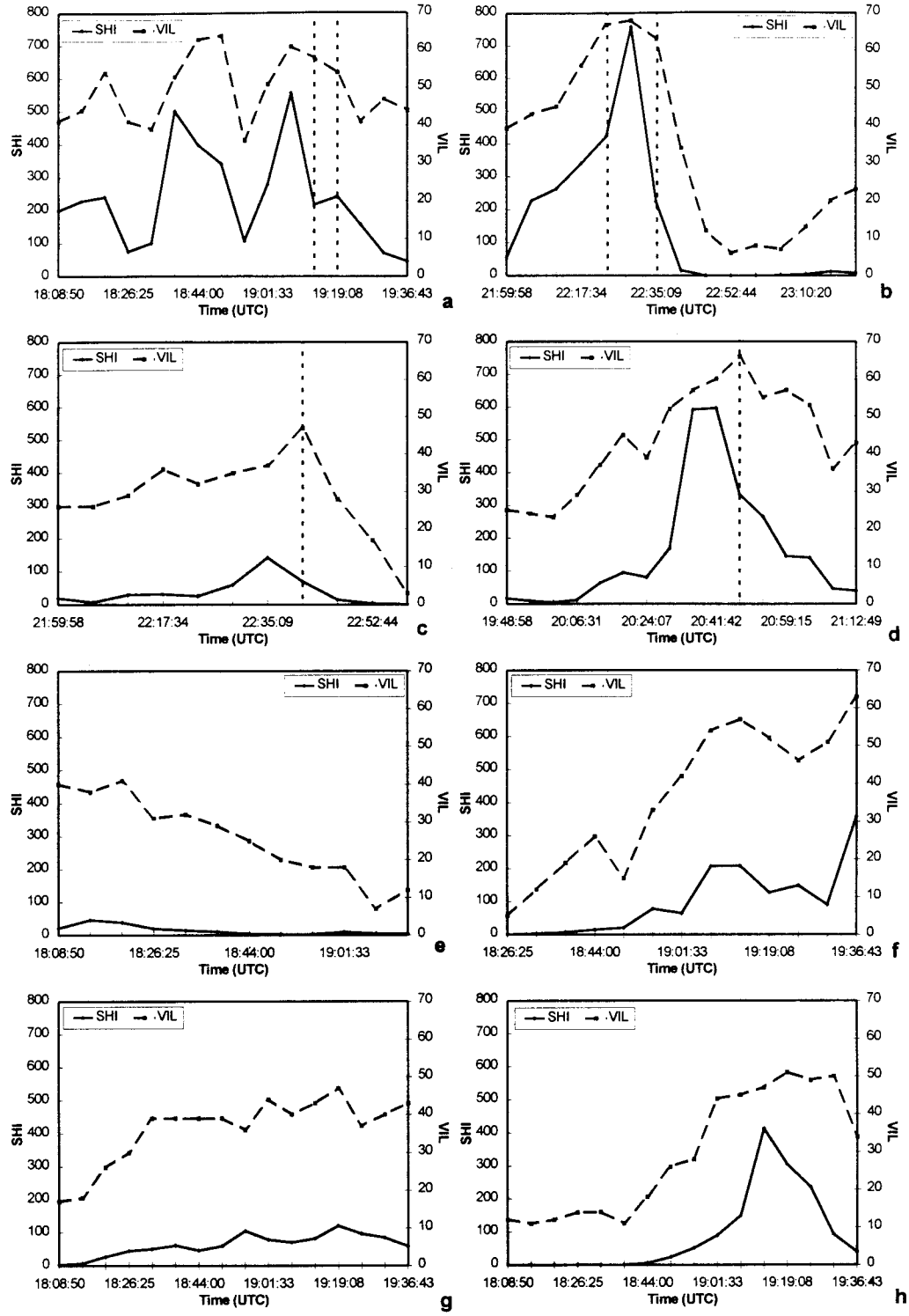

FIG. 7. Time plots of SHI (solid line) and VIL (dashed line) for (a)–(d) four hail-producing and (e)–(h) four non-hail-producing cells on 6 Mar 1996. Dotted vertical lines show times of hail reports.

this time, early results based on 200 algorithm predictions suggest that the recommended POSH threshold of 50% does provide useful guidance for issuing severe hail warnings in the TLH CWA. Furthermore, the WTSM, which is related to the height of the freezing level, seems to provide a reasonably good estimate of the best SHI for any given day. However, the early results also suggest that the average wet-bulb temperature between 1000 and 700 mb might provide an even better indication of the proper SHI for the TLH area. The threshold for VIL also was found to be highly correlated to the LLWB. Results suggest that the VIL algorithm will perform as well as the POSH algorithm if the VIL threshold for severe hail can be properly forecast. Additional studies, based on an expanded level II database, are needed to confirm the validity of these early results.

Further analysis into the relationship between the WT and the freezing level may not significantly alter the default build 9.0 WTSM equation. However, if another parameter such as LLWB produces a more accurate WT prediction (i.e., a better correlation to the best SHI), a WTSM based on this parameter should be considered. Given today's computing power, a particular parameter (e.g., freezing level) should not be used simply because it is easy to calculate.

This study has sought to minimize the effects of population distribution and verification efficiency on algorithm performance statistics. Its results indicate that 50% is the appropriate POSH threshold for issuing warnings in the TLH CWA. This does not mean, however, that a POSH threshold of 50% will necessarily yield the best performance statistics in an operational environment, where verification efficiency and population distribution are important and real factors.

Lenning et al. (1996) noted that NWS warning verification efforts have increased significantly since 1980. It is important that local NWS offices now refine their verification efforts so they can use storm reports more effectively in their local algorithm studies. An improved verification database is even more necessary now that level II data are being archived so that stations can perform on-site studies using WATADS. Of course, improved reliability of the level II archive recorders also will assist offices in building a radar database.

We recommend that the verification methodology presented here be used by other NWS offices in their evaluations. It is time-consuming to manually determine which cells pass over densely populated areas, especially when there are dozens of cells and intertwining cell tracks. However, software that would automate or greatly simplify this task would not be difficult to develop. We believe these procedures for selecting events give a more accurate measure of algorithm performance and are worth the extra effort.

The methodology presented in this paper is adaptable for scoring a variety of algorithms. Future efforts should evaluate the other new algorithms in build 9.0 and future builds. As more and more data become available to the forecaster, these algorithms become increasingly important. They help the forecaster interpret a large amount of information very quickly. Local evaluations and optimizations of these algorithms will lead to a genuine increase in their overall usefulness and to a better understanding of the atmosphere in general.

Acknowledgments. The authors thank Arthur Witt and Mike Eilts for providing assistance with the build 9.0 hail detection algorithm; Steve Courton for assistance with WATADS; Tom Ross for providing tapes of archived radar data; and Gary Beeler, Jeffrey Medlin, Paul Duval, Bob Goree, Christopher Herbster, and Paul Ruscher for many helpful suggestions.

This research was funded from Subawards S94-49193, S96-71863, and S97-86993 from the Cooperative Program for Operational Meteorology, Education and Training under a cooperative agreement between the National Oceanic and Atmospheric Administration (NOAA) and the University Corporation for Atmospheric Research (UCAR). The views expressed herein are those of the authors, and do not reflect the views of NOAA, its subagencies, or UCAR.

REFERENCES

Browning, K. A., and G. B. Foote, 1976: Airflow and hail growth in supercell storms and some implications for hail suppression. *Quart. J. Roy. Meteor. Soc.,* **102,** 499–534.

Bureau of the Census, 1992: *TIGER/Line File Alabama*: CD-TGR92-39; *TIGER/Line File Florida*: CD-TGR92-13 and CD-TGR92-14; *TIGER/Line File Georgia:* CD-TGR92-29 and CD-TGR92-30; and *TIGER/Line File Mississippi*: CD-TGR92-41. [Available from U.S. Dept. of Commerce, Bureau of the Census, P.O. Box 277943, Atlanta, GA 30384-7943.]

Burgess, D. W., 1993: Results of an experiment to evaluate WSR-88D algorithms. Preprints, *26th Int. Conf. on Radar Meteorology,* Norman, OK, Amer. Meteor. Soc., 160–162.

Crum, T., 1995a: Enhancing and regionalizing the WSR-88D meteorological algorithms. Preprints, *27th Int. Conf. on Radar Meteorology,* Vail, CO, Amer. Meteor. Soc., 167–169.

——, 1995b: An update on WSR-88D level II data availability. *Bull. Amer. Meteor. Soc.,* **76,** 2485–2487.

Donaldson, R. J., R. M. Dyer, and M. J. Kraus, 1975: An objective evaluator of techniques for predicting severe weather events. Preprints, *Ninth Conf. on Severe Local Storms,* Norman, OK, Amer. Meteor. Soc., 321–326.

Foster, D. S., and F. C. Bates, 1956: A hail size forecasting technique. *Bull. Amer. Meteor. Soc.,* **37,** 135–141.

Greene, D. R., and R. A. Clark, 1972: Vertically integrated liquid water—A new analysis tool. *Mon. Wea. Rev.,* **100,** 548–552.

Hales, J. E., 1993: Biases in the severe thunderstorm data base: Ramifications and solutions. Preprints, *13th Conf. on Weather Forecasting and Analysis,* Vienna, VA, Amer. Meteor. Soc., 504–507.

Kessinger, C. J., and E. A. Brandes, 1995: A comparison of hail detection algorithms. Final Rep. to the FAA, 52 pp. [Available from NCAR, P.O. Box 3000, Boulder, CO 80307.]

Lenning, E., H. E. Fuelberg, and B. Goree, 1996: An analysis of severe weather reports for the new Tallahassee, Florida, county warning area. Preprints, *18th Conf. on Severe Local Storms,* San Francisco, CA, Amer. Meteor. Soc., 330–334.

Marshall, J. S., and W. M. Palmer, 1948: The distribution of raindrops with size. *J. Meteor.,* **5,** 165–166.

Miller, R. C, 1972: Notes on analysis and severe-storm forecasting procedures of the Air Force Global Weather Central. Tech. Rep. 200 (Rev.), Air Weather Service (MAC), U.S. Air Force, 190 pp.

Moore, J. T., and J. P. Pino, 1990: An interactive method for estimating maximum hailstone size from forecast soundings. *Wea. Forecasting,* **5,** 508–525.

NOAA, 1995: *Storm Data.* Vol. 37, Nos. 5–12, NESDIS, National Climatic Data Center, Asheville, NC.

——, 1996: *Storm Data.* Vol. 38, Nos. 1–6. NESDIS, National Climatic Data Center, Asheville, NC.

Paxton, C. H., and J. M. Shepherd, 1993: Radar diagnostic parameters as indicators of severe weather in central Florida. NOAA Tech. Memo. NWS-SR 149, 12 pp. [Available from NTIS, U.S. Department of Commerce, 5285 Port Royal Road, Springfield, VA 22161.]

Rasmussen, E. N., J. M. Straka, R. Davies-Jones, C. A. Doswell III, F. H. Carr, M. D. Eilts, and D. R. MacGorman, 1994: Verification of the Origins of Rotation in Tornadoes Experiment: VORTEX. *Bull. Amer. Meteor. Soc.,* **75,** 995–1006.

Rasmussen, R. M., and A. J. Heymsfield, 1987: Melting and shredding of graupel and hail. Part II: Sensitivity study. *J. Atmos. Sci.,* **44,** 2764–2782.

Sanger, S. S., 1994: An interactive Doppler radar and weather detection algorithm display system. Preprints, *10th Int. Conf. on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology,* Nashville, TN, Amer. Meteor. Soc., 7–10.

Waldvogel, A., and W. Schmid, 1982: The kinetic energy of hailfalls. Part III: Sampling errors inferred from radar data. *J. Appl. Meteor.,* **21,** 1228–1238.

——, W. Schmid, and B. Federer, 1978a: The kinetic energy of hailfalls. Part I: Hailstone spectra. *J. Appl. Meteor.,* **17,** 515–520.

——, B. Federer, W. Schmid, and J. F. Mezeix, 1978b: The kinetic energy of hailfalls. Part II: Radar and hailpads. *J. Appl. Meteor.,* **17,** 1680–1693.

Winston, H. A., and L. J. Ruthi, 1986: Evaluation of RADAP II severe storm detection algorithms. *Bull. Amer. Meteor. Soc.,* **67,** 145–150.

Witt, A., 1993: Comparison of the performance of two hail detection algorithms using WSR-88D data. Preprints, *26th Int. Conf. on Radar Meteorology,* Norman, OK, Amer. Meteor. Soc., 154–156.

——, and J. T. Johnson, 1993: An enhanced storm cell identification and tracking algorithm. Preprints, *26th Int. Conf. on Radar Meteorology,* Norman, OK, Amer. Meteor. Soc., 141–143.

——, M. D. Eilts, G. J. Stumpf, J. T. Johnson, E. D. Mitchell, and K. W. Thomas, 1998a: An enhanced hail detection algorithm for the WSR-88D. *Wea. Forecasting,* **13,** 286–303.

——, ——, ——, E. D. Mitchell, J. T. Johnson, and K. W. Thomas, 1998b: Evaluating the performance of WSR-88D severe storm detection algorithms. *Wea. Forecasting,* **13,** 513–518.