# Development of a Hail-Detection-Product

I. Holleman, H. R. A. Wessels, J. R. A. Onvlee, and S. J. M. Barlag

Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands

# Development of a Hail-Detection-Product

**I. Holleman, H. R. A. Wessels, J. R. A. Onvlee, and S. J. M. Barlag**

Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands

**Abstract.** From literature five methods for the detection of summer hail have been selected. These five different methods have been tested on severe weather events in the Netherlands that occurred during the summer of 1999. The general trends in the scoring parameters of the detection methods as a function of the warning threshold are rather similar, but there are substantial quantitative differences. Using a simple model, the effects of missing ground truth data on the scoring parameters of the detection methods has been described qualitatively. It is concluded that, of all hail detection methods considered, the method of Waldvogel performs best and is suited best for display of the "probability of hail".

## 1 Introduction

Currently, a tool for the detection and display of severe weather phenomena related to convective systems, like wind-gusts and summer hail, is being developed at the KNMI. Radar reflectivity and Doppler winds will be the primary source of information, and it will be complemented with other observations and data from Numerical Weather Prediction (NWP) models. The KNMI operates two Gematronik C-band Doppler radars which are performing low-elevation volume scans every 5 minutes and extensive volume scans every 15 minutes. From the low-elevation volume scans, a pseudoCAPPI product and an echotop product are extracted. Ground clutter is removed from the pseudoCAPPI image using a statistical method (Wessels and Beekhuis , 1994). The first new product under consideration is a tool for the detection and display of summer hail.

From the literature five methods for the detection of summer hail have been selected. First of all, the methods consisting of CAPPI and Vertically-Integrated-Liquid images with a warning threshold have been considered. Next, the method developed recently by Auer (1994), in which reflectivity is combined with cloud-top temperatures, has been studied.

*Correspondence to:* I. Holleman

Finally, two methods which are currently in use within the framework of WSR-88D radar program, the method of Wald-vogel and the Severe Hail Index (SHI), have been selected (Witt et al. , 1998). These five different methods have been tested on severe weather events in the Netherlands that occurred during the summer of 1999. As ground truth, observations by the 321 volunteers of the (rainfall) observer network of the KNMI as well as detailed hail damage reports from insurance companies have been taken. In the validation of the detection methods against the ground truth data, both the effect of incomplete ground truth data and the influence of possible spatial mismatches between radar observations well above ground and surface hail reports have been investigated.

## 2 Methods

Nowadays, the most direct way to distinguish between hail and rain is by using the dual-polarization radar technique which can make a direct distinction between the spherical, rotating hail stones and the non-spherical rain droplets (Aydin et al. , 1986; Höller et al. , 1994; Smyth et al. , 1999). As operational implementation of the dual-polarization technique is rare up to now, detection methods for operational use still have to rely on single-polarization radar data in general.

The first method which may be used to distinguish hail from rain using a single-polarization radar is based on a plan-position indicator of the radar reflectivity at constant altitude (CAPPI display). At the KNMI, these CAPPIs are calculated for an altitude of 0.8 km above mean-sea-level (MSL). Because the radar reflectivity increases dramatically with increasing diameter of the scattering particles, larger hail stones (>10 mm) potentially give rise to higher reflectivities than would be possible for rain droplets which have a maximum diameter of 6.5 mm. Mason (1971) has suggested to use a reflectivity threshold of 55 dBZ for distinguishing between rain and hail when using this CAPPI method.

Recently, Auer (1994) has reported on the detection of hail using a combination of radar reflectivity at low altitude

and cloud-top temperatures. Using a nomogram for hail and heavy rain events, a dependence of the optimum reflectivity threshold on the cloud-top temperature is determined. This optimum threshold varies between 37 and 54 dBZ for cloud-top temperatures in the range between $-5$ and $-55\ ^\circ$C. The cloud-top temperature is determined either from the infrared imagery of Meteosat or by combining radar echotops with temperature information from a radio sounding or from an NWP model. The method of Auer has been verified during an all-season operational evaluation in New Zealand, and it was seen to perform a lot better than the ordinary, fixed threshold CAPPI method. In a sequel to this study, Hardaker and Auer (1994) have attempted to separate the contributions of rain and hail to the total reflectivity signal.

The use of the entity "Vertically Integrated Liquid water" (VIL) for the detection of thunderstorms has been introduced by Greene and Clark (1972). Via conversion of the radar reflectivity to liquid-water content and subsequent vertical integration of this water content, the three-dimensional radar data is converted to a plan-position indicator of the amount of liquid water present in a vertical column above a certain position. It has been observed that a high value of this "potential rainfall indicator" correlates well with the occurence of severe thunderstorms and hail. There is no agreement in literature on the best warning threshold for the detection of hail with the VIL method. American forecasters often use a "VIL of the day" threshold which is determined either by using the temperatures at 400 and 500 hPA via an empirical equation or by taking the VIL-value corresponding to the first hail storm of that day (Lenning et al. , 1998). Amburn and Wolf (1997) have proposed to circumvent this threshold problem by use of the entity "VIL-density" which is defined as the ratio of the VIL value and the radar echotop height. They suggest a universal warning threshold for the detection of hail of 3.5 g/m$^3$. However, the advantage of the use of VIL-density over just VIL is disputed (Edwards and Thompson , 1998).

The original version of the "Hail Detection Algorithm" (HDA) for the WSR-88D radar network in the USA (Smart and Alberty , 1985), uses a combination of seven hail indicators. The most important indicators are the presence of a reflectivity core of 50 dBZ or higher between 5 and 12 km altitude and the presence of radar echotops higher than 8 km. The version of the HDA that is currently in use is developed by Witt et al. (1998), and it is based on the hail criterium as proposed by Waldvogel et al. (1979). The method of Waldvogel for the detection of hail uses the maximum altitude at which a reflectivity of 45 dBZ is found. When this strong reflectivity extends to 1.4 km or more above the freezing level, the presence of hail is likely, and the probability of the presence of hail increases with increasing height of this reflectivity core above the freezing level. The height of the freezing level is determined from a radio sounding or from an NWP model. In the current HDA, the maximum height of the 45 dBZ reflectivity above the freezing level is converted to a probability of hail, where a height difference of 1.4 km corresponds to 0% and of 6.0 km to 100% probability (Witt

et al. , 1998).

The current HDA also contains an algorithm that attempts to estimate the probability of severe hail. For this, a semi-empirical relationship between the kinetic energy flux of the hail stones and the radar reflectivity is used (Waldvogel et al. , 1978a,b). A "Severe Hail Index" (SHI) is calculated by vertically integrating the kinetic energy flux weighted by a reflectivity-based and a temperature-based gating function. Subsequently, a SHI warning threshold is calculated from the height of the freezing level using an empirical relationship. Finally, the probability of severe hail is calculated from the obtained SHI and this warning threshold using again an empirical relationship (Witt et al. , 1998).

## 3 Comparison with ground truth

A systematic comparison of the output of the five selected methods, i.e., CAPPI, VIL, Auer, Waldvogel, and SHI, for the detection of hail to on-ground observations of hail has been conducted using radar scan data of selected days in the summer of 1999. Due to the small spatial extent of most hail events related to summertime thunderstorms, the 19 synops observers in the Netherlands will only report a minor fraction of the total number of hail events. Therefore, the ground truth data have been completed with hail observations by the 321 volunteers of the (rainfall) observer network of the KNMI, detailed hail damage reports from agricultural insurance companies, and reports of hail in newspapers.

The hail events can be classified using a 2-by-2 contingency table. Hail detected by radar which is confirmed by ground truth observations will be classified as a hit ($H$), hail detected by radar which is not confirmed by observations as a false alarm ($F$), observed hail that is not detected by radar as a miss ($M$), and no event as a none ($N$). Although much effort has been put into the collection of ground truth data, some hail events will remain unnoticed. By assuming that only a fraction $\eta$ of the occurring hail events are reported, the effect of the missing ground truth data on the classification of the hail events can be investigated. In Table 1, the four classes of a modified contingency table ($H'$, $M'$, $F'$, and $N'$) are expressed in terms of the original classes and the fraction $\eta$. Using the modified contingency table, the apparent Probability Of Detection (POD$'$), False Alarm Rate (FAR$'$), and Critical Success Index (CSI$'$) can be expressed in terms of the fraction $\eta$ and the true POD, FAR, and CSI:

$$\mathrm{POD}' \;\; = \;\; \frac{H'}{H' + M'} \;\; = \;\; \mathrm{POD}$$

**Table 1.** Modified contingency table, which is valid when only a fraction $\eta$ of hail events, that have occurred, is reported.

| | | Hail | No Hail |
|---|---|---|---|
| Detection: | Yes | $\eta \cdot H$ | $F + (1 - \eta) \cdot H$ |
| | No | $\eta \cdot M$ | $N + (1 - \eta) \cdot M$ |

$$\text{FAR}' = \frac{F'}{H' + F'} = (1 - \eta) + \eta \cdot \text{FAR}$$

$$\frac{1}{\text{CSI}'} = \frac{H' + M' + F'}{H'} = \frac{1}{\text{CSI}} + \frac{1 - \eta}{\eta \cdot (1 - \text{FAR})}$$

In this simple $\eta$-model, incompleteness of the ground truth data results in an increase of the apparent FAR' and a concomitant decrease of the apparent CSI'. As the maximum distance to the radar is roughly 150 km, the possible reduction in POD by undersampling of hail cells or by attenuation of the radar radiation is expected to be less significant than the effect of the missing ground truth data on the FAR, and therefore it is not explicitly taken into account in this model. A more sophisticated, two-parameter model for treating the effects of imperfect reporting on the verification of weather warnings is presented by Smith (1999).

Because most hail observations are only available per day and per municipality or zip-code area, the pixels in the radar images have been grouped into municipality-areas and subsequently these images have been combined to daily composites. Using a scoring program, every municipality in the Netherlands (538 municipalities, average area of 63 km$^2$) is classified for each day as an $H'$, $M'$, $F'$, or $N'$. The final POD', FAR' and CSI' scores are obtained by combining the day-by-day results for 15 days with thunderstorms in the Netherlands during the summer of 1999.

## 4   Results

In the assignment of a particular group of radar pixels having values above the warning threshold to a municipality where hail has been observed, allowance has been made for a certain spatial mismatch. The maximum allowed distance between a 2.4 km radar pixel above threshold and the boarder of a municipality with the on-ground confirmation is defined as the "positioning tolerance". When the positioning tolerance is increased, the observed POD' for the detection methods steadily increases and their observed FAR' steadily decreases. In Fig. 1 the resulting CSIs for the method of Waldvogel, the VIL method, and the CAPPI method are shown as a function of the positioning tolerance. The two other detection methods show similar behavior. The warning thresholds used in the method of Waldvogel (1.75 km), the VIL method (15 kg/m$^2$), and the CAPPI method (49 dBZ) are set to optimum performance, i.e., the highest CSI', at a positioning tolerance of 12.5 km (*vide infra*). It is evident that the performance of all detection methods as indicated by their CSI increases substantially when increasing the positioning tolerance. The method of Walvogel, however, seems to gain the most from the increase of the positioning tolerance from 0 to roughly 15 km. In contrast to, e.g., the CAPPI method (altitude is 0.8 km above MSL), the method of Waldvogel uses strong radar echoes at higher altitudes (4-8 km), and therefore the horizontal spread of the on-ground hail occurrences with respect to the radar echoes is expected to be larger.

The average area of municipalities in the Netherlands sets a lower limit to the attainable position tolerance of about
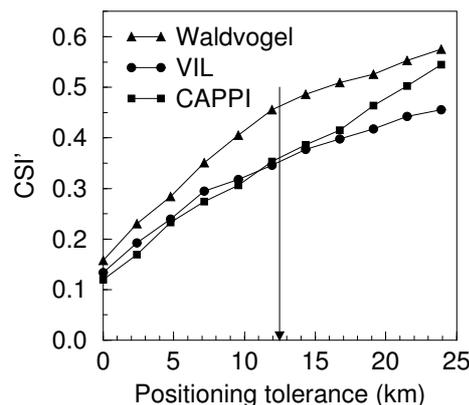


**Fig. 1.** The dependence of the apparent Critical-Success-Index (CSI') of three different hail detection methods on the allowed spatial mismatch between a radar pixel and an on-ground hail observation. The warning threshold of Waldvogel is set at 1.75 km, that of VIL at 15 kg/m$^2$, and that of CAPPI at 47 dBZ. The vertical arrow marks the positioning tolerance that is used throughout this study.

5 km. Taking into account a reasonable region of influence for a summertime thunderstorm and to be consistent with other studies of this kind (Kessinger et al. , 1995), however, a positioning tolerance of 12.5 km has been applied throughout this study. The applied positioning tolerance is marked in Fig. 1 with a vertical arrow.

The scoring parameters, i.e., the apparent POD', FAR', and CSI' of the five selected hail detection methods have been determined as a function of their warning thresholds. The results are shown in the five sub-plots of Fig. 2. For the method of Auer, the difference between the observed radar reflectivity and the reflectivity threshold of Auer, which is determined from the observed cloud-top temperature, is used as a hail indicator. Using this reflectivity difference, a warning threshold of 0 dBZ corresponds to an exact reproduction of Auer's method. The warning threshold for the method of Waldvogel is set at the difference between the maximum height of the 45 dBZ reflectivity and the height of the freezing level (in km). For the method based on the Severe Hail Index, the warning threshold is directly set on the SHI value in J/ms value, and thus the conversion of the SHI to probability of severe hail is not implemented.

Although there are large differences, the general trends in the scoring parameters as a function of the warning thresholds are rather similar for all methods. They show a decrease of both the POD' and the FAR' with increase of their warning threshold and a maximum of the CSI' at a certain threshold. The decrease of the FAR' with increase of the warning threshold, implies that, in accordance with expectations, the reliability of a detected event will increase when the warning threshold is raised.

As virtually all hail events will be accompanied by some precipitation and thus by a (weak) radar echo, a measure for the quality of the ground truth data can be deduced from the maximum POD of the CAPPI method. For this method, the observed POD is maximum 0.99 which indicates that at most
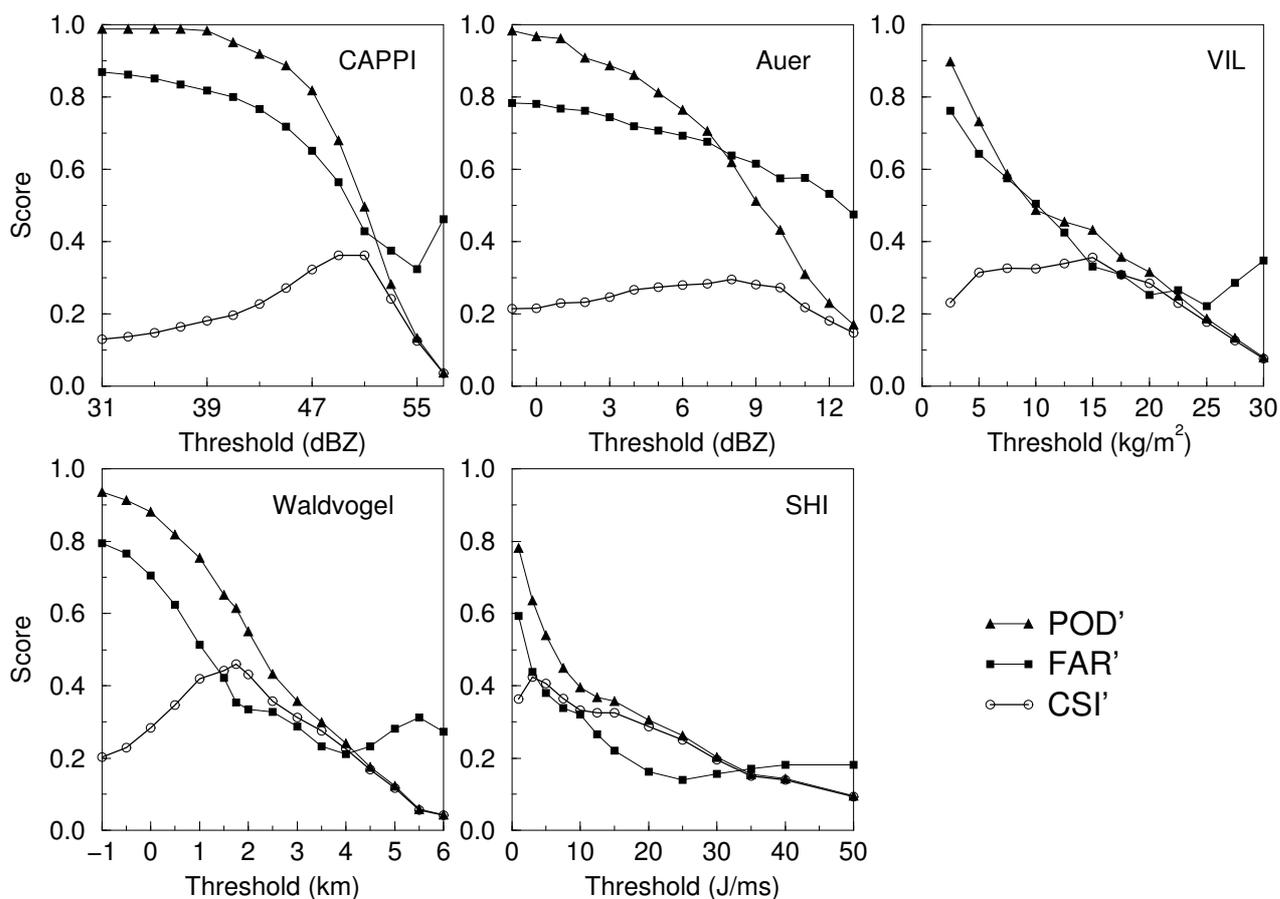
**Fig. 2.** The scoring parameters (POD$'$, FAR$'$ & CSI$'$) of the five different hail detection methods as a function of the warning threshold. The position tolerance is set at 12.5 km.

1% of the ground truth hail reports is suspect of being inaccurate or false. The lowest FAR$'$ is observed for the SHI method with a warning threshold of 25 J/ms, and it is about 0.14. The equation for the apparent FAR$'$, as given previously, states that this FAR$'$ cannot go below $(1 - \eta)$ even when the true FAR of a detection method is 0. Therefore, a lower limit for the fraction of reported hail events $\eta$, which is a property of the set of ground truth data only, can be determined from the lowest, apparent FAR$'$. Within the accuracy of the $\eta$-model and the positioning tolerance used, it is found that at least a fraction $\eta = 0.86$ of the hail events that have occurred are contained by the ground truth data used in this study.

The highest CSI$'$, i.e., the best performance, of 0.46 is observed for the method of Waldvogel using a warning threshold of 1.75 km. The method of Auer does not live up to the expectations. On our data it actually performs poorest of all methods considered. In addition, the optimum performance is shifted away substantially from a warning threshold of 0 dBZ which would correspond to an exact reproduction of Auer's method. The optimum performances of the CAPPI, VIL, and SHI methods are found at (much) lower warning thresholds than those reported in literature (Mason , 1971; Edwards and Thompson , 1998; Witt et al. , 1998).

The observed discrepancies may be explained by differences in both radar calibration and climatological conditions and, for the SHI method, by the fact that it is originally designed for detection of large hail and not for detection of hail of all sizes. In addition to the highest CSI$'$, the difference between the minimum and maximum values of the FAR$'$ as a function of the warning threshold is largest for the method of Waldvogel as well. A value of a radar pixel above a certain threshold indicates a "probability of hail" which is equal to $1 -$ FAR at that warning threshold. Therefore, the large difference between the minimum and maximum FAR$'$ enables the definition of several different thresholds with distinct warning properties, i.e., FAR$'$ and resulting POD$'$.

The fraction of reported hail events $\eta$ can be changed systematically by selecting municipalities with a certain probability of hail damage and considering the hail damage reports of the insurance companies only. Using a database of landuse in the Netherlands obtained from satellite observations between 1993 and 1995, the fraction of a municipality with hail-sensitive landuse, like crops, orchards, greenhouses, etc., is calculated. Subsequently, only the municipalities having a certain minimum fraction of hail-sensitive landuse are taken into account when the comparison of the detection methods against the hail damage reports is made.
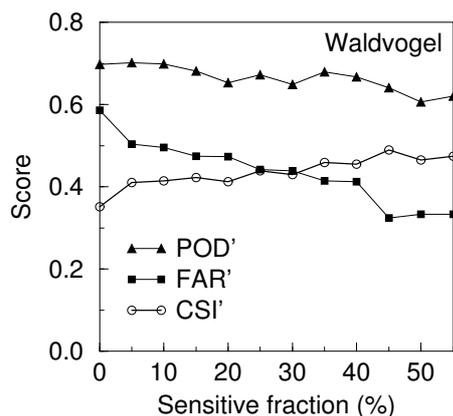
**Fig. 3.** The scoring parameters of the method of Waldvogel using a optimum, warning threshold of 1.75 km as obtained on the ground truth data of hail damage reports only. The minimum, required coverage fraction of hail-sensitive landuse of a municipality in order to be taken into account in the scoring parameters is varied.

In Fig. 3 the scoring parameters of the method of Waldvogel, obtained using the optimum warning threshold of 1.75 km, are shown as function of the minimum hail-sensitive landuse fraction of the selected municipalities. In accordance with the $\eta$-model, the POD is more or less constant, the apparent $FAR'$ is decreasing steadily, and the apparent $CSI'$ is increasing gradually when the minimum hail-sensitive fraction is increased, i.e., when the fraction of reported hail $\eta$ is increased. A maximum $CSI'$ of roughly 0.49 is obtained for the method of Waldvogel in this way. This is slightly higher than the maximum $CSI'$ observed using all ground truth data and all municipalities (see Fig 2), and probably it is close to the true CSI of Waldvogel's method.

The POD of roughly 0.70 is higher than that obtained for the method of Waldvogel using all ground truth data and the same warning threshold. On average, the hail damage reports will refer to more severe hail events than those reported by the observer network. The higher POD indicates that the method of Waldvogel and probably all methods will detect severe hail more effectively than light hail.

## 5 Conclusions

Five methods for the detection of summer hail using radar have been compared against on-ground hail observations and reports of hail damage. Altough there are substantial quantitative differences, the general trends in the apparent $POD'$, $FAR'$, and $CSI'$ of the detection methods as a function of the warning threshold are rather similar. The effects of missing ground truth data on the $POD'$, $FAR'$, and $CSI'$ of the methods has been described qualitatively using a simple model. Of all methods considered the one of Waldvogel scores best and that of Auer scores poorest. The CSI of 0.49 for the method of Waldvogel, as obtained in the present study, is somewhat higher than the CSI of 0.46 which can be deduced from the original verification results as presented by Wald-

vogel et al. (1979). The obtained CSI compares favorably to the results found for the verification of the Hail-Detection-Algorithm, which is based on Waldvogel's method, against events of hail larger than 6 or 13 mm in diameter (Kessinger et al. , 1995). Due to the large variation of the FAR as a function of warning threshold, the warning properties of the method of Waldvogel can be altered over a wide range to fulfill the needs of different kinds of users. Detection of hail using the method of Waldvogel will improve the performance significantly as compared to that using the CAPPI method, which is the present-day practice at the KNMI.

## References

Amburn, S. A., and Wolf, P. L., VIL Density as a Hail Indicator, *Wea. Forecasting*, *12*, 473–478, 1997.

Auer Jr., A. H., Hail Recognition through the Combined Use of Radar Reflectivity and Cloud-Top Temperatures, *Mon. Wea. Rev.*, *122*, 2218–2221, 1994.

Aydin, K., Seliga, T. A., and Balaji, V., Remote Sensing of Hail with a Dual Linear Polarization Radar, *J. Appl. Meteor.*, *25*, 1475–1484, 1986.

Edwards, R. E., and Thompson, R. L., Nationwide Comparisons of Hail Size with WSR-88D Vertically Integrated Liquid Water and Derived Thermodynamic Sounding Data, *Wea. and Forecasting*, *13*, 277–285, 1998.

Greene, D. R., and Clark, R. A., Vertically Integrated Liquid Water—A New Analysis Tool, *Mon. Wea. Rev.*, *100*, 548–552, 1972.

Hardaker, P. J., and Auer Jr., A. H., The Separation of Rain and Hail using Single Polarization Radar Echoes and IR cloud-top temperatures, *Met. Apps.*, *1*, 201–204, 1994.

Höller, H., Bringi, V. N., Hubbert, J., Hagen, M., and Meischner, P. F., Life Cycle and Precipitation Formation in a Hybrid-Type hailstorm Revealed by Polarimetric and Doppler Radar Measurements, *J. Atmos. Sci.*, *51*, 2500–2522, 1994.

Kessinger, C. J., Brandes, E. A., and Smith, J. W., A Comparison of the NEXRAD and NSSL Hail Detection Algorithms, *27th conf. on Radar Meteorology (AMS)*, 603–605, 1995.

Lenning, E., Fuelberg, H. E., and Watson, A. I., An Evaluation of WSR-88D Severe Hail Algorithms along the Northeastern Gulf Coast, *Wea. Forecasting*, *13*, 1029–1044, 1998.

Mason, B. J., The Physics of Clouds, Clarendon Press, Oxford UK, 1971.

Smart, J. R., and Alberty, R. L., The NEXRAD Hail Algorithm Applied to Colorado Thunderstorms, *14th conf. on Severe Local Storms (AMS)*, 244–247, 1985.

Smith, P. L., Effects of Imperfect Storm Reporting on the Verification of Weather Warnings, *Bull. Amer. Meteor. Soc.*, *75*, 1099–1105, 1999.

Smyth, T. J., Blackman, T. M., and Illingworth, A. J., Observations of oblate hail using dual polarization radar and implications for hail-detection schemes, *Q. J. R. Meteorol. Soc.*, *125*, 993–1016, 1999.

Waldvogel, A., Schmid, W., and Federer, B., The Kinetic Energy of Hailfalls. Part I: Hailstone Spectra, *J. Appl. Meteor.*, *17*, 515–520, 1978a.

Waldvogel, A., Federer, B., and Schmid, W., The Kinetic Energy of Hailfalls. Part II: Radar and Hailpads, *J. Appl. Meteor.*, *17*, 1680–1693, 1978b.

Waldvogel, A., Federer, B., and Grimm, P., Criteria for the Detection of Hail Cells, *J. Appl. Meteor.*, *18*, 1521–1525, 1979.

Wessels, H. R. A., and Beekhuis, J. H., Stepwise procedure for suppression of anomalous ground clutter, *COST-75 Seminar on Advanced Radar Systems*, *EUR 16013 EN*, 270–277, 1994.

Witt, A., Eilts, M. D., Stumpf, G. J., Johnson, J. T., Mitchell, E. D., and Thomas, K. W., An Enhanced Hail Detection Algorithm for the WSR-88D, *Wea. Forecasting*, *13*, 286–303, 1998.