Hail detection using single-polarization radar

Iwan Holleman

Scientific Report, KNMI WR-2001-01, 2001

Other publications about the hail detection project are:

- I. Holleman: 2000, Detectie van zomerhagel met radar. *Meteorologica*, **2**, 21–25 (Dutch).
- I. Holleman, H. R. A. Wessels, J. R. A. Onvlee, and S. J. M. Barlag, 2000: Development of a hail detection product. *Phys. Chem. Earth B*, **25**, 1293–1297.
- I. Holleman, H. R. A. Wessels, J. R. A. Onvlee, and S. J. M. Barlag, 2000: Development of a radar-based hail-detection-product. *20th conference on severe local storms*, AMS, 477–480.

Contents

Chapter 1

Introduction

Within the framework of the project "Development of a product for detection of severe weather phenomena using non-Doppler radar data" (KNMI project 2301), a tool for the detection and display of severe weather phenomena related to convective systems, like wind gusts and summer hail, is being developed. Currently, 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 "pseudo constant-altitude plan-position indicator" (pseudoCAPPI) of the radar reflectivity and an echotop product, which presents the maximum height of the echo for each pixel, are extracted. Ground clutter is removed from the pseudoCAPPI image using a statistical method (Wessels and Beekhuis, 1997). This report describes the development of the first new sub-product: a tool for the detection and display of summer hail. This hail detection product is to be used for nowcasting of the development and movement of summertime thunderstorms. In addition, an archive of this hail detection product could be a useful reference for e.g. insurance companies.

1.1 Summer hail

Hail related to summertime thunderstorms is a small-scale phenomenon, and it often has a short time duration. Hail will be produced by a thunderstorm for typically a fraction of an hour, and the width of the track of hail damage can be as narrow as a few hundred meter. Due to the wide spacing of manned stations observing synoptically every hour (in the following refered to as "synop" stations), about 1 station every 1800 km^2 in the Netherlands, events of summer hail are often not observed at these stations. On average the 19 synop stations in the Netherlands have reported hail only once per station during the summer of 1999, but there have been about 60 days with one or more hail events during that summer. Due to the potentially strong updrafts in a thunderstorm and the possibly high water vapor content of the atmosphere on a warm summer day, hail produced by a summertime thunderstorm can reach fairly large diameters. A thunderstorm producing hail with a diameter of 2 cm or more has been reported

Table 1.1: Description and the maximum terminal velocity of hail as a function of its diameter. The desciptions are taken from Ludlam (1980), and the velocities from WMO (1966).

Kind	Diameter (cm)	Velocity (m/s)	Description
Small	<0.5		Grain
	0.5-1.0		Pea
	1.0-1.5		Mothball; small marble
	1.5-2.0	19	Cherry; marble
Large	2.0-2.5		Large marble
	2.5-3.0	24	Walnut
	3.0-4.0	28	Golfball
	4.0-5.0	31	Small egg
Giant	5.0-6.2	34	Egg
	6.2-7.5	38	Tennis ball

during that summer on seven days only. In Table 1.1 the description and terminal velocity of hail as a function of its diameter is given. The descriptions, compiled by Ludlam (1980), are used for estimation of the size of hail when no measuring device is available. Large summer hail with its large mass and high terminal velocity can cause severe damage and is potentially dangerous. Fortunately, hail with large diameters is rare: the probability is reduced by a factor of 10 with every cm of increase of the diameter (Ludlam, 1980). Not only large hail can cause damage, however. Approximately 2100 reports of hail damage have been received by three agricultural insurance companies in the Netherlands during the summer of 1999, and this is not an exceptionally high number.

In order to gain insight in the probability of observing hail at a certain location and the seasonal variation thereof, five years of upper-air sounding data at De Bilt and synop observations of hail in the Netherlands have been analysed. The height of the freezing level has been calculated from the upper-air sounding data of 12 UTC for each day. The number of days and the number of hail observations that fall within certain freezing-level-height classes (four classes per km) have been counted. Subsequently, the probability of a hail observation by a synop station has been calculated from the ratio between the number of hail observations and the number of days within a certain freezing-level-height class, and it has been normalized by the number of synop stations in the Netherlands (19). The probability of observing hail is shown as a function of the height of the freezing level in the upper frame of Figure 1.1, and the number of days per year in these height classes is shown in the lower frame. It is evident from this figure that the probability of hail depends strongly on the height of the freezing level.

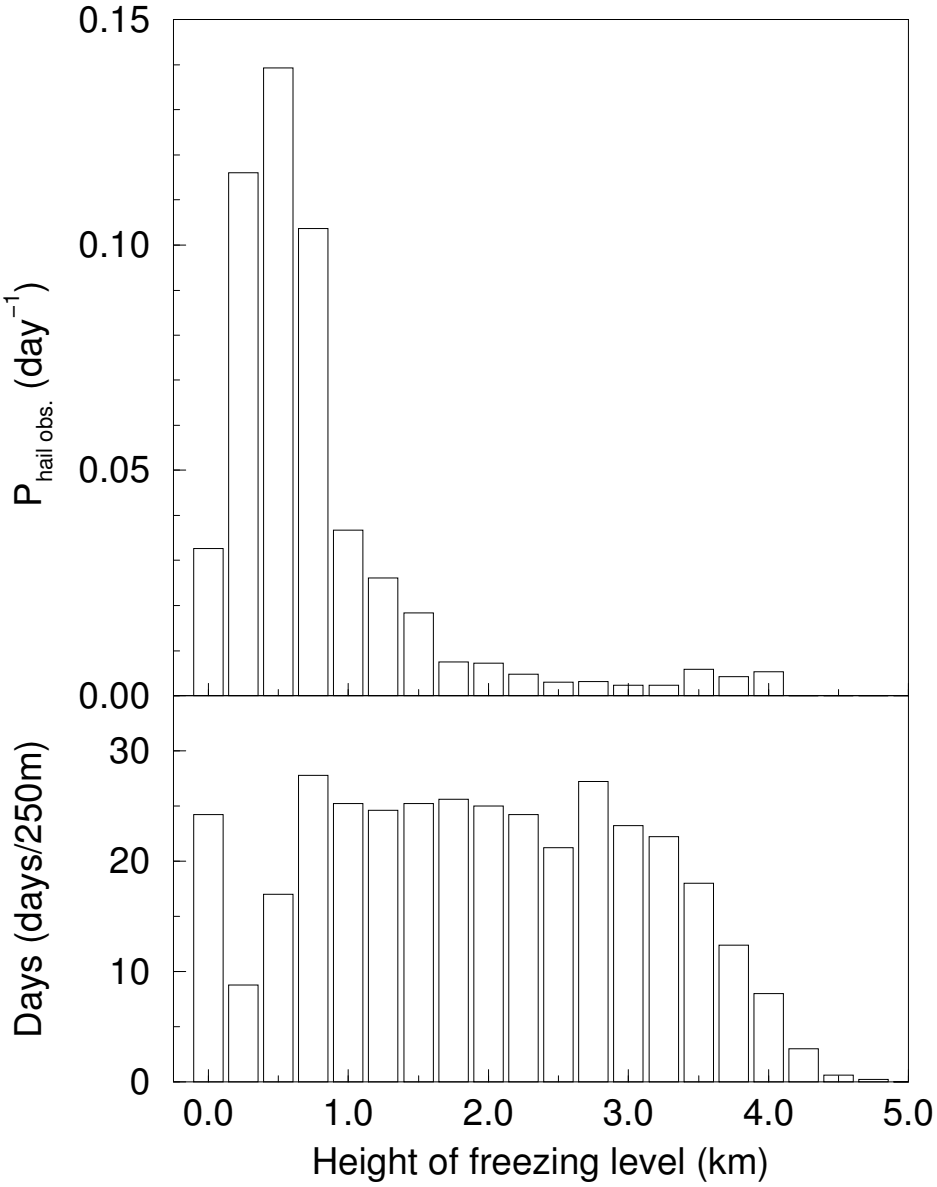

Figure 1.1: The upper frame of this figure shows a histogram of the daily probability that hail will be observed by a synop station in the Netherlands as a function of the height of the freezing level at that day. The lower frame shows a histogram of the number of days per year as a function of this height. Five years, ending on November 7 of 2000, of upper-air sounding data at De Bilt (12 UTC) and synop reports of all manned stations in the Netherlands have been used to compile this figure. The number of days in the 0 km-class is relatively high because days where the freezing level is below mean sea level have been counted in this class as well.

For days when the freezing level is at 1.5 km or lower, the probability of observing hail is relatively high, and it increases strongly when the freezing level is descending to 0.5 km. When the freezing level is this close to the surface, small hail formed due to a large-scale (weak) vertical motion, e.g. due to the passage of a front, will be able to reach the surface without melting. This type of small hail on a large scale, which is mostly occurring in winter, is dubbed "winter hail". For days when the freezing level is at 1.5 km or higher, the probability of observing hail is relatively low and constant, and it seems to have a secondary maximum between 3.5 and 4.0 km. When the freezing level is at these high altitudes, small hail will melt before it reaches the surface and consequently mostly larger hail will be observed. For such large hail to grow, strong updrafts are required which can exist locally in thunderstorms provided that the temperatures at the surface are high, i.e., when the freezing level is at high altitudes. This type of larger hail on a small scale, related to summertime thunderstorms, is dubbed "summer hail". Using the height of the freezing level, the type of hail can, therefore, be identified .

1.2 Outline of this report

The development of the radar-based hail detection product has been divided into the following actions. From the literature eight methods for the detection of summer hail using single-polarization radar have been selected. In chapter 2 of this report the selected detection methods and some other methods are described in detail. Five of the selected methods use radar data only, while two others use additional information on the vertical temperature profile and one other uses additional information on the cloud-top temperature determined by a geostationary satellite, e.g Meteosat. The performances of these eight different hail detection methods have been compared using data of 15 selected days with thunderstorms in the Netherlands during the summer of 1999.

In chapter 3 the available data of the selected days of the summer of 1999 is described. For the construction of the different detection methods, three-dimensional radar data, Meteosat calibrated infrared data, analyses of the HiRLAM numerical weather prediction model (Machenhauer, 1988), and in situ observational data, like upper-air soundings and synops, have been gathered for the selected days. For verification of the methods, hail observations by synop stations and precipitation stations, reports of hail damage from three agricultural insurance companies, and reports by weather amateurs and newspapers have been collected.

In chapter 4 the calculation of the different detection methods from the radar and model data is described. In addition, the processing of the obtained hail detection images and of the hail observations and damage reports prior to the verification process is outlined. Finally, the classification of the hail events during the verification process, the deduced verification scores, and the effect of missing hail reports on the verification scores are discussed. In chapter 5 the results of the verification of the eight hail detection methods are presented. The performances

of the different methods are compared in detail, and a choice for a certain hail detection method is made.

The results of the extended verification during the summer of 2000 of the chosen detection method, i.e., the method of Waldvogel, are presented in chapter 6. From early May till end of September of 2000, a hail detection method based on Waldvogel's method has been run semi-operationally. This run has resulted in an extended set of radar data compromising 135 days. The obtained dataset has been used to investigate and characterize the warning behavior of the method of Waldvogel in much more detail. In the last chapter, the final conclusions and recommendations for operational implementation of a hail detection product based on Waldvogel's method are given.

Chapter 2

Methods for detection of hail

2.1 Introduction

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 hail stones and the non-spherical rain droplets. Apart from the standard reflectivity for horizontal polarization (Z_H), a dual-polarization radar also measures the ratio between horizontal and vertical polarization (Z_V) reflectivities. The differential reflectivity (Z_{DR}) is defined as:

$$Z_{DR}(dB) = 10 \cdot^{10} \log(\frac{Z_H}{Z_V}) \tag{2.1}$$

In rainfall Z_{DR} is always positive, generally varying between 0 and 4 dB and being correlated with Z_H, while in hail storms Z_{DR} is approximately 0 dB and Z_H has a high value. Using these differences Aydin et al. (1986) have derived a hail signal (H_{DR}) which is based on both Z_H and Z_{DR}, and they have used it to study two major hailstorms in Colorado. Recently, dual-polarization radar has been used to study properties of hail stones and hailstorms (Smyth et al., 1999; Höller et al., 1994).
As only single-polarization radars are currently in operation at KNMI and operational implementation of the dual-polarization technique at other meteorological institutes is rare up to now, hail detection methods for operational use still have to rely on single-polarization radar data possibly complemented with data from other sources.

2.2 CAPPI-method

The first method which is used to distinguish hail from rain using a single-polarization radar is based on a plan-position indicator (PPI) of the radar reflectivity at constant altitude (CAPPI display). At short (large) ranges from the radar site, no reflectivity data can be obtained at the

desired altitude due to the too low (high) altitude of the radar beam. When the missing data at short (large) ranges is completed with data from the highest (lowest) available elevation, the resulting product is a so-called "pseudoCAPPI". At KNMI, these pseudoCAPPIs are calculated for an altitude of 0.8 km above mean sea level, but beyond a range of 100 km data from higher altitudes is presented. Assuming that the diameters (D_i) of the scattering particles are (much) smaller than the wavelength of the radar radiation, implying that Rayleigh scattering wil be dominant, the radar reflectivity Z can be written as:

$$Z = \sum_i n_i \cdot D_i^6 \tag{2.2}$$

where n_i is the number of particles per unit volume having diameter D_i. Because the radar reflectivity increases dramatically with increasing diameter of the scattering particles, larger hail stones (>10 mm) will give rise to higher reflectivities than would be possible for rain droplets, which have a maximum diameter of about 6.5 mm. A thunderstorm producing hail stones with a diameter of 10 mm gives rise to a radar reflectivity of 54 dBZ (Auer, 1994), assuming that the hail stones are produced over the entire volume of the radar beam. From Table 2.1 it can be estimated that a reflectivity of 54 dBZ would correspond to a rainfall rate of about 87 mm/h, which is a rare event. Mason (1971) has suggested to use a reflectivity threshold of 55 dBZ for distinguishing between severe rain and hail when using this CAPPI method.

2.3 Maximum-reflectivity method

In the Rainbow processing software of the Gematronik radars, a hail warning product is present which is a simple extension of the CAPPI method. Instead of taking the reflectivity at a fixed altitude, the maximum observed reflectivity between a minimum and maximum altitude is used. The display of this "maxPPI" is then used to issue hail warnings. In this way, a (developing) reflectivity core above the level of the low-altitude CAPPI is still detected.

Table 2.1: Radar reflectivity Z in dBZ and the approximate rainfall rate R on the ground.

Z [dBZ]	7	15	23	31	39	47	55
R [mm/h]	0.1	0.3	1	3	10	30	100

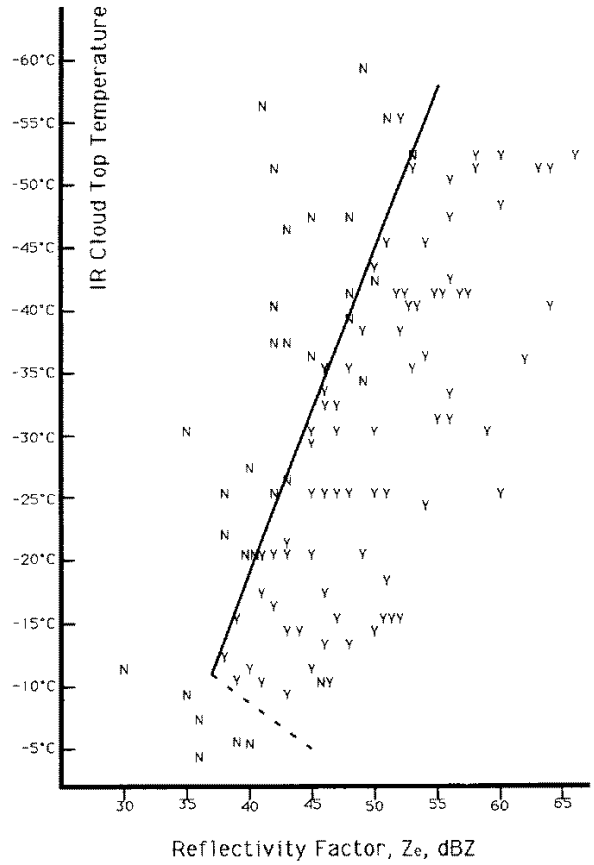

Figure 2.1: The distribution of hail (Y) or rain (N) events as a function of low-level CAPPI reflectivity and convective cloudtop temperature. The best-fit line, both the solid and the dashed part, which discriminates between the events, is presented in equation 2.3 in the text (figure taken from Auer (1994)).

2.4 Method of Auer

Auer (1994) has reported on the detection of hail using a combination of radar reflectivity at low altitude and cloud-top temperatures. The cloud-top temperatures can be determined from the infrared radiance measured by a geostationary satellite, e.g. Meteosat. The nomogram for hail and heavy rain events as obtained by Auer (1994) is depicted in Figure 2.1. The nomogram has been constructed using over 100 hail or rain cases during all seasons throughout New Zealand. An impressive differentiation between hail and heavy rain events is evident from this figure. The best-fit warning threshold for the CAPPI reflectivity (Z_{Th}) as a function of the cloud-top temperature (T_{top}) is given by Hardaker and Auer (1994):

$$Z_{Th} = \begin{cases} -0.38 \cdot (T_{top} - 85.0) & \text{if } T_{top} \leq -11°\text{C} \\ 1.33 \cdot (T_{top} + 38.8) & \text{if } T_{top} > -11°\text{C} \end{cases} \tag{2.3}$$

This optimum threshold varies between 36 and 53 dBZ for cloud-top temperatures in the range between -11 and -55 °C. The method of Auer has been verified during an all season operational evaluation in New Zealand, and it performs much better than the ordinary, fixed threshold CAPPI method. In addition, Auer (1994) has been able to estimate the diameter of the hail stones near the melting layer or at the surface. In a sequel to this study, Hardaker and Auer (1994) have attempted to separate the contributions of rain and hail from the total reflectivity signal. The part of the observed radar reflectivity that exceeds the warning threshold is attributed to hail and the part up to the warning threshold is attributed to rain. The maximum possible rain rate is thus determined by the cloud-top temperature.

2.5 NEXRAD hail detection algorithm

Within the framework of the operation of the network of WSR-88D radars in the United States, the NEXRAD-project, several hail detection algorithms have been developed and tested. Kessinger et al. (1995) have compared several different algorithms used or to be used by NEXRAD. In Figure 2.2, the designs of the different hail detection algorithms are shown schematically.

The original version of the NEXRAD hail detection algorithm, which is the left schematic in Figure 2.2, has been developed by Petrocchi et al. (Smart and Alberty, 1985; Kessinger et al., 1995) and uses a combination of seven hail indicators. The most important indicators are the presence of a reflectivity core of 50 dBZ or higher somewhere between 5 and 12 km altitude and the presence of radar echotops higher than 8 km, but also a mid-level overhang of more than 4 km is included. The seven hail indicators are combined, using different weights, into one hail index. The output of the algorithm has only four possible outcomes: hail, probable hail, no hail, and insufficient data. This NEXRAD hail detection algorithm has recently been replaced by an algorithm which is developed by Witt et al. (1998) and which produces a probability of hail. This new NEXRAD algorithm, which is actually more straightforward than the old one, is seen to perform significantly better (Kessinger et al., 1995). In the center schematic of Figure 2.2, the design of the new hail detection algorithm, based on work by Waldvogel et al. (1979), is shown.

While examining one of the Soviet seeding criteria used to detect hail cells at an early stage of development, Waldvogel et al. (1979) reached the conclusion that the criterion can be replaced by a simpler and slightly more efficient method. The method of Waldvogel for the detection of hail uses the maximum altitude at which a reflectivity of 45 dBZ is found (H_{Z45}) in relation to the height of the freezing level (H_{T0}). In Figure 2.3 the frequency distribution of

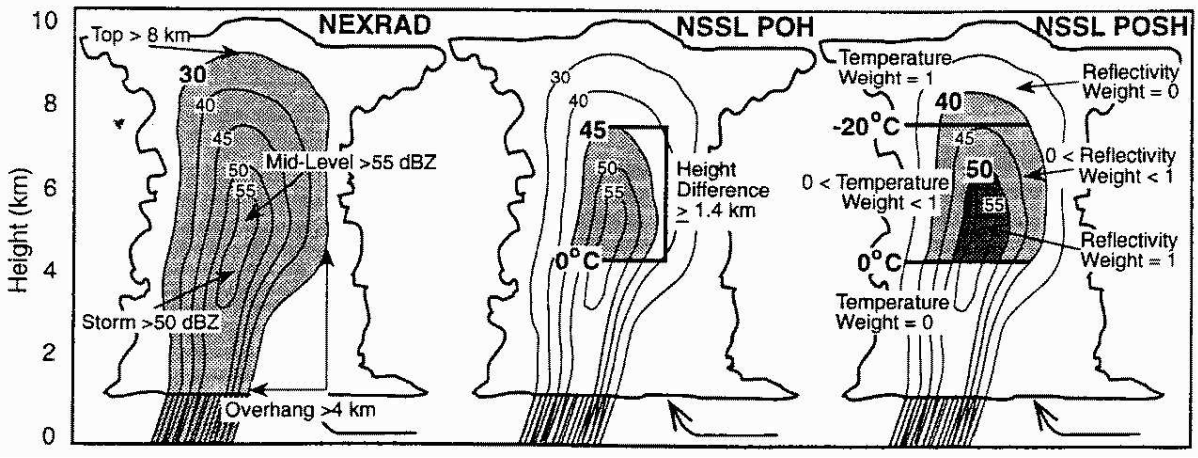

Figure 2.2: Schematic of a hail storm and the design of each Hail Detection Algorithm (HDA). The old NEXRAD HDA examines particular characterisctics, such as maximum storm reflectivity and overhang, and applies a weight. The new NEXRAD HDA computes the height between the freezing level and the maximum height of the 45 dBZ reflectivity and applies a probability curve. The possibility of severe hail is calculated using the severe hail index (SHI). The Severe Hail Index (SHI) is calculated by vertical integration of the product of the hail kinetic energy with a temperature function and a reflectivity function (figure taken from Kessinger et al. (1995)).

the parameter ($H_{Z45} - H_{T0}$) for strong rain cells and strong hail cells as observed by Waldvogel et al. (1979) is shown. It is evident from this figure, that when the 45 dBZ-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 method of Waldvogel combines an indicator for the presence of a substantial updraft, the height of the strong reflectivity core (45 dBZ), with that for a large amount of undercooled water and/or ice, the reflectivity core above the freezing level, to detect (developing) hail. Waldvogel et al. (1979) used radiosonde data to obtain the height of the freezing level. Nowadays, however, this height can also be determined from a numerical weather prediction model. In the current NEXRAD hail detection algorithm, the maximum height of the 45 dBZ reflectivity above the freezing level is converted to a probability of hail using the curve depicted in Figure 2.4. A height difference of 1.6 km corresponds to 10% probability of hail and one of 6.0 km to 100% probability of hail (Witt et al., 1998).

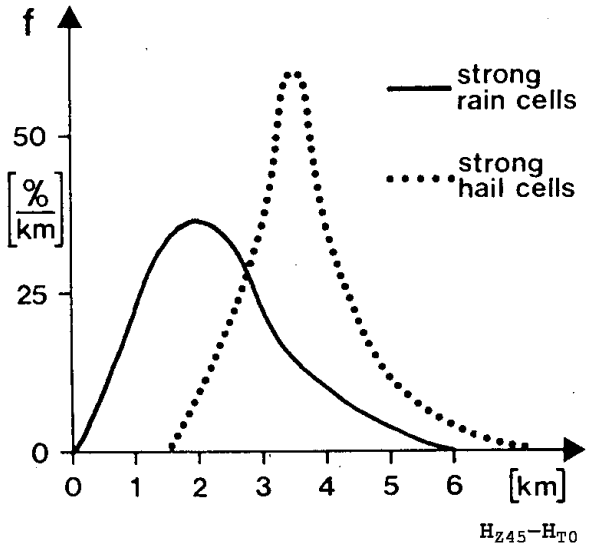

Figure 2.3: The normalized frequency distributions of the parameter ($H_{Z45} - H_{T0}$) for strong rain cells and strong hail cells (figure modified from Waldvogel et al. (1979)).

2.6 NEXRAD severe hail algorithm

The NEXRAD hail detection algorithm as developed by Witt et al. (1998) also contains a part that attempts to estimate the probability of severe hail. For this, a semi-empirical relationship between the kinetic energy flux of the hail stones ($\dot{E}$) and the reflectivity as found by Waldvogel et al. (1978a,b) is used. By measuring size distributions at six different sites for four severe hail storms, Waldvogel et al. (1978a,b) have obtained a total of 175 hailstone size distributions. In Figure 2.5 the quantities $\dot{E}$ and Z, calculated from the measured size distributions while assuming a certain dependence of the terminal velocity on the diameter of the stones, have been plotted. The best-fit line gives the relation between hail kinetic energy flux in J/m^{2}s and the reflectivity Z in mm^6/m^3:

$$\dot{E}(Z) = 5.0 \times 10^{-6}\, Z^{0.84} \tag{2.4}$$

Using this hail kinetic energy flux, a "Severe Hail Index" (SHI) is calculated by vertically integrating the obtained flux weighted with a reflectivity-based $W(Z)$ and a temperature-based $W_T(H)$ function:

$$\text{SHI} \equiv \frac{1}{10} \int_0^{H_{top}} W(Z(h)) \cdot W_T(h) \cdot \dot{E}(Z(h))\, dh \tag{2.5}$$

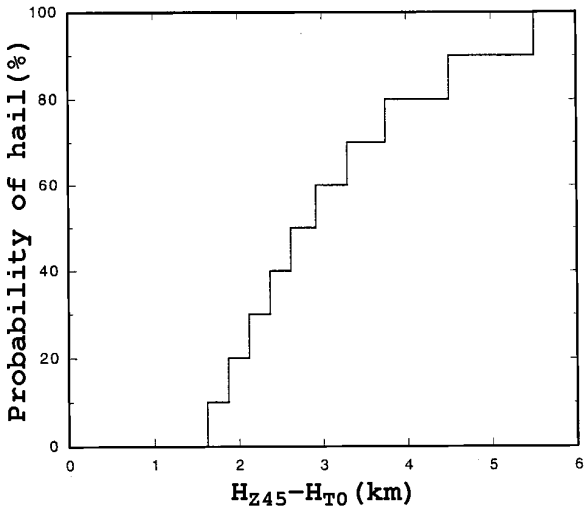

Figure 2.4: Probability of hail at the ground as a function of $(H_{Z45} - H_{T0})$. Here H_{Z45} is the height of the 45 dBZ echo above radar level, and H_{T0} is the height of the melting level above radar level (figure modified from Witt et al. (1998)).

where the reflectivity-based weighting function is defined as:

$$W(Z) = \begin{cases} 0 & \text{for } Z \leq Z_L \\ \frac{Z - Z_L}{Z_U - Z_L} & \text{for } Z_L < Z < Z_U \\ 1 & \text{for } Z \geq Z_U \end{cases} \tag{2.6}$$

and the temperature-based weighting function as:

$$W_T(H) = \begin{cases} 0 & \text{for } H \leq H_{T0} \\ \frac{H - H_{T0}}{H_{Tm20} - H_{T0}} & \text{for } H_{T0} < H < H_{Tm20} \\ 1 & \text{for } H \geq H_{Tm20} \end{cases} \tag{2.7}$$

where the reflectivity cut-off values Z_L and Z_U are set at 40 and 50 dBZ, respectively, and H_{T0} (H_{Tm20}) is the height of the $0°$C ($-20°$C) environmental temperature level. In this way, the severe hail index is primarily sensitive to high reflectivities at temperatures near $-20°$C or colder where most hail growth occurs. The heights of the $0°$C and the $-20°$C levels can be determined using either radiosonde data or data from a numerical weather prediction model. Subsequently, a warning threshold for the severe hail index is calculated from the height of the freezing level using an empirical relationship. Finally, the probability of severe hail is calculated from the severe hail index and the obtained warning threshold using again an empirical relationship. In a verification and comparison study performed by Kessinger et al. (1995), it is found that the NEXRAD severe hail algorithm detects "large" hail (diameter$\geq$13 mm) somewhat better than the "normal" NEXRAD hail detection algorithm.

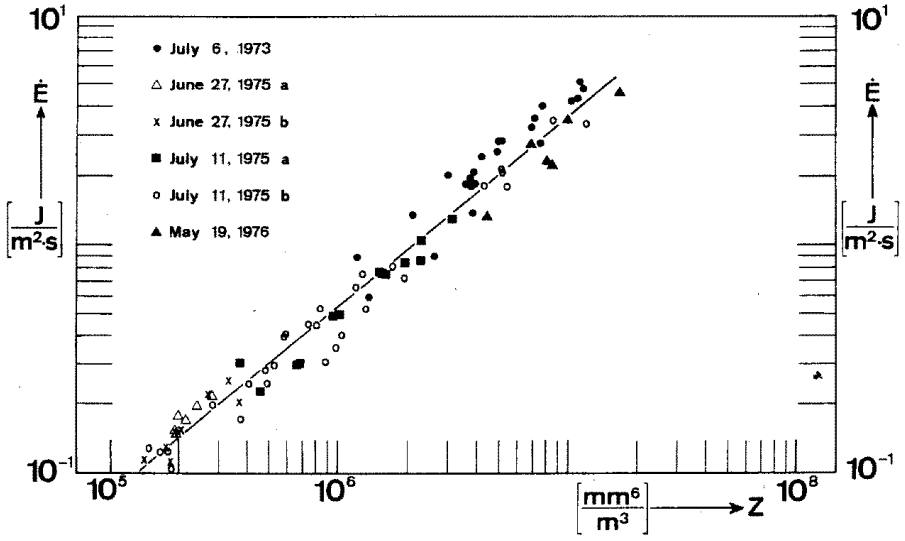

Figure 2.5: The real Z-$\dot{E}$ relations from four different hailstorms measured at six different measuring sites; 86 correlation points of the hailstone spectra each heaving an $\dot{E}$ value of > 0.1 Jm^{-2}s^{-1} are plotted. The coefficients of the regression line are given in equation 2.4 in the text (figure taken from Waldvogel et al. (1978a)).

2.7 Vertically Integrated Liquid water

The use of the entity "Vertically Integrated Liquid water" (VIL) as a new analysis tool has been introduced by Greene and Clark (1972), and they anticipated that this technique would be useful for both severe storm and hydrological applications. Using the VIL display in combination with the standard CAPPI and/or echotop display, it is possible to get an accurate view of the three-dimensional characteristics of a storm cell. The first step in the calculation of VIL is to convert all reflectivities to liquid water content (M) using the semi-empirical relation between M in g/m^3 and Z in mm^6/m^3:

$$M = 3.44 \times 10^{-3}\, Z^{4/7} \tag{2.8}$$

Subsequently, the obtained liquid water content at each location is integrated vertically:

$$\text{VIL} \equiv \int_0^{H_{top}} M \cdot dh = 3.44 \times 10^{-3} \int_0^{H_{top}} Z^{4/7}\, dh \tag{2.9}$$

where the VIL is expressed in kg/m^2 or in mm of "potential rainfall" and the height in km. The three-dimensional radar data is thus converted to a plan-position indicator of the amount of liquid water present in a vertical column above a certain position. Because the rate of precipitation formation is roughly proportional to the updraft velocity, VIL is a function of

both updraft and cloud depth (Kitzmiller et al., 1995). A high value of VIL correlates well with the occurrence of severe thunderstorms and hail. In stratiform situations VIL rarely exceeds a value of 10 kg/m^2, in thunderstorms, however, VIL is usually (much) higher. There is, however, no agreement in literature on the best warning threshold for the detection of hail with the VIL method. Forecasters in the United States 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).

2.8 VIL-density

In an attempt to eliminate the problems with thresholds for VIL-based hail warnings, Amburn and Wolf (1997) have proposed to "normalize" the VIL value using echotop heights of a certain reflectivity threshold, for instance 7 dBZ. This would capture the observation that some high-topped thunderstorms do not produce hail and some low-topped thunderstorms with low VIL values do produce hail. The "VIL-density" is defined by Amburn and Wolf (1997) as follows:

$$\text{VIL-density} \equiv \frac{\text{VIL}}{H_{top}} \tag{2.10}$$

where the VIL-density will be in g/m^3 when VIL is given in kg/m^2 and H_{top} in km. Amburn and Wolf (1997) have proposed a universal VIL-density threshold of 3.5 g/m^3 for issuing hail warnings. Unfortunately, the advantage of the use of this "universal" VIL-density threshold over just VIL is disputed. Edwards and Thompson (1998) note that the use of a warning threshold of 38 kg/m^2 for VIL on the data of Amburn and Wolf (1997) would result in the same performance as the use of the VIL-density threshold. In addition, they note that for VIL values exceeding 43 kg/m^2 hail is always observed independent of echotop height. Currently, both the severe hail index (SHI) and the VIL are used by forecasters in the United States to detect (severe) hail. Lenning et al. (1998) have evaluated these two methods for detection of hail, and they have concluded that the VIL method performs as well as the severe hail index when the best VIL threshold can be determined in advance.

2.9 Regression equations

Several attempts to predict hail and hail stone diameters using statistical analysis of sounding data and radar data have been made. Billet et al. (1997) have derived multiple regression equations and logistic regression equations for the prediction of hail size. The variables used in this study are vertically integrated liquid water (VIL) computed from two WSR-88D radars in the Washington D.C. area and several convective parameters derived from upper-air soundings taken in the same area. The convective parameters, that have been used, are amongst

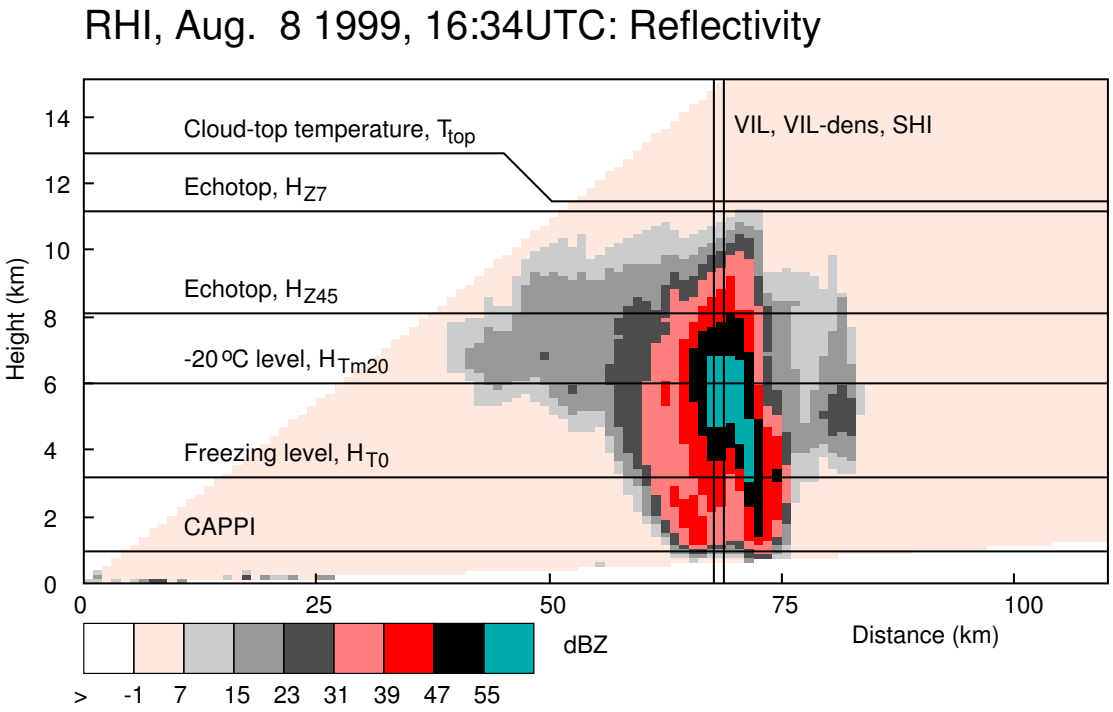

Figure 2.6: Vertical cross-section through volume data of the De Bilt radar at an azimuth of 187 degrees, a so-called range-height indicator (RHI). Several parameters used in the different hail detection methods have been indicated. Note that the lowest level (-1 dBZ) is merely used to indicate the reach of the radar scan.

others: the lifted index, the convectively available potential energy (CAPE), the temperature at 850 hPa, the height of the freezing level, and the mean storm-relative inflow. Billet et al. (1997) have found in this study that the most significant predictor variable for hail size is VIL. In addition, Billet et al. (1997) have developed a logistic regression equation which indicates the probability of large hail (>19 mm), and again it turned out that VIL was the most significant predictor. Billet et al. (1997) have concluded that it is possible to develop a site-specific (for the Washington D.C. area in this case) equation to predict the probability of large hail, and that regression equations for prediction of hail size are of limited use only. Edwards and Thompson (1998) have attempted to develop a regression equation for the prediction of the hail size as well, but they have used WSR-88D vertically integrated liquid water data and analysed upper-air sounding data from the entire United States. As a result of the analysis of 426 severe hail events in the United States, some parameters have been found that on average increase with larger hail-size categories, but the specific hail size varies widely across the spectra of VIL and sounding variables. Edwards and Thompson (1998) have concluded that on a nationwide basis, commonly used hail predictors showed little or no skill in predicting hail size.

Table 2.2: An overview of the hail detection algorithms which will be compared and verified in the present study.

Method	Short description
CAPPI	Reflectivity at constant (low) altitude
maxPPI	Maximum reflectivity observed in vertical column
Auer	CAPPI method with threshold depending on cloud-top temperature
Waldvogel	Echotops of 45 dBZ reflectivity relative to height of freezing level (NEXRAD)
Echotop	Echotops of 45 dBZ, i.e., Waldvogel without temperature data
SHI	Severe Hail Index, reflectivity and temperature weighted integral of hail kinetic energy flux (NEXRAD)
VIL	Vertically Integrated Liquid, potential rainfall
VIL-dens	VIL density, ratio between VIL and height of 7 dBZ echotops

2.10 Summary

From the literature survey, it appears that several methods for the detection of hail and estimation of hail size using single-polarization radar exist. In addition, it seems that the methods for the detection of hail are in general more successful than those estimating the hail size. Several methods use temperature data from upper-air soundings, numerical weather prediction models or calibrated infrared measurements by a geostationary satellite as an additional source of information. The NEXRAD hail detection algorithm based on the method of Waldvogel is widely used and has proved to be quite reliable. The method developed by Auer in New Zealand, which is currently being used at, amongst others, the UK Met Office, is promising. Although there are difficulties in determining the threshold, the display of the vertically integrated liquid is often being used to issue hail warnings as well. In Table 2.2 an overview of the eight hail detection methods that will be compared and verified in this study is given. In addition, a short description of the essence of the methods is given. To give a graphical overview of these methods, a vertical cross-section through a volume dataset, a so-called range-height indicator (RHI), of the De Bilt radar is plotted in Figure 2.6. Several of the parameters used in the different hail detection methods, like the height of freezing level and the height of 7 dBZ and 45 dBZ echotop, are shown together with a schematic view of some of the methods themselves. As most hail detection methods have been developed elsewhere, the methods probably have to be re-tuned to account for differences in the climatological conditions and in the radar systems.

Chapter 3

Data from the summer of 1999

A systematic comparison and verification of the eight selected hail detection methods has been performed. For this, a database had to be compiled consisting of three-dimensional radar data, Meteosat calibrated infrared data, upper-air sounding data, synop observations, and analyses of the HiRLAM numerical weather prediction model (Machenhauer, 1988). This data was gathered for selected days with thunderstorms above the Netherlands. In Table 3.1 an overview is given of the days in the summer of 1999 that have been selected, as well as a short description of the synoptic situation.

In order to verify the outcome of the different hail detection methods, observations of hail events at ground level are needed. These data have been obtained from KNMI synop stations, from the volunteers of the precipitation observing network, from hail damage reports from three agricultural insurance companies, and from the observations by weather amateurs.

3.1 Radar data

KNMI operates two C-band Doppler radars manufactured by Gematronik (type Meteor 360 AC). One is located in De Bilt (5.179E, 52.103N, 44 m above mean sea level) and the other one is located in Den Helder (4.791E, 52.954N, 48 m above mean sea level). These radars transmit pulses, generated by a magnetron, with a peak power of 250 kW and a duration of 2 μs. The 4.2 m diameter parabolic reflector produces a beam with an 1.0 degree width for C-band radiation. The signal from the logarithmic receiver is preprocessed by a RVP-6 signal processor (Sigmet) and processed further by the Rainbow system (Gematronik). The sensitivity of the radars is such that a reflectivity of 7 dBZ can still be observed at a range of about 550 km from the radar site. The two radars are performing low-elevation volume scans every 5 minutes and extensive volume scans every 15 minutes. From these scans, a pseudo constant-altitude plan-position indicator (pseudoCAPPI) of the radar reflectivity and a display of the echotop heights are produced operationally. Ground clutter is removed from the CAPPI product using a dynamical cluttermap which is updated every 5 minutes using a statistical method.

Table 3.1: An overview of the days in the summer of 1999 used in the comparison of the different hail detection methods. For each day a short description of the synoptic situation is given.

Date	Description of synopsis
August 5	cold front system migrating from southwest to northeast.
August 8	unstable atmosphere, many thunderstorms.
August 10	same as August 8, 55 mm precipitation in Rotterdam.
August 14	cold front passing in the evening
August 15	jetstream just south of Netherlands, unstable atmosphere, many thunderstorms.
August 16	same as August 15.
August 17	same as August 15.
August 25	warm and humid air from the south, cold front migrating to northeast during late evening.
August 26	continuation of August 25.
September 6	small depression moves over Netherlands to northeast, thunderstorms, 95 mm precipitation in Oosterhout.
September 24	warm and humid air from the south, cold vortex aloft over Great Britain, unstable atmosphere, thunderstorms with significant amounts of precipitation.
September 26	same as September 24.
September 30	same as September 24.
October 3	passage of trough over warm North Sea in late evening, heavy thunderstorms, large amounts of precipitation (70 mm) in Westland.
October 4	continuation of October 3.

This statistical method is based on the difference between the pulse-to-pulse fluctuations of the signal originating from precipitation and those of the signal received from ground targets (Wessels and Beekhuis, 1997).

The extensive volume datasets with reflectivity data obtained by the radar in De Bilt have been collected for each of the selected days (96 per day). These volume datasets have been obtained using a pulse repetition frequency of 400 Hz, a maximum range of 320 km, and an azimuthal speed of 4 rpm. These datasets contain data as a function of azimuth and range taken at 14 different elevations, which are: 0.3, 0.8, 1.3, 1.8, 2.3, 2.8, 3.3, 4.0, 5.0, 6.0, 7.5, 9.0, 10.5, and 12.0 degrees. The azimuth and the range resolution of the volume dataset are 1.0 degree and 1.0 km, respectively. The dynamical cluttermaps produced by the radar in De Bilt have been collected for the selected days as well.

3.2 Meteosat data

The Meteosat satellite, currently Meteosat-7, is in a geostationary orbit 36,000 km above the crossing of the equator and the Greenwich meridian. The satellite is operated by the European Meteorological Satellite organisation (EUMETSAT). The spinning satellite takes line scans of the face of the earth from south to north every half hour using a visible channel (0.4-1.0 μm), a water vapor channel (5.8-7.0 μm) and an infrared channel (10.6-12.5 μm). The angular resolution of the detection system is about 1.25×10^{-4} rad which leads to a spatial resolution of roughly 5 by 8 km^2 at a latitude of 52 degrees. The Meteosat images are navigated and calibrated at the ground control office of Meteosat in Darmstadt (Germany). For the selected days, the half-hourly calibrated infrared images of Meteosat containing a part of the full image have been collected.

3.3 Other observational data

Other observational data like upper-air soundings and synoptic observations are collected as well. The soundings are, at least for De Bilt, available every six hours and the synop observations on an hourly basis. The sounding data, the synoptic observations, and other observational data covering the whole of Europe are stored in BUFR code. These BUFR files, which contain observations accumulated over three hours, are actually intended for assimilation in the KNMI numerical weather prediction model HiRLAM (*vide infra*).

3.4 Model data

KNMI uses operationally a High Resolution Limited Area Model (Machenhauer, 1988). The analysis of surface pressure, atmospheric temperature, wind, and humidity is based on an op-

timum interpolation scheme. A data assimilation cycle of three hours is used, and the lateral boundary values are given by the large-scale (global) model of the European Centre for Medium range Weather Forecasts (ECMWF). The HiRLAM model is a hydrostatic gridpoint model, and at KNMI the model is operated using a horizontal resolution of 55 km and 31 levels in the vertical. An upgrade to a higher horizontal resolution is planned for the near future. The coordinate systems used are a rotated latitude-longitude grid horizontally and a hybrid p-σ system, i.e., a combination of pressure and relative pressure, in the vertical.

Every three hours an analysis of the HiRLAM model is available, and every six hours forecasts in one-hour steps up to 48 hours ahead are produced. Only the eight analysis files, which are stored in the GRIB format, have been collected in the database for each of the selected days.

3.5 Verification data

3.5.1 Synop stations

In the chapter 1 is has been mentioned that, due to the small spatial extent of most hail events related to summertime thunderstorms, the 19 synop stations in the Netherlands will only report a minor fraction of the total number of hail events. During the summer of 1999, an average observer at a synop station has reported hail only once, although there have been about 60 days with thunderstorms which produced hail somewhere in the Netherlands. Therefore, the verification data from the manned synop station has been completed with hail observations and reports from other sources.

3.5.2 Precipitation stations

KNMI maintains a dense network of about 325 volunteers who report the amount of accumulated precipitation on a daily basis. The density of these "precipitation" stations is about one station every 100 km^2, while that of the synop stations is roughly one station every 1800 km^2. In Figure 3.1 a map of the Netherlands is shown where all synop stations (◯) and all precipitation stations (•) are marked. The precipitation stations are distributed fairly homogeneously across the Netherlands. Apart from the precipitation sums, the volunteers of the network are encouraged to report hail events as well. The hail events, like the accumulated precipitation, are reported daily at 08 UTC, and the reports refer to events that have occurred within the preceding twenty-four hours. A disadvantage of using these hail reports is that a report of "no hail" is not very reliable, because the precipitation observers are not obliged to report hail and they do not look for hail day-and-night. In the summer of 1999, all precipitation stations reported in total about 350 days with hail events. On average this is about one event per precipitation station per season, which is comparable to the number of events per season reported by a synop station.

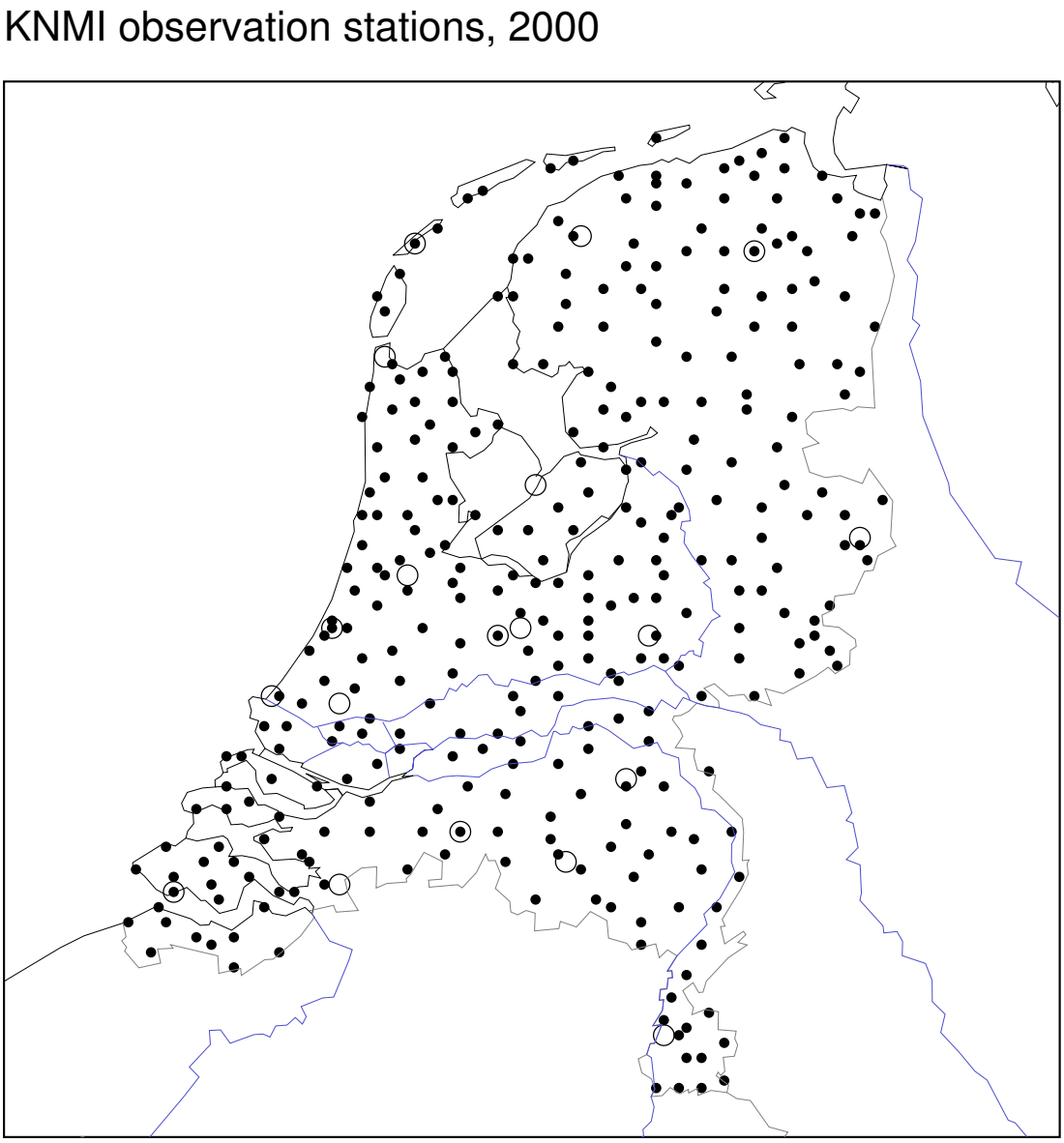

Figure 3.1: Map of the Netherlands showing the observation stations used in this study. In this map both the synop stations and the precipitation stations are marked. In the Netherlands, there are in total 19 manned synop stations and about 325 precipitation stations. The manned synop stations are indicated by a ◯ and the precipitation stations by a •.

3.5.3 Insurance companies

In addition to the observations of hail, reports of hail damage as received by three major agricultural insurance companies in the Netherlands have been used for the verification. The insurance companies which supplied the hail damage reports are:

- Hagelunie Agrarische Verzekeringen, in Leidschendam
- Onderlinge Fruittelers Hagelverzekeringsmaatschappij (OFH), in Den Haag
- AgriVer verzekeringen, in Hasselt

In the hail damage reports no time is specified. Only a date using local time for day transitions is given. The reports refer, therefore, to a period of twenty-four hours ending at 22 UTC of the day specified by the date of the report and starting at 22 UTC of the preceding day. The way in which the locations of the hail damage are specified depends on the origin of the damage reports: it can be given as a district number, i.e., the unique "gemeentecode" assigned by the "Centraal Bureau voor de Statistiek" (CBS) to each district, as a name of a city or a village, or as a part of the zip-code.

Even with a homogeneous distribution of hail events, the reports of hail damage to the insurance companies will not be distributed homogeneously across the Netherlands due to differences in the sensitivity of the local land use to hail events. Figure 3.2 shows a map of the Netherlands where the fraction of the land use, sensitive to hail events, is indicated for grid boxes having an area of 2.4 by 2.4 km^2. This map has been constructed using the LGN2 database of land use in the Netherlands obtained from satellite observations between 1993 and 1995 by determining for each grid box the fraction of "hail sensitive" land use classes, like crops, orchards, greenhouses, and bulbs. The presence of distinct areas where hail damage reports can be expected from and on the other hand the presence of areas where no damage reports are expected is obvious from this figure. This notion will be used in the verification of the hail detection methods when hail damage reports from insurance companies are used exclusively. During the summer of 1999, in total about 2100 cases of hail damage have been reported to these three insurance companies.

3.5.4 Other sources

Finally, reports of hail by weather amateurs or newspapers have been collected. The reports by the weather amateurs have been extracted from mailing lists dedicated to special weather events ("Bijzonder weer") and from the "Weerspiegel" magazine issued monthly by the "Nederlandse Vereniging voor Weeramateurs". Although the number of reports collected in this way was not very significant, i.e., about 35, these reports sometimes contain information on the size of the hail stones and the exact time and location of the events.

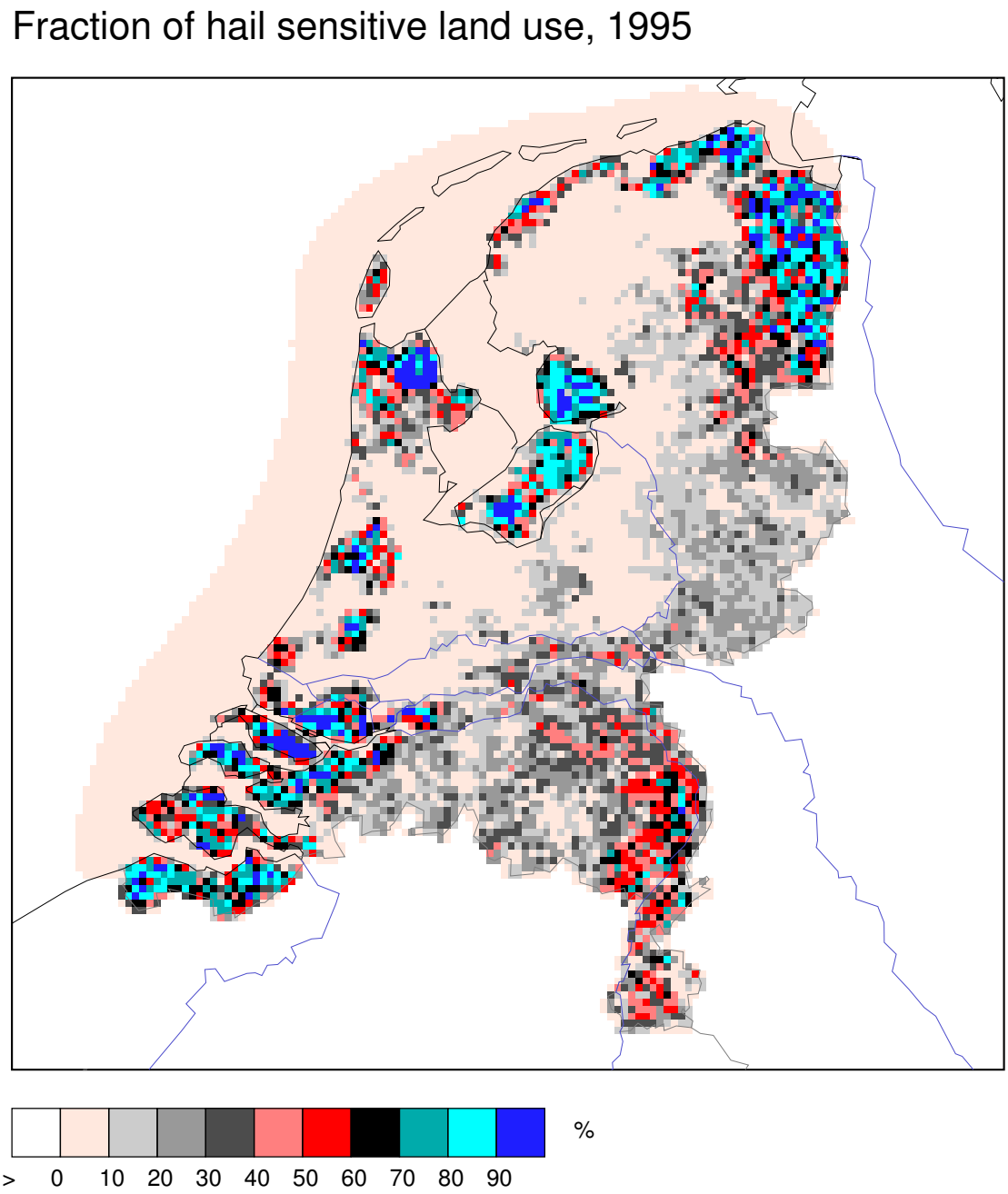

Figure 3.2: Map of the Netherlands showing the hail sensitive land use for each location. This land use has been calculated from the LGN2 database using satellite data from 1993 to 1995. The hail sensitivity for each pixel is determined from the fraction of "hail sensitive" land use classes, like agricultural crops, greenhouses, orchards, and bulbs.

Chapter 4

Methods of comparison

A systematic comparison of the output of the eight selected methods for the detection of hail to on-ground observations of hail has been conducted using radar volume data of the selected days in the summer of 1999. The calculation of the hail detection methods from the radar and temperature data is described first. Then, the grouping of the hail detection data and verification data into day bins and districts is discussed. Finally, the verification procedure and scores based on contingency tables is described, and a simple model to account for missing verification data is presented.

4.1 Calculation of detection methods

4.1.1 Temperature data

The method of Waldvogel and the severe hail index use information on the vertical temperature profile and the method of Auer uses cloud-top temperatures from calibrated infrared data of Meteosat. In this study, the information on the temperature profile has been obtained from the most recent HiRLAM analysis.

In order to be able to combine the model data with the radar data, the three-dimensional temperature information given on the grid and the 31 model levels has to be converted to a geometrical location and height. Using the pressure, the temperature, the mixing ratio at all model levels, the geopotential height of each model level at a certain gridpoint is calculated assuming hydrostatic equilibrium. Subsequently, the temperature at a certain height or the height of the freezing level or the $-20°$C level at a certain gridpoint is computed via linear interpolation between enclosing model levels. This procedure is repeated for all model gridpoints within a selected square enclosing the standard area covered by the radar products of KNMI (*vide infra*). In Figure 4.1, an example map of the height of the freezing level as obtained from an analysis of HiRLAM is shown for the selected area. To obtain the value of a temperature or height deduced from HiRLAM at a specific, non-gridpoint location, a bilinear interpolation between

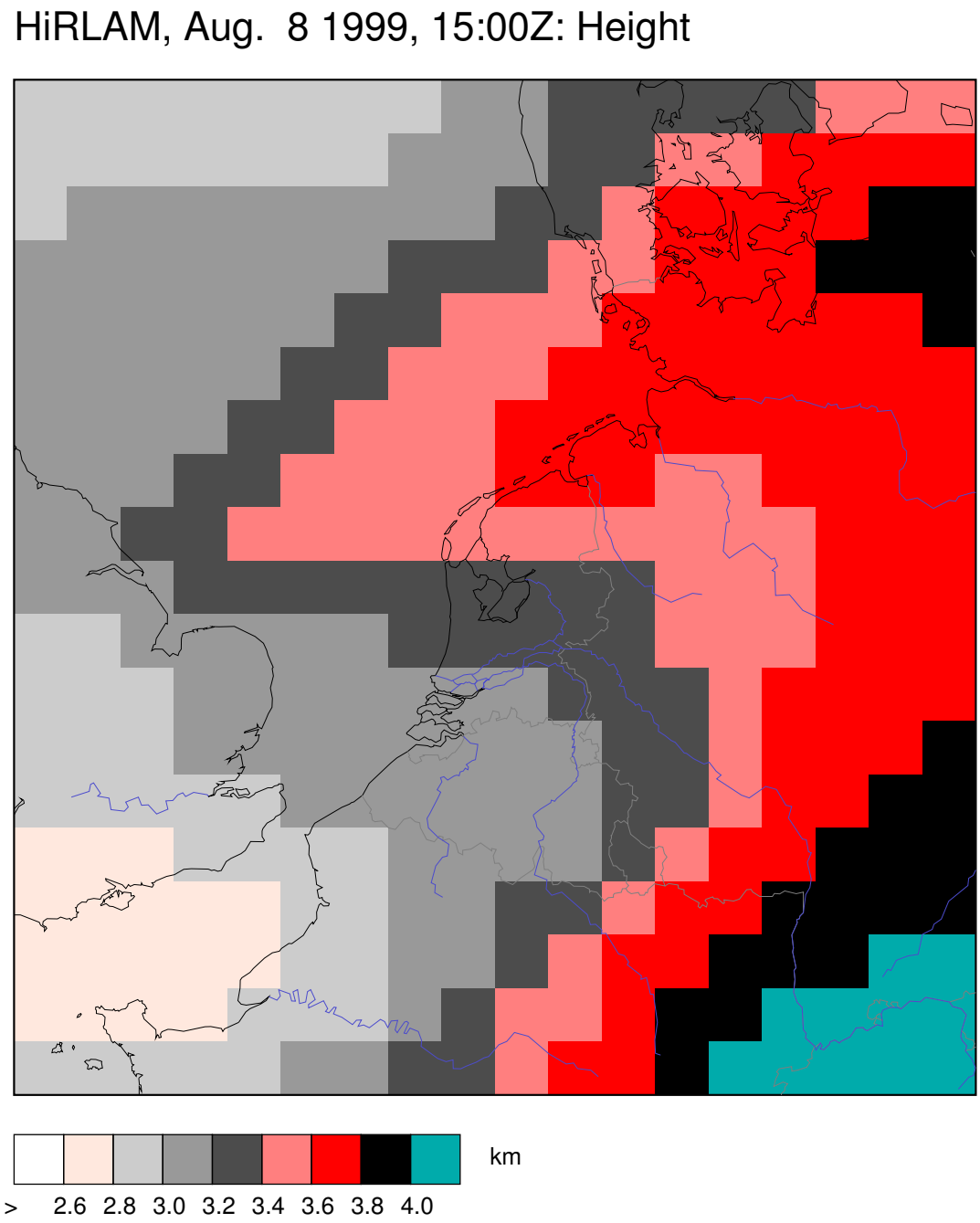

Figure 4.1: The height of the freezing level as obtained from the analysis of HiRLAM at 15 UTC on August 8 of 1999. Only a selected square of HiRLAM gridpoints is shown, i.e., 18 times 19 points spaced by 55 km, which is enclosing the standard area covered by the radar products of KNMI.

Table 4.1: Navigation details of the standard radar image of KNMI, and the resulting longitude and latitude of the four corners of the image.

Parameter	Value
Projection	Polar Stereographic North
Reference (lon,lat)	0.00E, 60.00N
Grid offset (x,y)	0.0, 1490.9
Pixelsize at reference (x,y)	2.50 km, 2.50 km
Earth radius (equator,polar)	6378.388 km, 6356.912 km
Number of lines	256
Number of columns	256
Corner	Location (lon,lat)
northwest	0.000E, 55.296N
northeast	9.743E, 54.818N
southeast	8.337E, 49.373N
southwest	0.000E, 49.769N

the four gridpoints enclosing the desired location is performed.

For the method of Auer, cloud-top temperatures have to be extracted from the calibrated infrared images of Meteosat. When a pseudoCAPPI pixel has a value exceeding the lowest reflectivity threshold of 36 dBZ (see Figure 2.3), the longitude and latitude of that radar pixel are calculated. Subsequently, the nearest Meteosat pixel is determined using the equations describing the "geostationary projection". Finally, the obtained cloud-top temperature is used to calculate the hail threshold according to Auer (1994).

4.1.2 Radar data

All hail detection methods were calculated from the volume data of the radar in De Bilt complemented with Meteosat infrared data and numerical weather prediction model data. The three-dimensional volume datasets contain the radar reflectivity as a function of range, azimuth, and elevation. From the volume data, polar datasets, i.e., data as a function of on-ground distance to the radar and azimuth, of (intermediate) quantities, such as reflectivity at constant altitude, maximum reflectivity in vertical, echotop heights, or vertically integrated (modified) reflectivity are calculated. The (intermediate) quantities are combined to obtain the eight different hail detection algorithms, as given in Table 2.2, in a polar projection with a resolution of 1 km and

1 degree for distance and azimuth, respectively.

Subsequently, the hail detection methods are reprojected from the radar-centered polar projection to a polar stereographic projection with the northpole in the center. The equations defining the polar stereographic projected grid and those for the computation of the coordinates of a target with given distance and azimuth, in both cases for an ellipsoid-earth model, are taken from Wessels (1990). The spatial resolution of the polar radar data in the tangential direction is changing with distance: at short distances the resolution is relatively high (0.2 km at 10 km), while at large distances the resolution becomes rather poor (5 km at 300 km). When this polar data is projected on a polar stereographic grid with a pixelsize of about 2.4 km, the data will be undersampled at short distances and oversampled at long distances. To make optimal use of the polar data, a forward projection is performed first. In this forward projection, all polar datapoints are projected to the new grid, and the data is averaged when more polar datapoints are available at a certain new gridpoint. Subsequently, a backward projection is performed where the value of the nearest polar datapoint is assigned to all new gridpoints that not have been assigned a value by the forward projection.

Finally, some cosmetic operations are applied to the obtained radar image. Single pixels, i.e., pixels solely surrounded by pixels having values below a certain threshold, are removed. The obtained radar image is double-checked using the standard radar display of reflectivity at constant (low) altitude (CAPPI) from which signal due to clutter has been removed by applying the corresponding dynamical cluttermap. Each pixel, in the obtained radar image, with a significant value should be confirmed by a nearby pixel in the corresponding CAPPI image having a value above a certain threshold (typically 15 dBZ), otherwise it is set to "no data". Both these cosmetic operations result in a considerable reduction of "noisy pixels" in the radar images.

In Figure 4.2 an example of a hail detection image which is obtained using the method of Waldvogel and the above mentioned procedures is shown. For the method of Waldvogel, the difference between the maximum height of the 45 dBZ radar echoes and the height of the freezing level (see Figure 4.1) is displayed. The displayed image has been enlarged from 256×256 pixels to 140×140 pixels, and it is shifted to just enclose the Netherlands. A thunderstorm in the south of the Netherlands producing severe hail ($\geq$3.5 cm) at that moment is clearly visible, and it has a maximum height difference of 4 km.

In Table 4.2 the relevant quantities and units are listed for all implemented hail detection methods. The choice for the use of a certain quantity is obvious for most detection methods or has been explained above. For the method of Auer, the difference between the observed radar reflectivity and the reflectivity threshold, which is determined from the observed cloud-top temperature, is used. The use of a value of 0 dBZ for this reflectivity difference is equivalent to the use of the same reflectivity threshold as proposed by Auer (1994). For the detection method based on the Severe Hail Index, the value of this quantity is used directly, and thus the conversion of the SHI to the probability of severe hail is not implemented.

Figure 4.2: A hail detection image using the method of Waldvogel for August 8 of 1999 at 16:34 UTC. The difference between the maximum height of the 45 dBZ radar echoes and the height of the freezing level as obtained from the HiRLAM model (see Figure 4.1) is depicted. The image is enlarged to 140×140 pixels, and it is just enclosing the Netherlands. The cross marks the position of the radar (De Bilt).

Table 4.2: An overview of the quantities and units that are used for each of the implemented hail detection algorithms.

Method	Quantity	Unit
CAPPI	Reflectivity	dBZ
maxPPI	(Maximum) reflectivity	dBZ
Auer	Reflectivity (difference)	dBZ
Waldvogel	Height (difference)	km
Echotop	Height	km
SHI	Kinetic energy flux	J/ms
VIL	Integrated water content	kg/m^2
VIL-dens	Water content	g/m^3

4.2 Daily bins and district groups

It has been detailed in section 3.5 that most of the hail observations and reports needed for verification of the hail detection methods are available with a limited temporal and spatial accuracy. In order to compare the hail detection radar images, which have a high resolution in both time and space, to the verification data, the temporal and spatial resolution of the radar images has to be degraded.

The precipitation stations report their daily hail observations at 08 UTC of the next morning, while the hail damage reports from the insurance companies are valid from 22 UTC of the previous day until 22 UTC of the specified day. For use in the verification of the detection methods, these hail observations and reports have to be combined. Assuming that hardly any of the hail-producing thunderstorms are occurring and/or observed between 22 UTC and 08 UTC, the hail observations by the precipitation stations and the hail damage reports from the insurance companies can be said to be valid between 00 UTC and 24 UTC. Note that in the case of the observations by the precipitation stations the valid day is the day before the specified one. To make the timing of all other verification data with a more specific timetag, for instance the data of the synop stations, equivalent to that of the data mentioned above, it is reduced to a date only.

The 96 radar images obtained for a specific hail detection method on every day are reduced to one daily image for that method. For this, a binned radar image is constructed that contains for each pixel the maximum value that has occurred at that location and at that day. In Figure 4.3 an example of a binned hail detection image constructed using all 96 hail detection images of the Waldvogel method of one day is shown. This figure should be compared with Figure 4.2

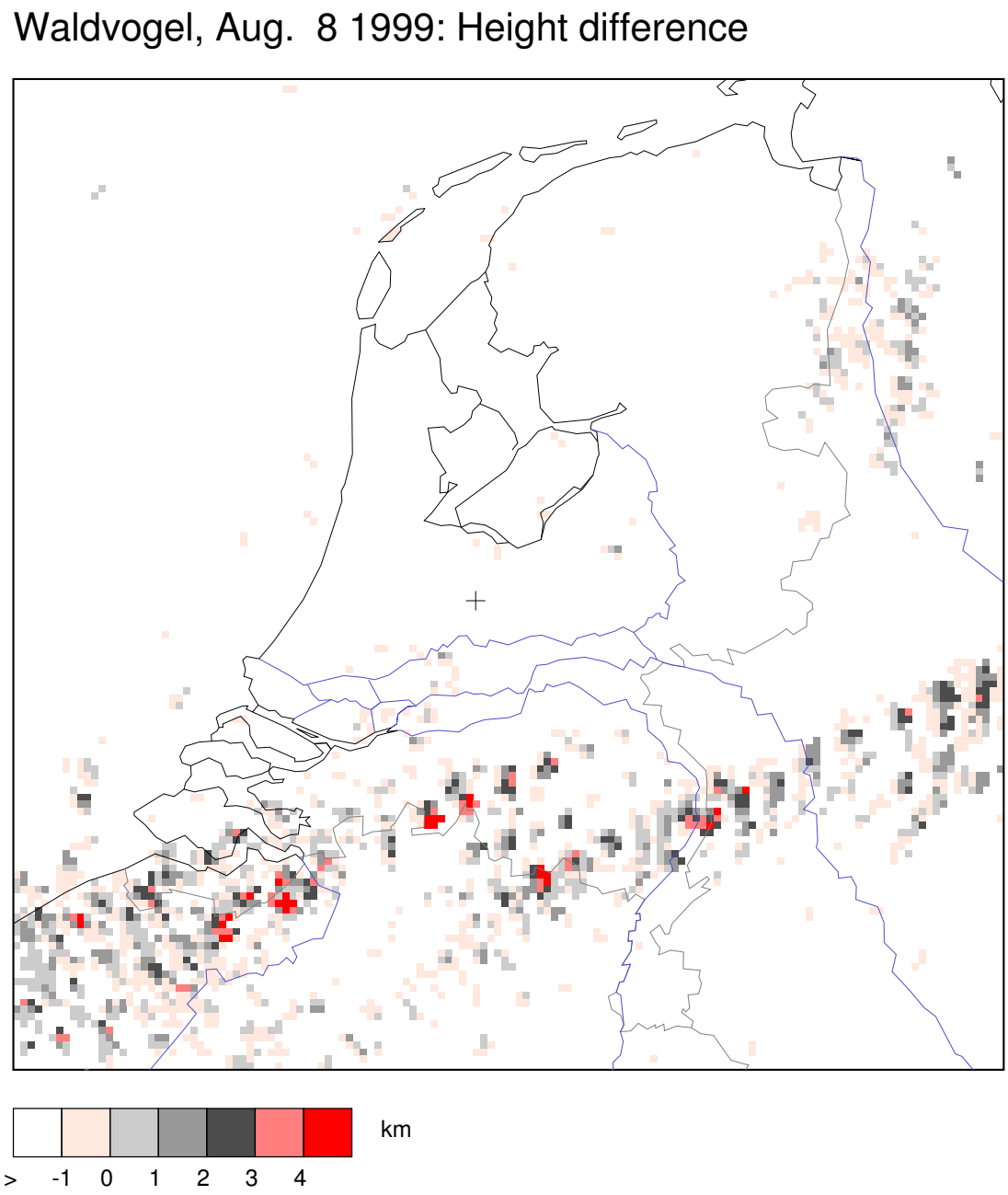

Figure 4.3: A binned hail detection image using all 96 radar images obtained with the method of Waldvogel on August 8 of 1999. This figure should be compared with Figure 4.2 of a single radar image of the method of Waldvogel. The tracks of several hail-producing thunderstorms observed at 15 minutes intervals can be identified as “jumping cells”. This figure is enlarged similarly to Figure 4.2.

which is actually part of the binned image. The tracks of several hail-producing thunderstorms which are observed only at 15 minutes intervals can be identified as "jumping cells". The heading and velocity of these cells can easily be inferred from this binned image. It is obvious from this figure that the 15 minutes interval between two consecutive radar scans is too long to get smooth-looking cell tracks. A problem that will not be touched in this study, but that could be solved by tracking of the cells and using the obtained motion vectors for interpolation between consecutive scans.

The locations of the hail damage reports from the insurance companies are given as district, as names of cities and villages, or as parts of zip-codes. The districts in the Netherlands have turned out to be convenient units for binning all verification data and radar data into the same "location units". In 1999, the Netherlands were divided into 538 districts having an average area of about 60 km^2 (537 districts in 2000). The "Centraal Bureau voor de Statistiek" (CBS) has assigned a unique number to each district, the so-called "gemeentecode" (CBS number), and provides lists of all districts and their numbers. The locations of all hail damage reports from insurance companies and all hail observations at stations have been characterized by the CBS number of the district in which the report or station is localized.

Using a (relatively) high-resolution map of the Netherlands containing on a 1×1 km grid the CBS district numbers, a district map has been created on the standard radar grid (see Table 4.1), where each radar pixel has been assigned to a CBS district number. This district map has been used as a look-up table for allocation of the district number for a certain pixel within a hail detection image.

Via the daily binning and the grouping into CBS district numbers, all verification data and radar data have been converted into the same units of time and space. Because all hail events are now characterized in a standard way, i.e., by giving a date and a CBS district number, a quantitative comparison between the different hail detection methods and the verification data has become feasible.

4.3 Verification scores

In the verification process, i.e., the comparison between the outcomes of the different hail detection methods and the verification data, the hail events will be classified using a 2-by-2 contingency table. Hail detected by a radar-based method which is confirmed by the verification data will be classified as a hit (H), hail detected by a radar-based method which is not confirmed by verification data as a false alarm (F), hail observations or reports in the verification data that are not detected by the radar-based method as a miss (M), and no event at all as a non-event (N). These four classes can be shown schematically in the 2-by-2 contingency table:

	Hail	No hail
Radar detection	H	F
No radar detection	M	N

In the case of (very) rare events, like summer hail, the number of "non-events" N will be the largest by far. If, in this case, one uses verification scores that include the number of non-events, like the fraction correct, the outcome of the score will be dominated by this large number. The dominance of the none events in the case of rare events can be circumvented by the use of verification scores that do not include N. The Probability Of Detection (POD), the False Alarm Ratio (FAR), the Critical Success Index (CSI), and the bias are defined as:

$$\text{POD} \equiv \frac{H}{H+M} \tag{4.1}$$

$$\text{FAR} \equiv \frac{F}{H+F} \tag{4.2}$$

$$\text{CSI} \equiv \frac{H}{H+M+F} = \left[\frac{1}{\text{POD}} + \frac{1}{1-\text{FAR}} - 1\right]^{-1} \tag{4.3}$$

$$\text{bias} \equiv \frac{H+F}{H+M} = \frac{\text{POD}}{1-\text{FAR}} \tag{4.4}$$

The probability of detection and the false alarm ratio always have to be used together to characterize the result of a verification, where the method with a high POD and a low FAR is preferred, while the critical success index characterizes the verification result in a single number, where the method with the highest CSI is preferred. The bias is the ratio between the number of detections or forecasts and the number of actual occurrences. A method that detects or predicts events with a too high (low) frequency has a bias greater (smaller) than one, and a method with a bias of one has no bias. The POD, FAR, and/or CSI are often used to characterize the performance of hail detection methods. A more elaborate discussion on the behavior of these and other verification scores can be found elsewhere (Kok, 2000; Doswell et al., 1990). The CSI has for instance the disadvantage that it rewards detection or prediction of an event with too high a frequency, i.e., overdetection or overforecasting, respectively. The bias should, therefore, be used to identify this overdetection or overforecasting by a method.

As detailed in section 3.5 much effort has been put into the collection of the hail observations and damage reports used for the verification. Due to the small spatial extent of most summertime hail events, however, some of these events will remain unnoticed. The effect of the missing verification data on the classification of the hail events and on the verification scores can be investigated by making the (simple) assumption that only a fraction η of the occurring hail events is actually present in the verification data. The four classes of a modified contingency table (H', M', F', and N') can be expressed in terms of the original classes and this fraction η as follows:

	Hail	No hail
Radar detection	$\eta \cdot H$	$F + (1 - \eta) \cdot H$
No radar detection	$\eta \cdot M$	$N + (1 - \eta) \cdot M$

Effectively, a fraction $(1 - \eta)$ of the events in the left column of the contingency table has been transfered to the right column of the table. Using this modified contingency table, the apparent Probability Of Detection (POD′), False Alarm Rate (FAR′), Critical Success Index (CSI′), and bias′ can be expressed in terms of the fraction η and the true POD, FAR, CSI, and bias:

$$\text{POD}' = \frac{\eta \cdot H}{\eta \cdot H + \eta \cdot M} = \text{POD} \tag{4.5}$$

$$\text{FAR}' = \frac{F + (1 - \eta) \cdot H}{H + F} = (1 - \eta) + \eta \cdot \text{FAR} \tag{4.6}$$

$$\frac{1}{\text{CSI}'} = \frac{H + F + \eta \cdot M}{\eta \cdot H} = \frac{1}{\text{CSI}} + \frac{1 - \eta}{\eta \cdot (1 - \text{FAR})} \tag{4.7}$$

$$\text{bias}' = \frac{H + F}{\eta \cdot H + \eta \cdot M} = \frac{1}{\eta} \cdot \text{bias} \tag{4.8}$$

In Figure 4.4 an example of the dependence of the apparent POD′, FAR′, CSI′, and bias′ on the fraction η is shown using a true POD of 0.70 and a true FAR of 0.30. It is evident from this figure that, under the assumptions made in the η-model, incompleteness of the verification data ($\eta < 1$) results in an increase of the apparent FAR′ and a concomitant decrease of the apparent CSI′. For $\eta < 1$, the apparent bias′ is always larger than the true bias. In addition, the results from the η-model imply that the POD can be determined more accurately than the FAR, CSI, or bias because it is independent of η. The apparent POD′ can, however, be reduced at longer distances due to undersampling of hail cells and due to attenuation of the radar radiation. As the maximum distance to the radar used in this study is only about 150 km, these effects on the POD′ are expected to be less significant than the effect of the missing verification data on the FAR′, and therefore they are not explicitly taken into account in the 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).

4.4 Automated verification

A program is used to calculate the number of (apparent) hits, false alarms, misses, and non-events from a daily-bin of a hail detection method (see Figure 4.3) and a list with the verification data. For this, an array containing the "verification status" of all districts is created and all

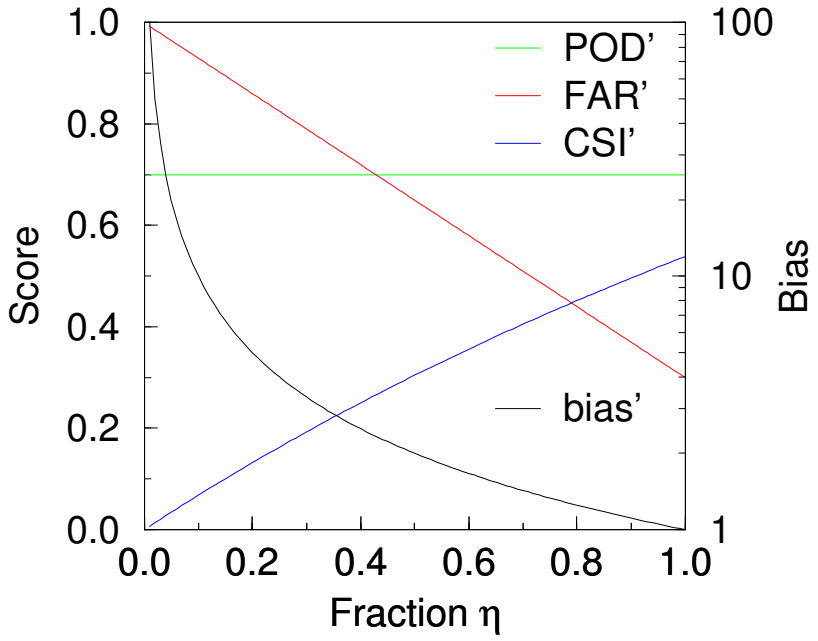

Figure 4.4: An example of the apparent POD′, FAR′, and CSI′ (left y-axis) and the apparent bias′ (right y-axis) as a function of the fraction η of reported hail events. This example has been calculated using a true POD of 0.70 and a true FAR of 0.30 (implying a true CSI of 0.54 and no true bias).

entries are marked as an N. Using the date extracted from the binned radar image and the list with the verification data, the entries of the array corresponding to the districts where hail has been observed or hail damage has been reported on that day are set to M. Subsequently, a loop is performed over all pixels in the binned image that have a value higher than a given warning threshold. Using the district map as mentioned in section 4.2, the districts within a given maximum allowable distance from the pixel, dubbed the "positioning tolerance", are checked. Depending on the verification status of these districts, the following action is taken:

- If one or more districts are present with a verification status of M, the status of the one which is nearest to the radar pixel is changed to H.
- If no district with a verification status of H, M, or F is present, the status of the district which is nearest to the pixel is changed to F.

Finally, there is an option to remove a selected set of districts from the resulting verification array. When, for instance, only hail reports from KNMI stations are considered, the districts where no observing stations are present would spoil the verification results. Those stations can be removed using this option.

A second program is used to bin the verification results obtained for each day and for each detection method and to calculate the final verification scores, like POD, FAR, and CSI, and the bias. The verification program and the binning program are run several times using a script while the warning threshold or the positioning tolerance is varied.

Chapter 5

Results of comparison

In this chapter, the results of the comparison and verification of the eight different hail detection methods are described. For this, the radar data and verification data of 15 selected days with thunderstorms during the summer of 1999 are used. First of all, the dependence of the performance on the positioning error is investigated for a few selected methods. Then, the performance and the warning characteristics as a function of the warning threshold are examined for all methods. Finally, the fraction η of reported hail events is changed systematically by a certain selection of districts that are taken into account in the verification scores. The resulting changes in the verification scores of the method of Waldvogel are evaluated, and an approximation of the true performance (CSI) of the method of Waldvogel is made.

5.1 Positioning tolerance

It has been mentioned previously that the verification program allows for a certain spatial mismatch in the assignment of a particular (group of) radar pixels having values above the warning threshold to a district where hail has been observed or reported. The maximum allowed distance between a radar pixel above threshold and the border of a district with the on-ground confirmation is defined as the "positioning tolerance". In Figure 5.1, the performances, i.e., the apparent CSIs, resulting for the method of Waldvogel, the VIL method, and the CAPPI method are shown as a function of the positioning tolerance. The warning thresholds used for the method of Waldvogel (1.75 km), the VIL method (15 kg/m^2), and the CAPPI method (49 dBZ) have been set to optimum performance, indicated by the highest CSI$'$, at a positioning tolerance of 12.5 km (*vide infra*).

It is evident from Figure 5.1 that the performances of all methods are rather poor at zero tolerance. The requirement of exact spatial coincidence, within a 2.4-by-2.4 km^2 radar pixel, between detection and verification puts a too high demand on the spatial accuracy of the detection methods and on the completeness of the verification data. In comparable studies a certain spatial mismatch has been used as well (Kessinger et al., 1995; Witt et al., 1998).

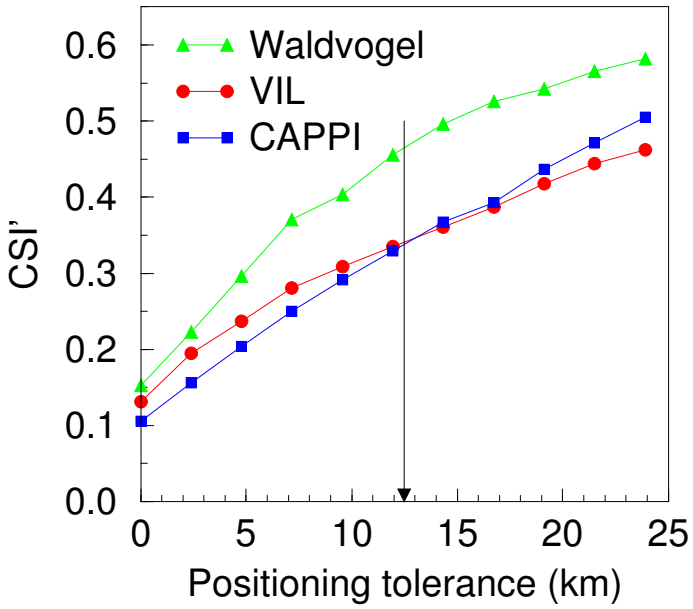

Figure 5.1: The dependence of the apparent Critical Success Index (CSI′) of three different hail detection methods on the allowed spatial mismatch between a radar pixel and an on-ground hail observation. The warning threshold of Waldvogel is set at 1.75 km, that of VIL at 15 kg/m^2, and that of CAPPI at 49 dBZ. The vertical arrow marks the positioning tolerance of 12.5 km that is used throughout this study.

When the positioning tolerance is increased, the observed POD′ for the method of Waldvogel, the VIL method, and the CAPPI method steadily increases and their observed FAR′ and bias′ steadily decrease (*not shown*). It is evident from Figure 5.1 that the performance of all detection methods as indicated by their CSI′ increases substantially when increasing the positioning tolerance. The decrease of the bias′ implies that the increase of the CSI′ is not due to (extra) rewarding of overforecasting (Kok, 2000). The five other detection methods show similar behavior. From the methods in Figure 5.1, the one 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 mean sea level), 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 high-altitude radar echoes is expected to be larger. This spreading effect is reflected in a stronger dependence of the performance of the method of Waldvogel on the positioning error.

The average area of districts in the Netherlands, which reflects the spatial accuracy of the available verification data, effectively sets a lower limit to the attainable position tolerance of about 5 km. Taking into account a reasonable region of influence for a summertime thunder-

storm and to be consistent with other studies of this kind, a positioning tolerance of 12.5 km has been applied throughout this study. Kessinger and Witt have used positioning tolerances ranging from 5 km up to 30 km in their verification of hail detection methods (Kessinger et al., 1995; Witt et al., 1998). The positioning tolerance applied in this study is marked in Figure 5.1 with a vertical arrow.

5.2 Dependence on threshold

By running the verification program repeatedly, the scoring parameters, i.e., the apparent POD′, FAR′, and CSI′, and the bias′ of the eight selected hail detection methods have been determined as a function of their warning thresholds. Apart from the biases, the obtained results are shown in Figures 5.2 and 5.3. The hail warning thresholds are set on different quantities depending on the hail detection method in question. An overview of these quantities and their units has been given in Table 4.2, and for the non-trivial ones it is reviewed here. 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 the use of the same temperature-dependent reflectivity threshold as proposed by Auer (1994). 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. For the method based on the severe hail index (SHI), the warning threshold is directly set on the SHI value, and thus the conversion of the SHI to probability of severe hail is not implemented.

Although there are large differences between the curves in Figures 5.2 and 5.3, the general trends in the scoring parameters as a function of the warning thresholds are rather similar for all hail detection 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. In addition, it is observed (*not shown*), for all detection methods, that the bias of a method decreases dramatically, typically by a factor of 10-100, when the warning threshold is increased. This decrease of the bias directly reflects the reduction of the warning frequency of a detection method when the warning threshold is raised. At high warning thresholds, the deduced value for the FAR′ of a method can become unreliable due to the low number of detected hail events.

In Table 5.1, two examples of contingency tables that have been obtained for the method of Waldvogel and of Auer are shown in the left and in the right table, respectively. It is evident from these tables, that the right colums (“No hail”) have the largest number of events by far, and thus summer hail is even on days with thunderstorms a rare event. The method of Auer has hardly any misses M and a large amount of false alarms F, and this results in a high POD′ and

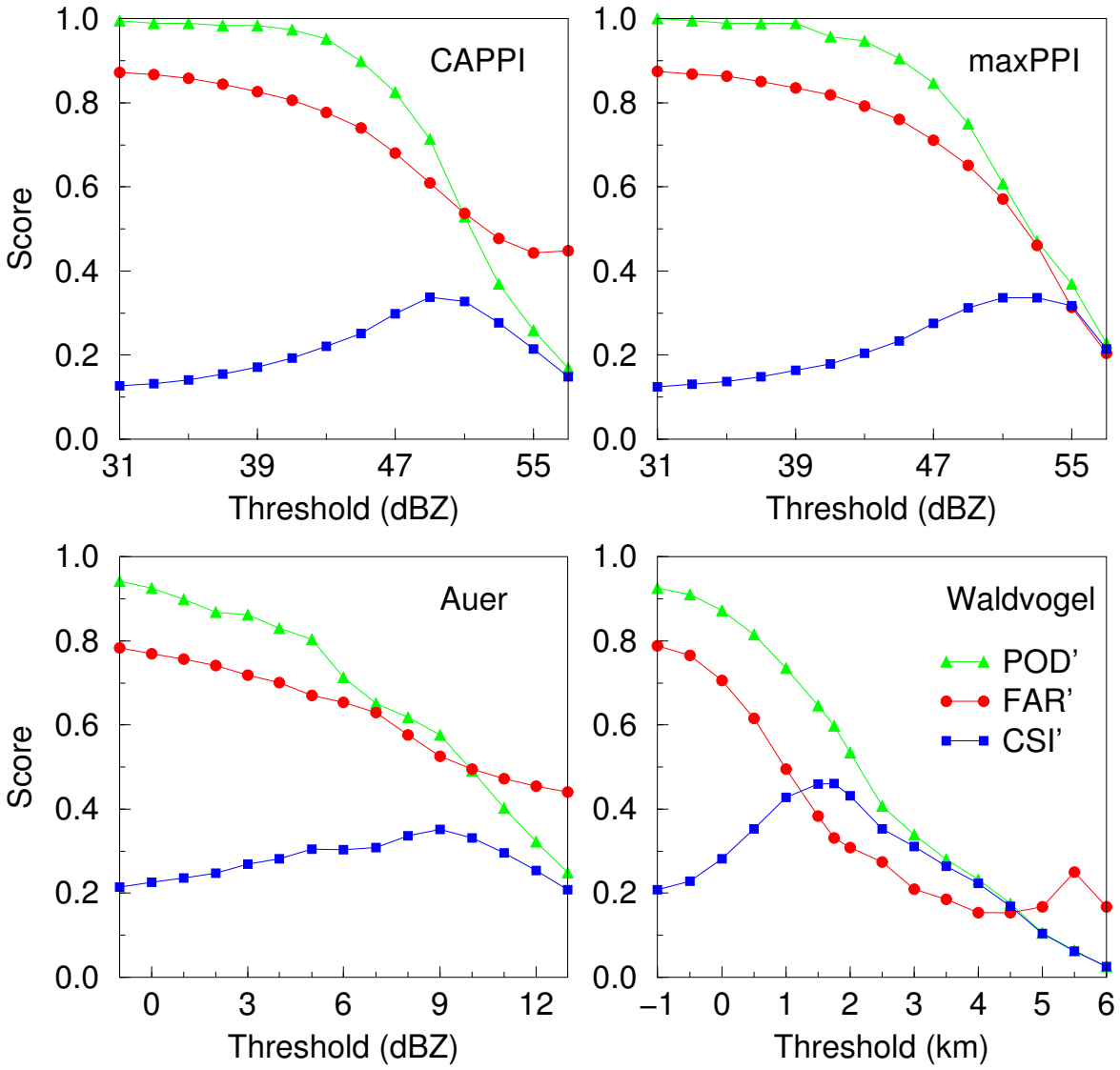

Figure 5.2: The scoring parameters (POD′, FAR′, and CSI′) for the first four hail detection methods as a function of the warning threshold. The scoring parameters are deduced from the comparison of the methods with the verification data. The positioning tolerance is set at 12.5 km.

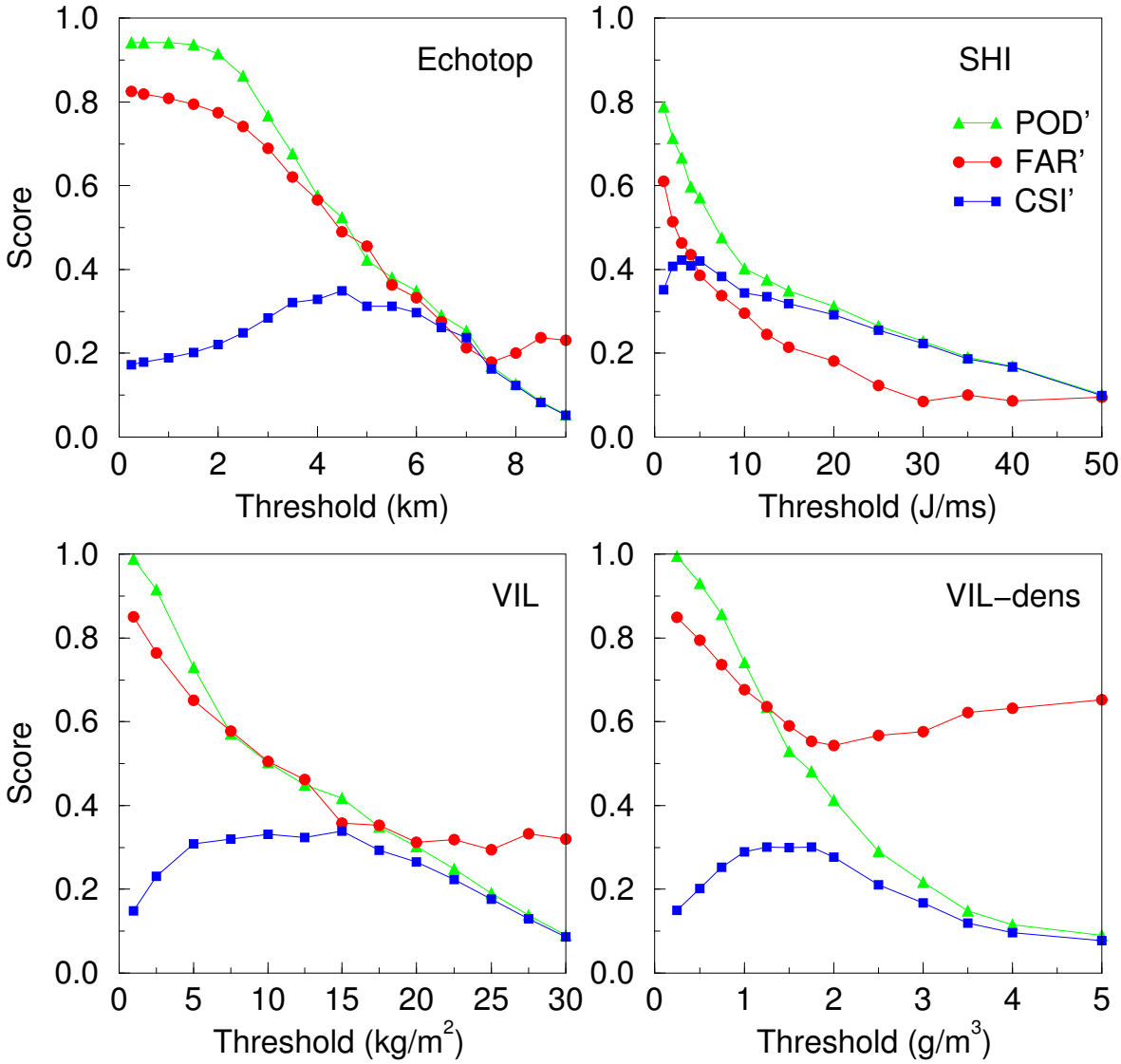

Figure 5.3: The scoring parameters (POD′, FAR′, and CSI′) for the other four hail detection methods as a function of the warning threshold. Everything else is similar to Figure 5.2

Table 5.1: Two examples of contingency tables that have been obtained from the verification runs. The left table contains the result for the method of Waldvogel using the optimum warning threshold of 1.75 km, and the right table contains the result for the method of Auer using a warning threshold of 0 dBZ, i.e., an exact reproduction of the method.

Method of Waldvogel

	Hail	No hail
Radar detection	113	56
No radar detection	76	7750

Method of Auer

	Hail	No hail
Radar detection	175	587
No radar detection	14	7219

a high FAR′ (see Figure 5.2). More details of these two contingency tables will be discussed below.

The lowest FAR′ is observed for the SHI method using a warning threshold of 30 J/ms, and it is about 0.1 (see Figure 5.3). Equation 4.6, which is the equation for the apparent FAR′ as a function of η, states that the FAR′ cannot go below $(1 - \eta)$ even when the true FAR of a detection method is zero. Therefore, a lower limit for the fraction of reported hail events η, which is a property of the verification dataset only, can be determined from the lowest, apparent FAR′. Within the accuracy of the η-model and the positioning tolerance used, it is found that at least a fraction $\eta = 0.9$ of the hail events that have occurred are contained by the verification dataset 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. From the corresponding contingency table (see Table 5.1), it can be deduced that the method of Waldvogel has a bias′ of 0.90 at this threshold. So, at this warning threshold the method of Waldvogel has the best performance and hardly any bias. Using equations 4.6, 4.7, and 4.8 and the lower limit for η of 0.9, a lower limit for the bias of 0.8 and an upper limit for the CSI of 0.49 are found for Waldvogel's method at this threshold. On our data, the method of Auer performs only slightly better than the straightforward CAPPI method. The optimum performance is shifted away substantially from a warning threshold of 0 dBZ, the threshold at which the method of Auer is reproduced exactly. From the contingency tabel of Auer's method (see Table 5.1), a bias of 4.0 is determined for the 0 dBZ warning threshold indicating that the method is overdetecting considerably. At a warning threshold of 9 dBZ where the optimum performance is observed, the observed bias′ of the method of Auer is close to one (1.2).

The optimum performances of the CAPPI, VIL, VIL-density, and SHI methods are found at (much) lower warning thresholds than those reported in literature (Mason, 1971; Edwards and Thompson, 1998; Amburn and Wolf, 1997; Witt et al., 1998). In addition, the optimum perfor-

mances are seen to correspond fairly well with the absence of a bias for these methods. The observed discrepancies in the optimum warning thresholds 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. On our dataset, the VIL-density method of Amburn and Wolf (1997) does not show any improvement over the original VIL method which is in agreement with the comments made by Edwards and Thompson (1998). The maximum reflectivity method (maxPPI), employed by the Rainbow software, shows no improvement with respect to the straightforward CAPPI method. From a comparison of the scoring curves of Waldvogel's method and those of the echotop method, it is evident that the addition of the temperature information, i.e., the height of the freezing level, improves the performance considerably.

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 one of the largest for the method of Waldvogel as well. A hail detection method can be used to produce a Probability Of Hail (POH) defined as:

$$\text{POH} \equiv \frac{H}{H+F} \;=\; 1 - \text{FAR} \tag{5.1}$$

So, the observation of a radar pixel value above a certain warning threshold indicates a probability of hail which is equal to $1 -$ FAR at that threshold. The large difference between the minimum and maximum FAR$'$ observed for the method of Waldvogel enables the definition of several different thresholds with distinct warning properties, which are the probability of hail or FAR$'$ and the POD$'$.

5.3 Varying the η-fraction

Although the fraction of reported hail events η is hard to be determined, it can be changed systematically by selecting districts with a certain probability of hail damage and by considering, during the verification, the hail damage reports from the insurance companies only. In Figure 3.2 the hail-sensitive land use, like crops, orchards, greenhouses, and bulbs, for each location obtained using a database of land use in the Netherlands obtained from satellite observations between 1993 and 1995 is shown. By overlaying this hail-sensitivity map with the district map as introduced in section 4.2, the fraction of the area with hail-sensitive land use can be calculated for each district. Subsequently, only the districts having a certain minimum fraction of hail-sensitive land use are taken into account when the comparison of the detection methods against the hail damage reports in the verification data is made.

In Figure 5.4 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 land use fraction of the selected districts. It can be seen from this figure that the POD$'$ is more or less

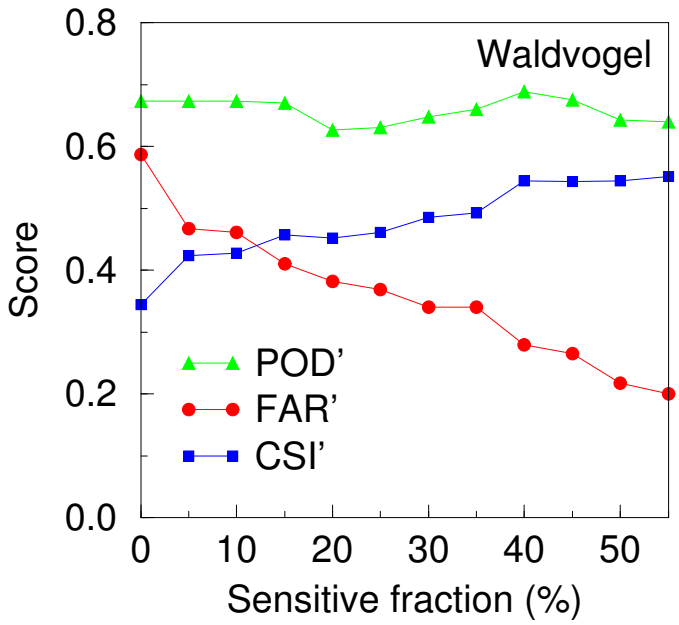

Figure 5.4: The scoring parameters of the method of Waldvogel as obtained on the verification data of hail damage reports of the insurance companies only using a warning threshold of 1.75 km . The minimum, required coverage fraction of hail-sensitive land use of a district in order to be taken into account in the scoring parameters is varied.

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 events η is increased. The apparent bias$'$ (*not shown*) decreases steadily when η is increased. These observed trends are in qualitative agreement with the trends that have been predicted by the η-model, which was presented in section 4.3. Quite some similarities are evident when Figure 5.4 is compared to the $0.6 < \eta < 1.0$ part of Figure 4.4. The number of districts participating in the verification as a function of the minimum hail-sensitive fraction decreases from 533 districts at zero down to 55 at 50% hail sensitivity. This dramatic decrease in the number of participating districts makes that the reliability of the verification scores decreases for increasing minimum hail-sensitive fractions. A maximum CSI$'$ of roughly 0.54 is observed for the method of Waldvogel via this selection of districts based on their hail sensitivity, and probably it is close to the true CSI of Waldvogel's method.

This CSI is somewhat higher than the maximum CSI of 0.49 determined previously using all verification data and a η-fraction of 0.9. On average, the hail damage reports will refer to more severe hail events than those reported by the observers of the precipitation network. Therefore, this higher CSI indicates that the method of Waldvogel and probably all methods will detect severe, damaging hail more effectively than light hail.

5.4 Choice of method

Eight different methods for the detection of summer hail using radar have been compared and verified using on-ground hail observations and reports of hail damage. Although there are substantial quantitative differences, the general trends in the apparent POD′, FAR′, and CSI′ of the detection methods as a function of their warning threshold are rather similar. The effects of missing reports and observations in the verification 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. The CSI of 0.54 for the method of Waldvogel, as obtained in the present study, is considerably higher than the CSI of 0.46 which can be deduced from the original verification results as presented by Waldvogel et al. (1979). The obtained CSI compares favorably to the results found for the verification of the NEXRAD 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′ and consequently the probability of hail as a function of the warning threshold, the warning properties of the method of Waldvogel can be altered over a wide range to fulfil the needs of different kinds of users. Detection of hail using the method of Waldvogel will improve the performance considerably as compared to that when using the straightforward CAPPI method.

Chapter 6

Extended verification of Waldvogel's method

Based on the results of the comparison and verification presented in the previous chapter, it was concluded that the method of Waldvogel is the most promising method for detection of summer hail. During the summer of 2000, in this case from early May till the end of September, an experimental hail detection product based on Waldvogel's method has been generated semi-operationally. This hail detection product has been presented real-time to the forecasters and researchers at KNMI via a webpage on the intranet. Using the archived hail detection product and the verification data which has again been gathered at the end of the summer season, an extensive verification of the method of Waldvogel has been performed.

6.1 Run during the summer of 2000

During the summer of 2000 a semi-operational hail detection product has been presented to the forecasters and researchers at KNMI. At the radarsite, an echotop image which displayed the maximum height of the 45 dBZ reflectivity was produced every 15 minutes. Subsequently, this echotop image was combined with the height of the freezing level as obtained from the most recent analysis of the HiRLAM model to produce the hail detection product. Instead of displaying the height difference between the 45 dBZ echotops and the freezing level, the height difference (ΔH) was converted to a probability of hail. For this, the relation between POH and FAR (equation 5.1) and the curve of the FAR versus the warning threshold for Waldvogel's method (see Figure 5.2) have been used. The three warning levels have been characterized by their POH and have been chosen as follows:

- POH=25% (ΔH=$-$0.3 km), level with a high probability of detection (POD) and false alarm ratio (FAR), and thus with a low certainty of hail actually occurring

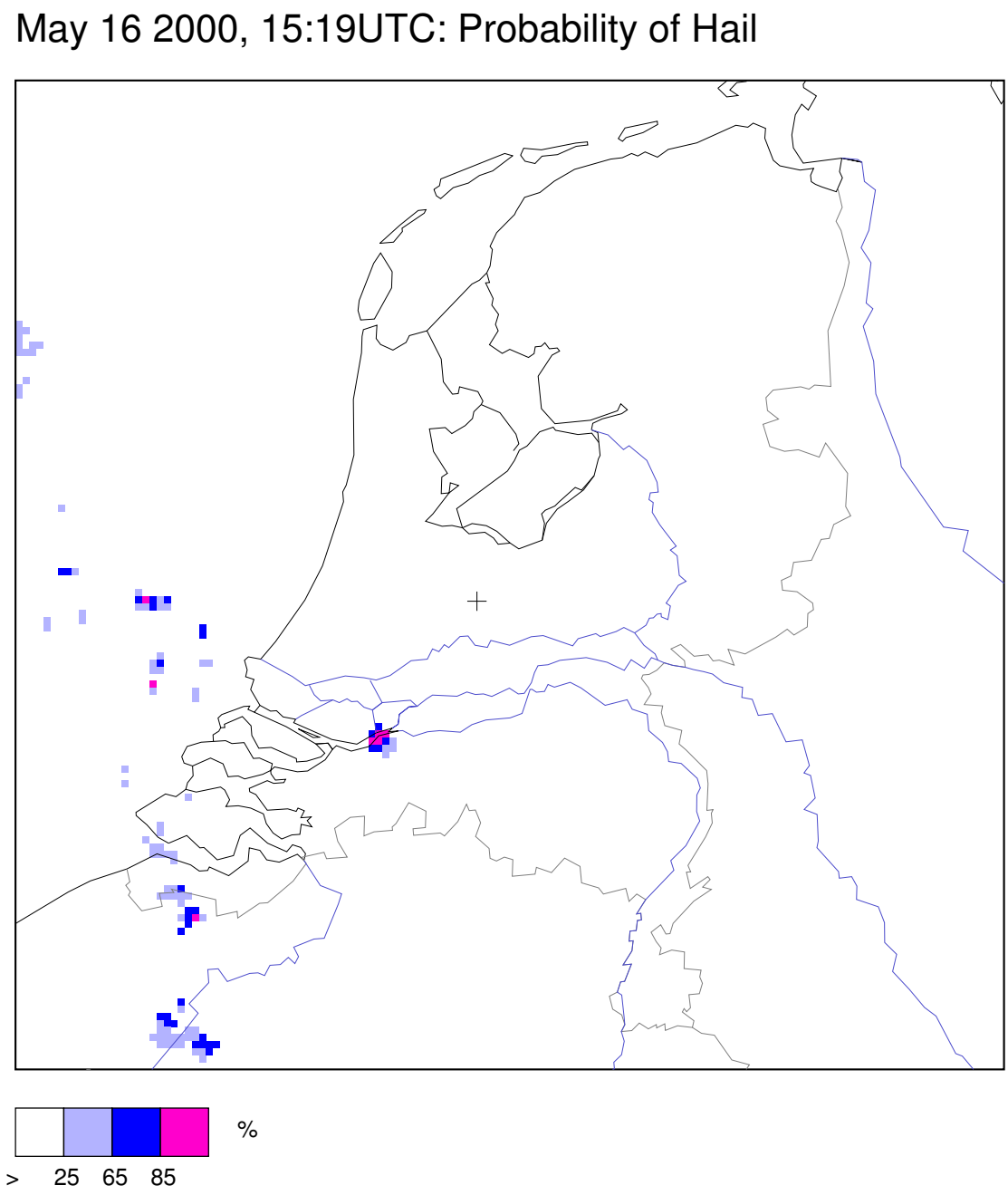

Figure 6.1: An example of the presentation of the hail detection product based on Waldvogel's method. Three probability of hail levels are presented to the forecasters. The hail cell in the southwest of the Netherlands (above Dordrecht) produced hail with a maximum diameter of 4 cm.

- POH=65% (ΔH=1.75 km), level with the best unbiased performance of the hail detection product
- POH=85% (ΔH=4.0 km), level with the lowest FAR and thus with the highest certainty of a hail event actually occurring

An example of the presentation of the semi-operational hail detection product for May 16 of 2000 is shown in Figure 6.1. The thunderstorm in the southwest of the Netherlands (close to Dordrecht) having a probability of hail exceeding 85% produced hail with a maximum diameter of 4 cm at that moment.

In addition to the real-time presentation of the semi-operational hail detection product, an archive of this hail detection product has been made available via the intranet. For this, the 96 images that are produced per day have been binned to single "daily bins" using the method described previously in section 4.2. Such a binned image gives a compact overview of the "hail activity" on a certain day, and an archive of these daily hail images will certainly be useful for some specific applications, like services to insurance companies. Furthermore, these binned images are used in the extensive verification of the hail detection method of Waldvogel (*vide infra*).

6.2 Data from the summer of 2000

6.2.1 Radar and model data

An extensive verification of the semi-operational run during the summer of 2000 of the hail detection product based on Waldvogel's method has been performed. For this, radar data, model data, and verification data has been collected in a similar way as described previously (see section 3.5). The radar and model data are, however, not collected for selected days but for the entire period between early May and end of September of 2000. To be able to store the data for this large number of days, only processed, i.e., two-dimensional, radar data of De Bilt and HiRLAM model data has been gathered. The following data has been stored:

- Radar: echotop heights of 45 dBZ reflectivity in standard projection (see Table 4.1)
- Radar: reflectivity at constant altitude (CAPPI) in standard projection
- Radar: dynamic cluttermap in standard projection
- HiRLAM: freezing-level map, the height of freezing level from the model analysis for a selected square enclosing the area covered by the standard projection of the radar products (see Figure 4.1)

The radar products (96 per day) and the HiRLAM freezing-level maps (8 per day) are available for a total of 135 days between May 10 and September 30 of 2000.

The calculation of the hail detection image of Waldvogel from the available radar data is straightforward. The radar image with the maximum heights of the 45 dBZ reflectivity supplies the basic radar data for Waldvogel's method. At each radar pixel of this image where an echotop of 45 dBZ is observed, the corresponding height of the freezing level is calculated via bilinear interpolation between the four gridpoints of the HiRLAM freezing-level map enclosing the radar pixel. Subsequently, the difference between the maximum height of the 45 dBZ reflectivity and the height of the freezing level is stored in the hail detection image. Finally, some cosmetic operations, the removal of single pixels and the double-check of pixels (see section 4.1) using the corresponding CAPPI image and cluttermap, are applied on the obtained hail detection image.

6.2.2 Verification data

Verification data, both hail observations and damage reports, has been collected again for the summer of 2000 (May till September). From the synop stations of KNMI, a mere total of 26 observations of hail were obtained.

From the KNMI network of precipitation observers, a total of about 500 hail observations during the summer of 2000 were obtained. This is substantially more than the 350 observations of hail obtained for the summer of 1999. Interestingly, about 350 out of the 500 observed hail events have occurred just in May. Assuming that the "hail alertness" of the precipitation observers is constant, a large fraction of the hail events during the summer of 2000 seems to have occurred in May.

Fortunately, the three agricultural insurance companies, which delivered the hail damage reports of 1999 (see section 3.5), have delivered these reports again for the summer of 2000. In total about 1900 reports of hail damage have been obtained from the insurance companies. Again, a large fraction, i.e., 1400 out of 1900, of the reported hail events has occurred in May. Because a large fraction of both the observations by the precipitations stations and the damage reports of the insurance companies refer to hail events in May, it seems that the largest fraction of hail events did occur in May. Sensitivity effects, like the fragility of just germinated crops, should, however, not be overlooked.

6.2.3 Binning of the data

Prior to the verification, the hail detection images and the verification data are converted to daily bins and grouped in districts in the same way as described in section 4.2. Via binning of the 96 images available for each day in the database of 2000, the radar data is reduced to only 135 binned hail detection images. In addition, an updated district map, valid for the year 2000, has been constructed which contains the assigned district number for each radar pixel.

6.3 Verification results

The verification of the extended run during the summer of 2000 of the hail detection product has been performed in the same way in which the underlying method of Waldvogel has been verified on the 15 selected days within the summer of 1999 (see chapters 4 and 5). Analogous to the results presented in chapter 5, the positioning tolerance has been fixed at 12.5 km for all results presented in this section.

6.3.1 Dependence on threshold

The scoring parameters, i.e., the apparent POD′, FAR′, and CSI′, and the biases of the hail detection product based on Waldvogel's method have been determined as a function of the warning threshold by running the verification program repeatedly. The resulting scoring parameters obtained for different selections of the data are shown in Figures 6.2 and 6.3, and the obtained biases are not shown. The results in the upper part of Figure 6.2 are obtained by verification of the method of Waldvogel over 135 days in the summer of 2000, while the results in the lower frame of this figure are obtained by verification over 15 selected days at the end of the summer of 1999. The results presented in the lower frame of this figure have been shown previously (see Figure 5.2). It is evident from Figure 6.2 that there are quite some differences between the apparent performance of the method on the selected days in 1999 and that during the summer of 2000. For the period in 2000, the POD′ is substantially lower and the FAR′ is somewhat lower than the ones determined for the days in 1999. In addition, the maximum CSI′ is less high and shifted towards lower thresholds for 2000 as compared to that for 1999. Because about three-quarters of the observed and reported hail events from 2000 occurred in May and because no days in May where considered in 1999, the observed differences between the results for 1999 and for 2000 could be due to the (too) large impact of the May-events on the verification results of 2000.

In order to assess the impact of the May-events, the dataset of 2000 has been split into two parts: May only and the period between June and September, In the upper frame of Figure 6.3, the verification results of the hail detection product are depicted for May only, and in the lower frame of this figure those for the remaining period between June and September are shown. It is obvious from this figure that the behavior of the hail detection method differs strongly between these two periods. The results for May look rather unusual: both the POD′ and the FAR′ are rather low and drop off rapidly when increasing the warning threshold, and the maximum of the CSI′ is much higher and shifted towards exceptionally low warning thresholds. In contrast, the verification results for the period between June and September 2000 (lower frame of this figure) are rather similar to those from 1999 (lower frame of Figure 6.2). The curves for the POD′ are almost identical, and the FAR′ for June till September 2000 just drops off less rapidly when increasing the warning threshold than the FAR′ of 1999. The optimum performance is found at the same threshold (1.75 km), but that for June till September 2000 is somewhat poorer

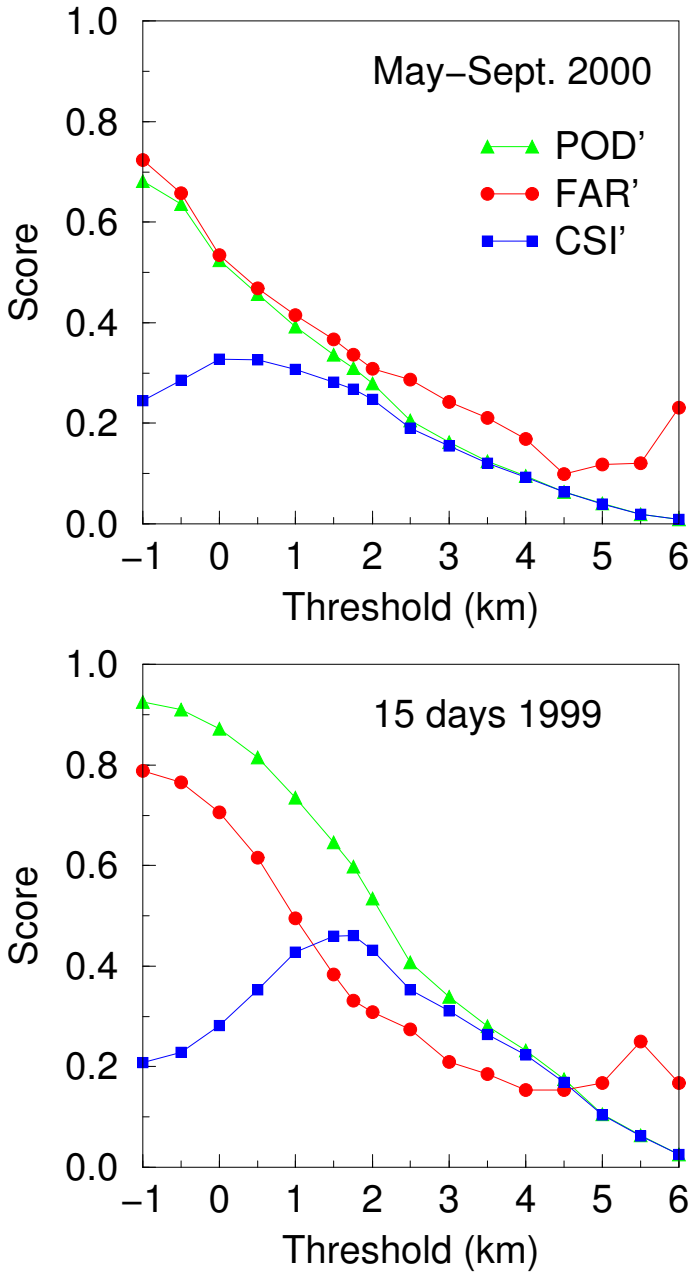

Figure 6.2: The scoring parameters (POD′, FAR′, and CSI′) for the method of Waldvogel for summer of 2000 (upper frame) and for selected days within summer of 1999 (lower frame). The verification has been performed using all available verification data, and using a positioning tolerance of 12.5 km. The lower frame of this figure has been shown previously as part of Figure 5.2.

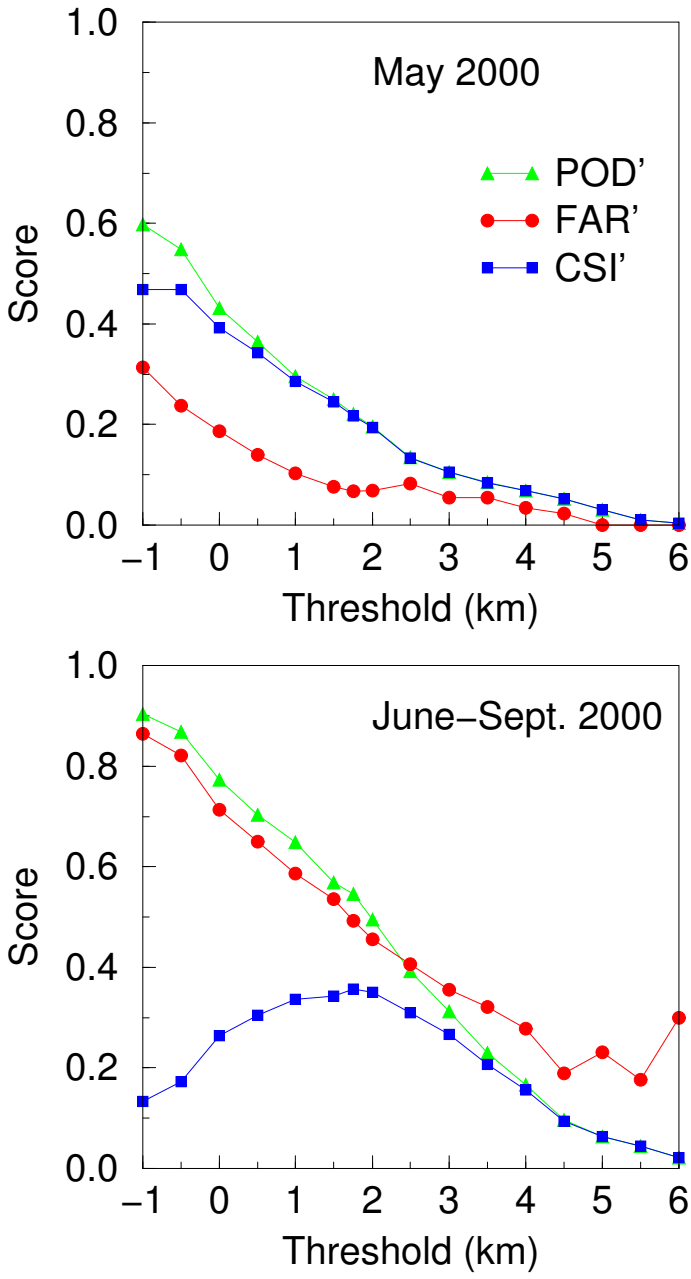

Figure 6.3: The scoring parameters (POD′, FAR′, and CSI′) for the method of Waldvogel for two different periods during the summer of 2000. In the upper frame of this figure, the results for verification during May 2000 only (21 days) are shown. In the lower frame, the results for the period between June and September 2000 (114 days) are shown.

Table 6.1: Two examples of contingency tables that have been obtained from the verification runs of the Waldvogel's method for the summer of 2000. The table contains the result using data of May only and a warning threshold of 1.75 km, and the right table contains the result using data between June and September and again a warning threshold of 1.75 km.

	May 2000	
	Hail	No hail
Radar detection	182	13
No radar detection	643	10334

	June–Sept. 2000	
	Hail	No hail
Radar detection	171	166
No radar detection	142	60169

due to the higher FAR′.

To further illustrate the differences observed between the two periods in the summer 2000, two examples of the contingency tables that have been determined from the verification of Waldvogel's method are presented in Table 6.1. Both tables have been obtained using a warning threshold of 1.75 km. It follows from these tables that the number of hail events in May only (left table) is much higher than that for the period between June and September (right table). In addition, it can be deduced that for the period between June and September the method of Waldvogel has hardly any bias (1.1), while for May a bias of 0.24 is found. Although the number of warnings issued by the hail detection product is about two times higher in May (195) than between June till September (about 80 per month), the number of reported or observed hail events in May is so high, that the method is still underdetecting by a factor of four. It appears that the main reason for the observed differences between the two periods in the summer of 2000 is the extraordinary large number of reported and observed hail events in May.

The large number of reports of hail damage to the agricultural insurance companies in May could be due to an increased sensitivity to hail events during this period. In May many crops have just germinated and many trees are blossoming, and therefore even small hail stones will severely damage this fragile vegetation. This sensitivity effect can be excluded by removing the hail damage reports from the verification data and thus primarily regarding the hail observations by the KNMI precipitation stations. In Figure 6.4 the results of the verification of Waldvogel's method using hail observations by KNMI stations only is shown for the same two periods in the summer of 2000 as before. From a comparison of this figure and Figure 6.3, it is apparent that the verification results using KNMI stations only are rather similar to those obtained using all verification data. For the period between June and September, the observed dependences of the POD′ on the warning threshold are nearly identical and the FAR′ for KNMI stations just falls off slower with increasing threshold than that obtained using all verification data. As a consequence, the maximum CSI′ value is lower for verification using KNMI stations

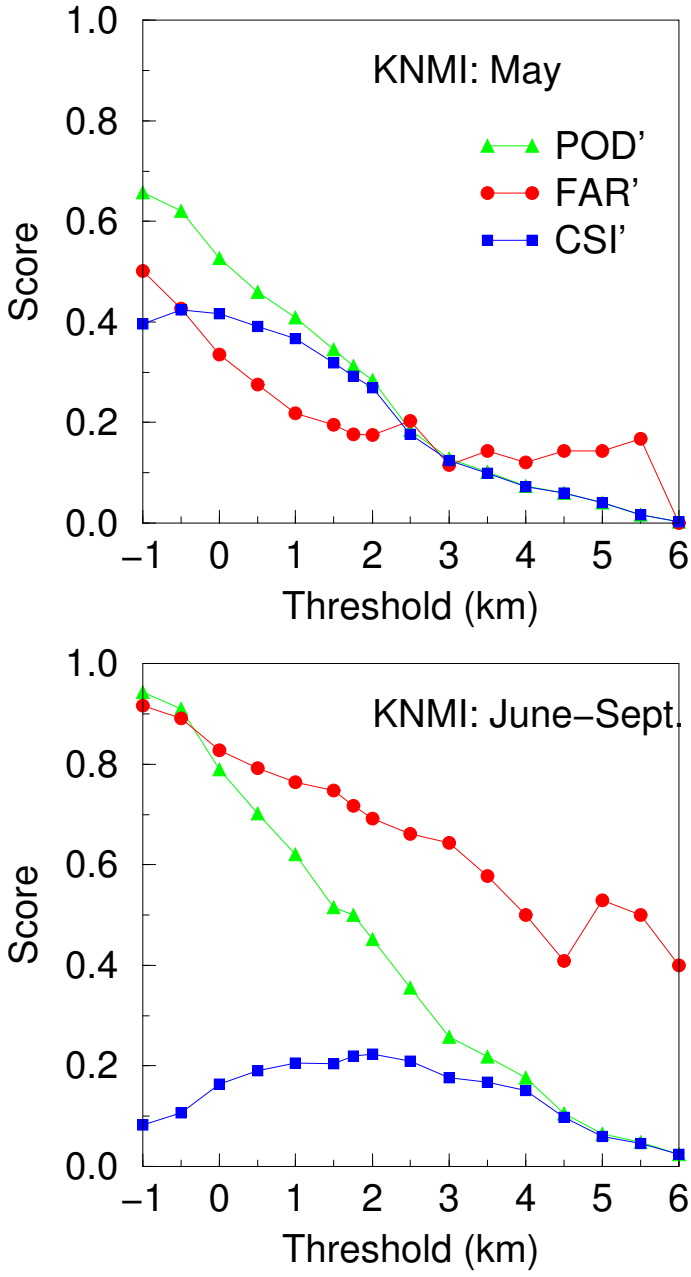

Figure 6.4: The scoring parameters (POD′, FAR′, and CSI′) for the method of Waldvogel from verification using hail observations by KNMI synop and precipitation stations only. In the upper frame the verification results for May only are shown, and in the lower frame those for the period between June and September are shown.

only, but the optimum performance occurs at the same warning threshold. These observed differences are in agreement with the expected trends in the scoring parameters when the fraction of reported hail events (η) is reduced (see Figure 4.4), and the removal of the hail damage reports from the verification data has, of course, led to such a reduction.

The verification results for May using KNMI stations only are presented in the upper frame of Figure 6.4. Both the POD$'$ and the FAR$'$ are somewhat higher than in the corresponding plot of Figure 6.3, but they are still rather low. Furthermore, the optimum performance is still shifted towards rather low warning thresholds, but actually not as far as observed when using all verification data. Finally, Waldvogel's method is still underdetecting the hail events in May (by a factor of 3) using the hail observations from KNMI stations only and a warning threshold of 1.75 km. The small differences between the verification results in May using KNMI stations only and using all verification data are probably due to the above mentioned difference in the fraction of reported hail events η.

Because the verification results for the two periods using just KNMI stations are quite similar to those when using all verification data, it seems unlikely that the main differences between the period in May and the period between June and September are due to large differences in the hail-sensitivity of the vegetation. Previously, the differences between these two periods have been attributed to the extraordinary large number of hail observations and damage reports. Because an increased hail sensitivity of the vegetation is not the main cause of the large number of hail reports in May, an extraordinary large number of hail events should have occurred during May of 2000.

In the introductory chapter, it has been detailed that the hail events occurring in the Netherlands can be divided in two categories: "winter hail" related to large-scale phenomena and "summer hail" related to (small-scale) thunderstorms. In contrast to an event of summer hail, an event of winter hail will be observed by a large number of KNMI stations and will cause damage on a large scale when the vegetation is fragile. In Figure 6.5 the height of the freezing level (upper frame) and the number of districts per day where hail is observed or hail damage is reported (lower frame) are shown as a function of the date. The dashed line separates the period in May from that between June and September. Two periods in May with a large number of hail events are seen to correspond with periods where the freezing level is at rather low altitudes ($\leq$1.5 km). At May 28, the day with the largest number of hail events (see marked period in Figure 6.5), hail has been observed in a total of 174 districts, which is about one-third of the total number of districts in the Netherlands. It is evident from the low altitude of the freezing level and the synoptic situation (several depressions passing the Netherlands) during these periods, that the extraordinary large number of hail events in May are caused by "winter hail".

From the previous discussions on the verification results for May 2000 (see upper frames of Figures 6.3 and 6.4), it is evident that the method of Waldvogel is not suitable, nor was meant to be suitable, for detection of winter hail events. Because the hail detection product under development is intended for the detection of (damaging) summer hail only, the days

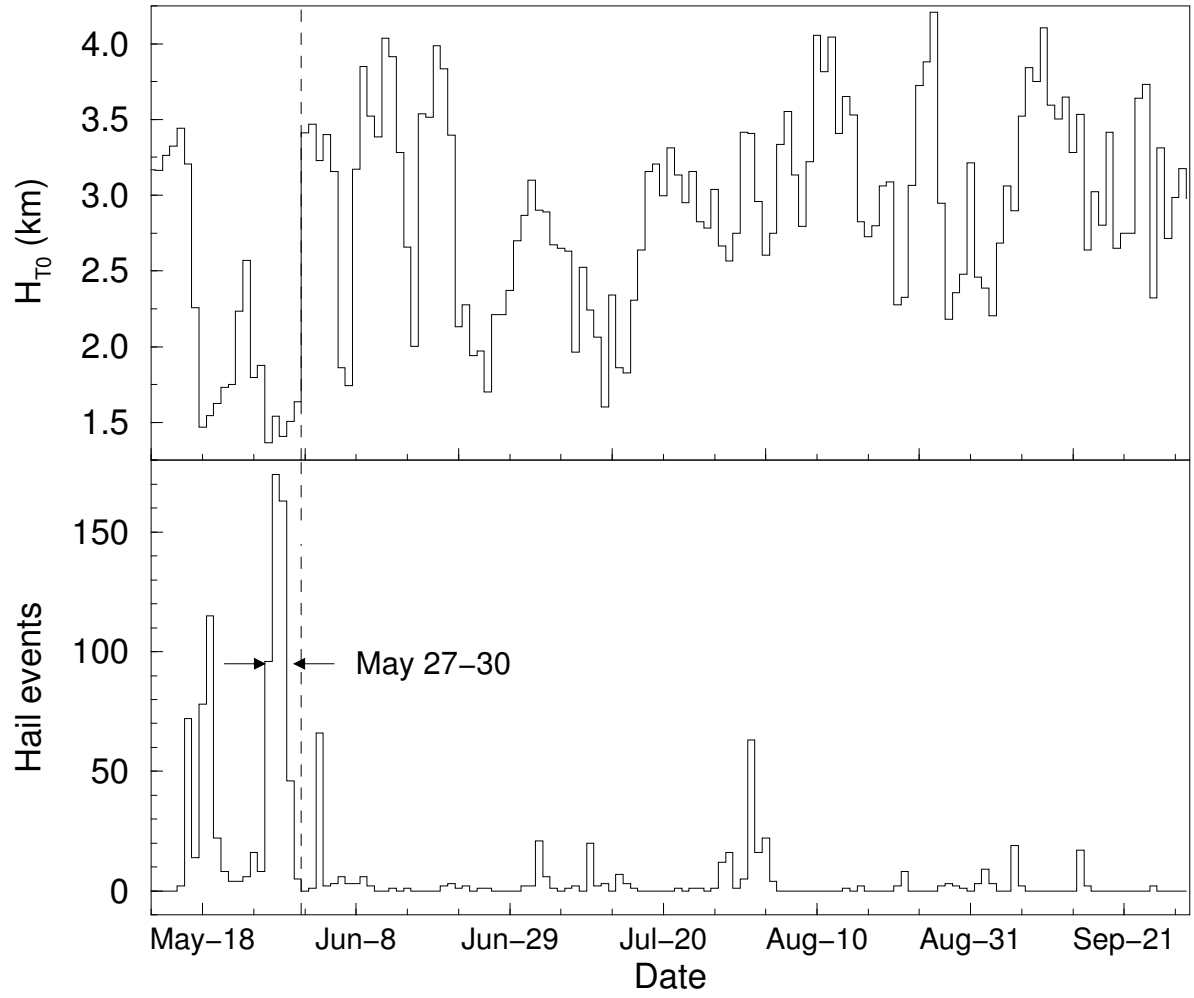

Figure 6.5: In the upper frame of this figure, the height of the freezing level at De Bilt is shown as a function of the date. These heights are determined using the analyses of the HiRLAM model at 12 UTC for all days in the dataset of 2000. The agreement between these heights and those obtained from the 12 UTC sounding data of De Bilt (not shown) is rather good. In the lower frame, the number of districts per day where hail has been observed is shown as a function of the date. The dashed line marks the transition between May and June.

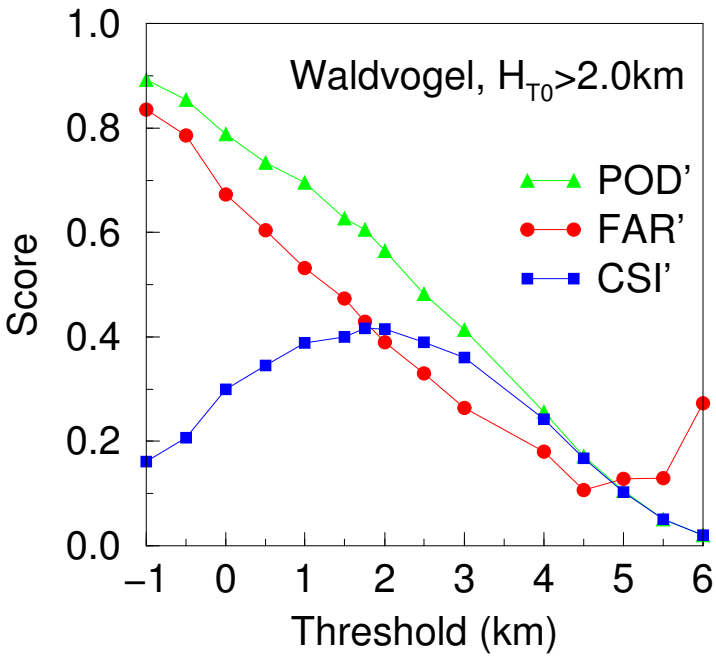

Figure 6.6: The scoring parameters (POD′, FAR′, and CSI′) for the method of Waldvogel from verification using all hail observations and damage reports. Only days where the freezing level at De Bilt is higher than 2.0 km at 12 UTC have been taken into account (115 days between May and September 2000).

where winter hail has occurred have to be excluded from the verification. The height of the freezing level can be used to distinguish between winter hail and summer hail. In Figure 6.6 the verification results for the method of Waldvogel using only days with a freezing level higher than 2.0 km are shown (115 days between May and September 2000). Both hail observations and damage reports have been used for the verification. The "summer hail" results presented in this figure are rather similar to the results for the period between June and September 2000 (lower frame of Figure 6.3). The best performance of the method of Waldvogel, characterized by a CSI′ of 0.42 and bias′ of 1.1, is found at a warning threshold of 1.75 km. From the lowest FAR′ of 0.1, it appears that, just as for 1999, at least a fraction $\eta = 0.9$ of the occurred hail events are contained by the verification dataset. The verification results presented in Figure 6.6 will play a crucial role in the final tuning of the hail detection product based on the method of Waldvogel.

6.3.2 Varying the η-fraction

In section 5.3 an estimation of the true performance, CSI, of the method of Waldvogel on the selected days in 1999 has been obtained by varying the fraction of reported hail events

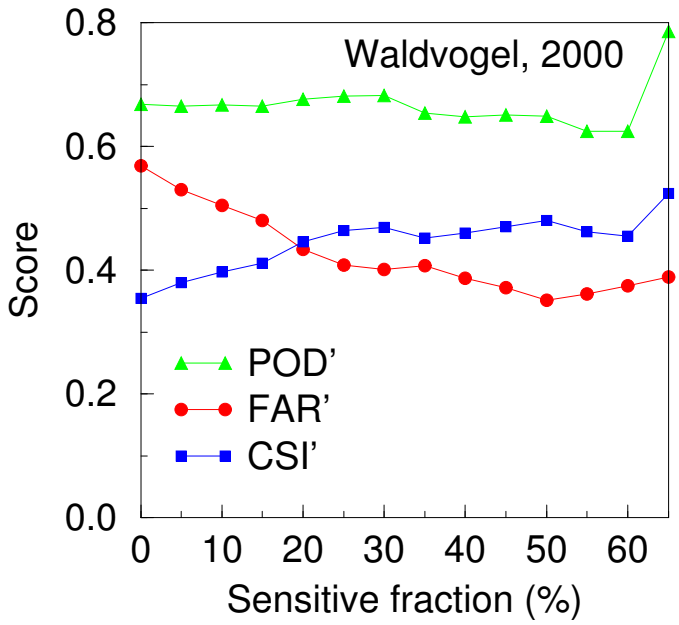

Figure 6.7: The scoring parameters of the method of Waldvogel as obtained on the verification data of hail damage reports of the insurance companies only and using a warning threshold of 1.75 km. Only days during the summer of 2000 with a freezing level higher than 2.0 km have been taken into account. The minimum, required coverage fraction of hail-sensitive land use of a district in order to be taken into account in the scoring parameters is varied.

(η) systematically. This variation was based on a selection of the districts participating in the verification based on their fraction of hail-sensitive land use. For the selected days in 1999, the verification scores of Waldvogel's method as a function of the minimum fraction of hail-sensitive land use of the selected districts have been shown in Figure 5.4. The corresponding results for the summer of 2000 using only days with a freezing level higher than 2.0 km are depicted in Figure 6.7. In agreement with the previous results shown in Figure 5.4, the results for the "summer hail" days in 2000 indicate that the apparent 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 and thus the fraction of reported hail events η is increased. These observed trends are in qualitative agreement with the trends predicted by the η-model, which was presented in section 4.3.

The number of participating districts is decreasing rapidly when the minimum hail-sensitive fraction is raised, and this will reduce the reliability of the verification scores at some point. The obtained verification scores for the summer of 2000 at high minimum hail-sensitive fractions become noisy above a minimum fraction of 55%, while those obtained for the selected days in 1999 become noisy above a fraction of 45% (see Figure 5.4). This improval of the statistics is

due to the larger number of days incorporated in the verification for 2000. From Figure 6.7 the maximum apparent CSI′ is determined to be roughly 0.48 for the method of Waldvogel. This is considerably higher than the maximum CSI′ observed using all verification data and all districts (see Figure 6.6), and it is somewhat lower than the CSI of 0.54 determined for Waldvogel on the selected days in 1999.

The small difference between the performance of the method of Waldvogel for the selected days in 1999 and that for the unselected period in 2000 is remarkable, because the selection of days with thunderstorms only for verification introduces an advantage for the detection method. The reduced performance for the summer of 2000 is caused by both a lowered POD′ and an increased FAR′ (see Figures 5.4 and 6.7) probably due to misses and false alarms, respectively, on days with only minor hail events or no hail at all.

6.4 Summary

A hail detection product based on the method of Waldvogel has been generated semi-operational during the summer of 2000, and it has been shown to forecasters and researchers via a web-page on the intranet of KNMI. In addition, this extended run has been used for a more detailed verification of the method of Waldvogel. Hail observations by KNMI stations and hail damage reports of insurance companies have been used as on-ground verification data again. About three-quarters of the hail events contained by the verification data have occurred in May 2000. The verification results of method of Waldvogel on the events in May only are rather peculiar, because they are dominated by an extraordinary large number of hail events. This large number of hail events appeared not to be related to events of summer hail. The events of winter hail have been excluded from further analyses using a threshold on the height of the freezing level.

The verification results of the method of Waldvogel for the "summer hail" days during the summer of 2000 are rather similar to those for the selected days in 1999, but the significance of the results has increased due to the larger number of days used for the verification. At optimum performance, a CSI of 0.48 is determined for the method of Waldvogel over the period in 2000, while a CSI of 0.54 has been found previously for 1999. The strong lapse of the FAR′ when increasing the warning threshold ensures that, just as for 1999, the warning properties of the hail detection product based on Waldvogel's method can be altered over a wide range to fulfil the needs of different kinds of users. The extended verfication results for the method of Waldvogel confirm the choice for this method, and they clearly demonstrate that a hail detection product based on this method is capable of detecting summer hail with a reasonable accuracy.

Chapter 7

Conclusions and recommendations

From the literature eight different methods for the detection of summer hail using single polarization radar have been selected. This selection includes both methods with a long standing reputation and promising methods that are relatively new. These eight methods have been intercompared using radar data and on-ground verification data for 15 days with thunderstorms in Netherlands during the summer of 1999. Much effort has been put into the collection of the verification data, hail observations by KNMI synop and precipitation stations and hail damage reports from agricultural insurance companies have been collected. The effect on the verification scores of unnoticed hail events, i.e., missed by the observers and without any reported damage, has been investigated in detail, and corrected verification scores have been estimated for the method of Waldvogel. The comparison between the hail detection methods and the verification data has been performed by grouping all data in daily bins and districts.

The performance of all hail detection methods was seen to strongly depend on the maximum allowed spatial mismatch between a radar-based detection of hail and its possible on-ground verification. A maximum spatial mismatch of 12.5 km, reflecting a reasonable radius of influence for a summertime thunderstorm, has been applied to all results presented here.

Although there are substantial quantitative differences, the general trends in the apparent POD$'$, FAR$'$, and CSI$'$ as a function of the warning threshold are rather similar for all eight hail detection methods. Of all methods considered the one of Waldvogel, however, is seen to perform most optimal on the data of the selected days of 1999. Because of this result and other previously mentioned considerations, it was decided that the hail detection product under development will be based on the method of Waldvogel. Subsequently, the effort has been put mainly into further characterizing and tuning of the performance of the method of Waldvogel.

During the summer of 2000 a hail detection product based on the method of Waldvogel has been run semi-operationally. In this way an extended dataset has been gathered containing 135 days between early May and the end of September. The verification results of Waldvogel's method on this large number of unselected days show that the trends of the apparent POD$'$, FAR$'$, and CSI$'$ are almost identical to those observed for the 15 days in 1999. The optimum

performance of the method of Waldvogel in 2000 (CSI=0.48) is somewhat reduced with repect to that in 1999 (CSI=0.54), however, probably caused by the use of unselected days in 2000. The result of this extended verification clearly justifies the choice for the method of Waldvogel, and in addition it confirms that the method is primarily suited for the detection of summer hail. The verification results from both 1999 and 2000 reveal that the FAR$'$ of the method of Waldvogel can be altered over large range by applying different warning thresholds. Therefore, the warning properties of the method of Waldvogel can be tuned over a wide range to fulfil the needs of different kinds of users. In conclusion, a hail detection product based on the method of Waldvogel performs significantly better and offers more functionality than the use of the CAPPI display for hail detection which is present-day practice at KNMI.

A few aspects of the behavior of the hail detection methods could not be addressed in this study due to limitations of the available verification data. Because of the lack of temporal information for most of the hail observations and damage reports, the hail detection methods could not be evaluated on the basis of the timeliness and punctuality of their hail warning. On a few cases, where the time of the hail events was known, the timing of the hail warning by Waldvogel's method has been examined. For the case of August 8 of 1999 (see Figure 4.2), hail with a maximum diameter of 3.5 cm has indeed been observed in Baarle-Nassau (on south border of the Netherlands) at the indicated time. Because the method of Waldvogel is based on radar echoes at high altitudes, it will be able to detect a developing thunderstorm in an early stage, i.e., before (damaging) hail reaches the ground. Just as for the time of hail events, there is hardly any data available on the (maximum) size of the hail in a certain event, and therefore the performance of the detection methods as a function of hail size could not be investigated systematically. For the cases in 1999 where very large hail ($>$3 cm) was observed, however, all detection methods indicated the presence of hail.

The operational implementation of a hail detection product based on the method of Waldvogel is rather straightforward. The construction of the hail detection product can be separated into three processes. First of all, an image with the echotops of the 45 dBZ reflectivity has to be constructed from the radar data. This echotop image can be obtained from mixing the data of both the Den Helder and the De Bilt radar of KNMI similarly to the standard low-threshold (7 dBZ) echotop product. Secondly, a field, which is just enclosing the area covered by the radar echotop product, has to be extracted from the HiRLAM numerical weather prediction model containing the geopotential height of the freezing level. The most recent analysis of HiRLAM has been used throughout this study for this purpose, but for operational implementation the nearest forecast based on the most recent analysis is recommended. Finally, the radar echotop data and the HiRLAM freezing level information have to be combined to obtain the hail detection product. The obtained difference between the maximum height of the 45 dBZ reflectivity and the height of the freezing level can be converted to the probability of hail using the curve of the FAR$'$ versus the warning threshold as depicted in the lower frame of Figure 6.6. In a similar way as for the semi-operational run during the summer of 2000, three probability of hail levels can be used for the presentation of the hail detection product to the users. Daily

bins of the hail detection products with the original 8-bit height resolution should be produced and archived. It is recommended to interpolate between the 15-minute images to avoid the presentation of "jumping cells". These daily bins can be valuable both for reference by for instance insurance companies and for additional verification of the product.

Acknowledgments

Kees Goedhart (Hagelunie), Gert-Jan van Dijk (OFH), and Gerold de Jonge (Agriver) are gratefully acknowledged for supplying the reports of hail damage in 1999 and 2000, and Frans van der Wel is acknowledged for providing land use data and geographical data in the proper format. The skilful assistance of Hans Beekhuis in various radar-related technical issues is appreciated. Herman Wessels is acknowledged for his interest and for many useful suggestions. The involvement of Sylvia Barlag, Jeanette Onvlee, and Robert Mureau with this project is highly appreciated. The detailed comments and suggestions of Kees Kok on the interpretation of the verification results have been of great use.

Bibliography

Amburn, S. A. and P. L. Wolf: 1997, Vil density as a hail indicator. *Wea. and. Forecasting*, **12**, 473–478.

Auer, A. H., Jr: 1994, Hail recognition through the combined use of radar reflectivity and cloud-top temperatures. *Mon. Wea. Rev.*, **122**, 2218–2221.

Aydin, K., T. A. Seliga, and V. Balaji: 1986, Remote sensing of hail with a dual linear polarization radar. *J. Appl. Meteor.*, **25**, 1475–1484.

Billet, J., M. DeLisi, B. G. Smith, and C. Gates: 1997, Use of regression techniques to predict hail size and the probablility of large hail. *Wea. and. Forecasting*, **12**, 154–164.

Doswell, C. A., R. Davies-Jones, and D. L. Keller: 1990, On summary measures of skill in rare event forecasting based on contingency tables. *Wea. and. Forecasting*, **5**, 576–585.

Edwards, R. and R. L. Thompson: 1998, Nationwide comparisons of hail size with wsr-88d vertically integrated liquid water and derived thermodynamic sounding data. *Wea. and. Forecasting*, **13**, 277–285.

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

Hardaker, P. J. and A. H. Auer, Jr.: 1994, The separation of rain and hail using single polarization radar echoes and ir cloud-top temperatures. *Meteor. Appl.*, **1**, 201–204.

Höller, H., V. N. Bringi, J. Hubbert, M. Hagen, and P. F. Meischner: 1994, Life cycle and precipitation formation in a hybrid-type hailstorm revealed by polarimetric and doppler radar measurements. *J. Atmos. Sci.*, **51**, 2500–2522.

Kessinger, C. J., E. A. Brandes, and J. W. Smith: 1995, A comparison of the nexrad and nssl hail detection algorithms. *27th conference on Radar Meteorology*, AMS, 603–605.

Kitzmiller, D. H., W. E. McGovern, and R. E. Saffle: 1995, The wsr-88d severe weather potential algorithm. *Wea. and. Forecasting*, **10**, 141–159.

Kok, C. J.: 2000, On the behavior of a few popular verification scores in yes/no forecasting. Scientific report 2000-04, Royal Netherlands Meteorological Institute (KNMI).

Lenning, E., H. E. Fuelberg, and A. I. Watson: 1998, An evaluation of wsr-88d severe hail algorithms along the northeastern gulf coast. *Wea. and. Forecasting*, **13**, 1029–1044.

Ludlam, F. H.: 1980, *Clouds and Storms: The behavior and effect of water in the atmosphere*. Pennsylvania State University Press.

Machenhauer, B.: 1988, Hirlam final report. Technical report No. 5, Danish Meteorological Institute, also: http://www.knmi.nl/hirlam.

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

Smart, J. R. and R. L. Alberty: 1985, The nexrad hail algorithm applied to colorado thunderstorms. *14th conference on Severe Local Storms*, AMS, 244–247.

Smith, P. L.: 1999, Effects of imperfect storm reporting on the verification of weather warnings. *Bull. Amer. Meteor. Soc.*, **75**, 1099–1105.

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

Waldvogel, A., B. Federer, and P. Grimm: 1979, Criteria for the detection of hail cells. *J. Appl. Meteor.*, **18**, 1521–1525.

Waldvogel, A., B. Federer, W. Schmid, and J. F. Mezeix: 1978b, The kinetic energy of hailfalls. part 2: Radar and hailpads. *J. Appl. Meteor.*, **17**, 1680–1693.

Waldvogel, A., W. Schmid, and B. Federer: 1978a, The kinetic energy of hailfalls. part 1: Hailstone spectra. *J. Appl. Meteor.*, **17**, 515–520.

Wessels, H. R. A.: 1990, Coordinate conversions for presenting and compositing weather radar data. Technical report 129, Royal Netherlands Meteorological Institute (KNMI).

Wessels, H. R. A. and J. H. Beekhuis: 1997, Stepwise procedure for suppression of anomalous ground clutter. *COST-75 Seminar on Advanced Radar Systems*, EUR 16013 EN, 270–277.

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

WMO: 1966, International meteorological tables.