14A.6 FORECASTING HAIL SIZE BY COMBINING A NWP MODEL WITH A HAIL GROWTH MODEL

Gerhard W. Reuter and Julian C. Brimelow*
Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, Alberta

Ron Goodson
Meteorological Service of Canada, Prairie and Northern Region, Edmonton, Alberta

And

Terrence W. Krauss
Weather Modification Inc., Fargo, North Dakota

1. INTRODUCTION

Hailstorms pose a serious economic problem to society. According to Changnon et al. (2000), in the U.S., losses from hailstorms exceeding $300 million (USD) have become frequent since the 1990s. Costly hailstorms have also been observed in Canada, with damages associated with a hailstorm that struck Calgary on 7 September 1991 estimated at $400 million (CAD) (Charlton et al. 1995). Severe weather algorithms, developed for weather radar, are routinely used to identify and nowcast the movement of hailstorms (Witt et al. 1998; Marzban and Witt 2001; Joe et al. 2004). However, lead times of radar-based warnings are very short and the algorithms show limited skill at forecasting the actual hail size on the ground (Edwards and Thompson 1998). An alternative to radar-based hail algorithms is to use upper-air soundings to relate the maximum hail size on the ground to the estimated maximum updraft velocity (Foster and Bates 1956; Renick and Maxwell 1977; Moore and Pino 1990). Doswell et al. (1982), however, found significant shortcomings with buoyancy-based methods used to forecast hail size.

Brimelow et al. (2002a) proposed a new approach to forecast maximum hail size on the ground using HAILCAST, which is a one-dimensional steady-state cloud model combined with a time-dependent hail growth model. In that study, HAILCAST was run using observed proximity soundings. The predicted hail sizes were compared against reports of maximum hail size gathered from a high-density observation network within the Alberta Hail Project area. Brimelow et al. (2002a) demonstrated that HAILCAST is skilful at forecasting the maximum expected hail size on the ground when initialized with representative proximity upper-air soundings and surface conditions. HAILCAST has subsequently been used in Argentina (Brimelow et al., 2002b), South Africa and the United States (Jewell and Brimelow 2004).

**Corresponding author address:* Julian Charles Brimelow Department of Earth and Atmospheric Sciences University of Alberta, Edmonton, Alberta T6G 2E3, Canada; Email: brimelow@ualberta.ca

The importance of using representative upper-air sounding data for predicting the maximum expected hail size cannot be overstated. Proximity soundings are usually identified by applying spatial and temporal constraints between the soundings and the observed storms (e.g., Brooks et al. 1994). However, in an operational setting, obtaining proximity soundings is complicated by the high spatial and temporal variability typically present in the pre-storm environment and by the coarse spacing of upper-air networks. This is particularly relevant over the Canadian prairies where there are only two sounding sites (see Fig. 1): Stony Plain (53.53°N, 114.10°W) and The Pas (53.96°N, 101.10°W). These sites are almost 900 km apart and soundings are made only twice a day. Consequently, the observed soundings often are not representative of the antecedent thunderstorm conditions.

This study introduces an innovative hail forecasting technique that predicts where and when the largest hail is expected to fall with a lead time of up to 12 hours. Our approach is based on using model soundings predicted by the Global Environmental Multiscale (GEM) model as input data for HAILCAST. GEM (Côté et al. 1998) is the operational weather prediction model used by Environment Canada for issuing public and aviation forecasts. For the purpose of this study, GEM-forecasted soundings of temperature, humidity and wind were generated on a 0.5-degree horizontal grid. These soundings provided the input data for HAILCAST.

The concept of using forecast soundings, generated by NWP models, for predicting the intensity of convection has been explored before (Hart et al. 1998; Thompson 1998; Hamill and Church 2000; Thompson et al. 2003). Specifically, Hamill and Church (2000) and Thompson et al. (2003) found that the Rapid Update Cycle model (RUC) model is generally capable of producing prognostic soundings, which can be used effectively to discriminate between environments that support thunderstorms of varying intensities. In this paper, we determine the feasibility of using prognostic GEM soundings to assist in forecasting the occurrence of hail over the Canadian prairies. We also examine the technique's skill in predicting the spatial distribution of hail and the maximum hail size.

For each day between 1 June and 31 August

2000, contour maps of the maximum forecast hail size on the ground (valid for 00 UTC) were generated by running HAILCAST on a grid consisting of about 1400 GEM 12-h forecast soundings. The forecast hail maps were compared against radar reflectivity data and surface hail reports. The skill of the forecast technique was quantified using the Probability of Detection, False Alarm Ratio, and Critical Success Index.

2. HAILCAST AND THE ENSEMBLE HAIL FORECASTING TECHNIQUE

HAILCAST consists of a steady-state cloud model linked to a hail growth model. The cloud model requires vertical profiles of ambient temperature, humidity and wind. These data are used to compute vertical profiles of liquid water content, updraft velocity and in-cloud temperature that are representative of the hail growth environment close to the updraft's near-adiabatic core. The time-dependent hail model then uses these data to simulate the growth of hail in the updraft. A drizzle-sized hail embryo is introduced at cloud base and allowed to grow by either wet or dry growth. Allowance is made for melting of the hailstone as it descends below the in-cloud freezing level. During wet growth (melting), excess accreted water (meltwater) on the surface of the stone is shed. Details of the HAILCAST model are provided in Brimelow et al. (2002a).

One of the challenges facing forecasters when predicting the initiation and strength of convection is the spatial distribution of moisture in the boundary layer. Low-level moisture and temperature fields are not homogeneous and can vary on spatial and temporal scales that are much smaller than those of most observation networks. Mueller et al. (1993) and Crook (1996) employed multi-dimensional cloud models to study the sensitivity of thunderstorm initiation and intensity to fluctuations of temperature and moisture in the boundary layer. Their experiments demonstrated that the modeled convection was sensitive to small changes in the surface input data. Specifically, Crook found that small variations in the surface temperature (1°C) and moisture (1 g kg^{-1}) could differentiate between no convection and intense convection. Similarly, model sensitivity experiments showed that model output from HAILCAST was sensitive to changes as small as 1°C in the surface temperature and dewpoint (Brimelow 1999).

To improve our ability to predict the evolution of the atmosphere, one can perform a number of model simulations (or ensemble members), each starting with slightly different initial conditions (Crook 1996). Brooks et al. (1992) suggested adopting a quasi-Monte Carlo, or probabilistic approach, when employing numerical cloud models to forecast convection. This approach involves varying the input data for the cloud model over a range of values expected in the area where convection is anticipated. Assuming that that the range of initial conditions spans the domain of expected error, the ensemble mean may provide a more skilful forecast than the majority of the individual forecasts. Adopting an ensemble approach has been found to mitigate the sensitivity of numerical weather prediction model forecasts to uncertainties in the input data (e.g., Stensrud et al. 2000).

HAILCAST is computationally efficient and is thus well suited for producing ensemble forecasts. In this study, the ensemble forecasts were prepared by varying both the temperature and dewpoint forecast at each surface grid point by the GEM model between –1.0 ºC, -0.5 ºC, 0 ºC, +0.5 ºC and +1.0 ºC. HAILCAST was then run for each combination of the temperature and dew-point, resulting in a total of 25 individual hail diameter forecasts. The ensemble diameter was determined by calculating the arithmetic mean of all 25 forecast hail diameters. The above ranges in temperature and dewpoint were selected to represent the surface temperature and moisture variations expected in the boundary layer on any given day. At this stage, the forecast profiles of temperature, dewpoint and wind above the surface were assumed to remain unchanged. The reason for this is that sensitivity testes conducted by Brimelow (1999) showed that of all the parameters considered, small variations in the surface temperature and dewpoint overshadowed the changes in the modeled hail size achieved by varying other microphysical parameters in the hail model.

3. PROGNOSTIC SOUNDING DATA

Our study focused on three Canadian prairie provinces: Alberta, Saskatchewan and Manitoba (Fig. 1). The forecast technique was evaluated for 92 days between 1 June and 31 August 2000.

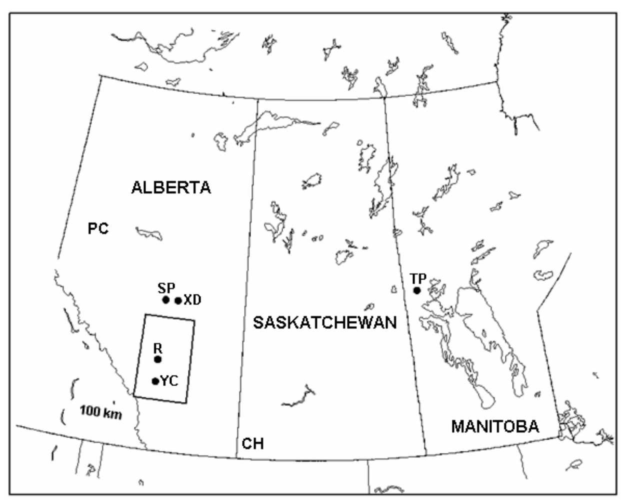

Fig. 1. Forecast domain of the GEM/HAILCAST hail forecasting technique for the summer of 2000. SP and TP are the locations of the operational upper-air sounding sites at Stony Plain and The Pas, respectively. The city of Edmonton is located at XD, the city of Calgary at YC, the radar site at R, the Peace country near PC and the Cypress Hills near CH. The large rectangle indicates the outline of the Alberta Study Area (ASA).

To capture the pre-storm environment on each day, we used prognostic upper-air soundings produced by the 12 UTC run of the 24-km resolution GEM model.

In this study, a total of 1403 GRIDded Binary format (GRIB) soundings were generated for 00 UTC at 0.5-degree intervals between 49º N and 60º N and 120º W and 90º W. The forecast pressure, height, temperature, dewpoint, wind speed and direction were available at 14 vertical levels between the surface and 150 hPa. HAILCAST was run at each grid point and the resulting field of hail diameters was then contoured to create spatial maps of the maximum expected hail size.

It is important to note that the forecast profiles of moisture and temperature in the low-levels of the atmosphere can differ from those observed in nature due to shortcomings in the NWP models. For example, steep orography, model boundary layer parameterizations, data assimilation methods and land use category schemes represent only a few of the complex factors that can have an important impact on NWP output (Hanna and Yang 2001). The parameterization of convection in NWP models is another potential source of uncertainty in the forecast profiles (Grell 1993).

Model prognostic soundings may contain inaccuracies on account of errors in the numerical data fields that specify the amount of instability prior to convective initiation, erroneous prediction of the timing and location of convection, and errors in the expected evolution of the convection (e.g., Grell 1993). In addition to the aforementioned errors, errors can also arise from modifications to the temperature and moisture profiles once the convection parameterization scheme is initiated. For example, the Betts-Miller-Janjić (BMJ) parameterization scheme (Betts 1986; Betts and Miller 1986; Janjić 1994) used in the National Centers for Environmental Prediction Eta model has been found to remove small-scale vertical features, such as capping lids or stable layers, once convection is initiated (Baldwin et al. 2002). In other words, using certain CP schemes can modify prognostic soundings, at grid points where convection has been triggered, resulting in forecast profiles that are not representative of the pre-storm environment. The version of the GEM used in this study employed the Fritsch-Chappell scheme (Fritsch and Chappell 1980) to simulate convection.

4. VERIFICATION DATA

4.1 Surface hail report database

In total, 533 reports of hail were collected for the period between 1 June and 31 August 2000 for the purpose of verifying the accuracy of the HAILCAST hail-size forecasts. Using surface hail reports has inherent problems, including uncertainty regarding the accuracy of the time and location the hail occurred, as well as the accuracy of the hail size measurements (Lenning et al. 1998). The latitude, longitude and time of a given hailstorm were only available for the U.S. Storm Prediction Center (SPC) and Meteorological Service of Canada (MSC) reports, which accounted for less than a third of all the reports collected. SPC reports of severe hail made within half a degree latitude (~ 55 km) of the U.S.-Canada border were included in the database. For the remaining reports, the latitude and longitude of the town where the hail was reported were used to specify the location of the hail events. The SPC reports were obtained from the preliminary severe weather log, but were corroborated using the final storm log. Lightning data were used to ensure that convection occurred in the immediate vicinity of the hail reports and to determine whether the hail fell between 20 and 05 UTC. This time frame was selected to increase the likelihood that the 00 UTC GEM forecast soundings would be representative of the convective environment.

The majority of the surface hail reports (332 or 62%) was obtained from the Alberta Agriculture Financial Services Corporation, with 153 (29%) from the MSC severe weather database, 28 (5%) from Weather Modification Inc. and 20 (4%) from the SPC severe weather database. Of the 533 hail reports in the database, 425 (almost 80%) of the hail reports were from Alberta, compared to 77 (14%) from Saskatchewan and 31 (6%) from Manitoba.

If, on a given day, more than one hail report was received from the same location, the largest size was used. A hail day was classified as severe when the reported (or forecast) hail diameter was 2.0 cm or larger. According to surface reports, 61 (66%) of the 92 days in the dataset were identified as hail days. Of these, 44 (72%) were classified as severe hail days. There were 268 reports of non-severe hail and 265 reports of severe hail. Reports of grape-sized hail were the most frequent (28% of all reports), followed by walnut-sized hail (25%). The number of reports of pea-sized hail and golfball-sized hail were almost the same (22%). Only 3% of the reports were for larger than golfball-sized hail.

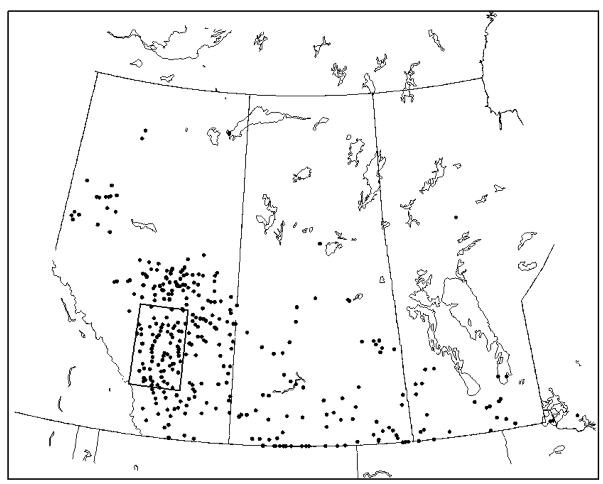

Fig. 2. Hail reports for the period 1 June to 31 August 2000.

Figure 2 shows all of the locations where hail was reported between 1 June and 31 August 2000. The high concentration of hail reports over south-central Alberta is evident. The reports also reflect the presence of the main highway between Edmonton and Calgary. Very few reports of hail were received north of 55° N. This is probably, in part, a result of the low population density which is less than 1 km^{-2} over the majority of the Canadian prairie provinces and only exceeds 10 km^{-2} over a few localized areas.

On account of the low population density over the majority of the Canadian prairies, many hailstorms go undetected. Thus surface hail reports are not suitable for verifying high spatial resolution forecast hail maps. We therefore focused our verification of the hail forecasts over a 2.5º by 2.5º box (175 km east-west by 275 km north-south) over south-central Alberta (Fig. 1), because continuous radar data were available over the Alberta Study Area (ASA) during the summer of 2000. The ASA also encompasses the climatogically preferred region for hailstorms (Etkin and Brun 1999). In addition, the area of the ASA (48 000 km^2) is similar in area to the typical watch area issued by the US Storm Prediction Center (40 000 to 50 000 km^2) and Canada's Prairie Storm Prediction Centre.

4.2 Radar reflectivity data

To avoid the problems associated with relying on a sparse surface observation network to report hail, we used radar data to infer the presence of hailstorms. The advantage of weather radar is that, by virtue of its continuous space and time coverage, it detects all precipitating cells within its viewing area.

Radar reflectivity data for the period 1 June to 31 August 2000 were collected by a C-band radar located at the Olds-Didsbury Airport (51.71°N, 114.11°W, elevation 1024 m) in southern Alberta (Fig. 1). Volume scans were performed at 5-min intervals throughout the summer. The radar transmits at a peak power of 250 kW, has a beamwidth of 1.65º, and a range gate size of 900 m. Most of the ASA was located within 150 km of the radar site, which would minimize sampling errors. Radar data were displayed using the TITAN cell tracking system (Dixon and Wiener 1993). The radar performance parameters were checked weekly and a complete calibration was performed each month during the summer.

Lenning et al. (1998) found that the use of radar-derived Vertically Integrated Liquid (VIL) water content was promising for indicating the presence of hail in thunderstorms near Tallahassee, Florida. Specifically, they indicated that selecting VIL thresholds can be useful for distinguishing between significant and marginal hail events, and between marginal and non-events. Brimelow et al. (2004) also found VIL data useful for identifying observed hailfall over central Alberta. In particular, comparisons with surface hail reports indicated that a VIL threshold of 25-30 kg m^{-2} was effective at correctly identifying those storms associated with reports of severe hail over central Alberta. Comparisons of VIL measurements with hail reports over Alberta suggest that an appropriate lower VIL threshold for hail is 10 kg m^{-2}. Likewise, Kitzmiller et al. (1995) noted that VIL values in organized convective cells usually exceeded 10 kg m^{-2}.

Using the guidelines of Brimelow (2004), a day in the ASA was classified as a probable hail day if the VIL at one or more pixels (~ 1 km^2) was larger than or equal to 10 kg m^{-2}. If the VIL exceeded 25 kg m^{-2} at two or more contiguous pixels (~3 km^2), then the day was classified as a severe hail day. The maximum VIL maps were created by recording the maximum VIL observed at each pixel in the radar's domain. Figure 3 shows an example of a maximum VIL map (valid for 4 July 2000) that was used to verify the hail forecasts and hail reports. The locations where hail was reported are also indicated on the map. The location and severity of the 10 surface hail reports corresponded well with the radar-derived VIL data. However, there were several locations without surface hail reports, yet the VIL values were the same or higher than at those locations where hail was reported. Specifically, there were at least 15 locations over the ASA where the VIL data suggested that hail was likely present but no surface reports of hail were received. This day is a typical example of cases when using surface hail reports alone would have underestimated the occurrence of hail in sparsely populated areas.

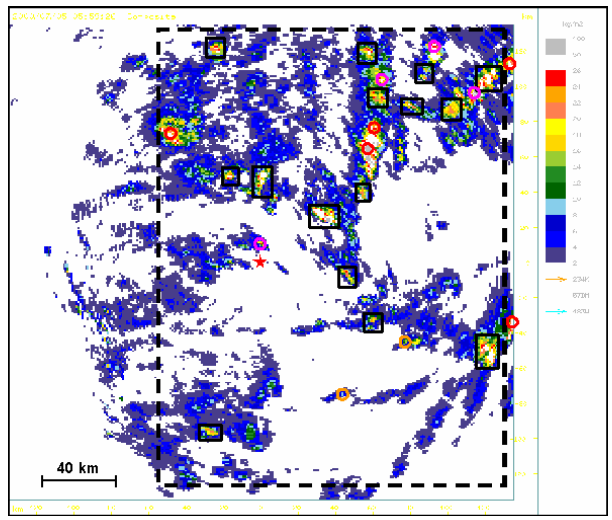

Fig. 3. Maximum VIL map for the ASA (large dashed rectangle) on 4 July 2000. Green circles represent locations where pea-sized hail was reported, pink circles the locations where grape-sized was reported and red circles locations where walnut-sized hail or larger was reported. The red star indicates the location of the Olds-Didsbury radar. Black rectangles represent locations where the VIL field indicated the presence of hail but no hail was reported (see text for details).

The daily maximum VIL maps included the locations of observed hail reports with few exceptions. In fact, there were only three occasions (of 151 hail reports) when reports of hail were received over the ASA, but the VIL maps did not indicate the presence of hail. For one event, attenuation of the radar signal by a severe storm located between the radar and the report probably was responsible for the radar underestimating the strength of the distant hailstorm. The two remaining reports were both located more than 120 km from the radar, where VIL values were ~9 kg m^{-2}, which is just below the 10 kg m^{-2} threshold used for hail.

Radar indicated that hail was likely present on nine days when no reports of hail were received over the ASA. Similarly, radar indicated the potential for severe hail over the ASA on four days when no hail reports were received. Radar also indicated severe hail on 11 days when reports of non-severe hail were received.

On nine of these days, grape-sized hail was reported, while on the remaining two days pea-sized hail was reported. The absence of severe hail reports on certain days could be because the observers failed to measure the largest hail size, or because severe hail fell over a sparsely populated area.

Studies investigating the utility of VIL as an indicator of severe hail have shown that severe VIL thresholds display a strong regional dependence (Wagenmaker 1992; Kitzmiller et al. 1995; Edwards and Thompson 1998). Consequently, Paxton and Shepherd (1993) proposed an algorithm to calculate an appropriate VIL threshold for severe hail that is dependent on the airmass characteristics. This threshold is referred to as the VIL of the day or VOD. On all days when the observed VIL was larger than or equal to 26 kg m^{-2}, the maximum VIL observed over the ASA exceeded the VOD calculated from the 00 UTC Stony Plain sounding. It is thus unlikely that a day was incorrectly classified as a severe hail day. In addition, it is unlikely that the threshold of 26 kg m^{-2} to identify severe hail was too high, as none of the 48 severe hail reports over the ASA were associated with maximum VIL values less than 26 kg m^{-2}.

5. COMPARISON OF HAILCAST MAP AND OBSERVATIONS FOR 4 JULY 2004

Figure 4 shows a map of the forecast hail size valid for 00 UTC 5 July 2004. The colour scale for the maximum largest expected hail diameter is labelled in cm. HAILCAST was initialized using the prognostic GEM soundings from the 12 UTC model run. On this day, several severe thunderstorms produced golfball-sized hail over east-central Alberta. The location and size of the hail reports suggested that the HAILCAST guidance was good, with eight of the nine severe hail reports located in close proximity (< 50 km) of the forecast 2 cm hail contour. Also, the maximum forecast hail diameter of 4 to 5 cm agreed well with the reports of golfball-sized hail. The forecast indicated that the potential for severe hail decreased as one moved southwestwards from the region of large hail forecast over central Alberta. This tendency was supported by surface reports and radar data (Fig. 4), with no severe hail being reported or indicated by radar over the southern ASA.

The model also forecast that the potential existed for large hail over far southeastern Saskatchewan. Although no reports of severe hail were received from this area, weather radar identified several strong thunderstorms in the area enclosed by the circle in Fig. 4 between 00 and 03 UTC. The maximum reflectivity observed in these storms was between 50 and 55 dBZ, which suggests that hail may have been present at the surface (Barge 1974; Changnon 1992).

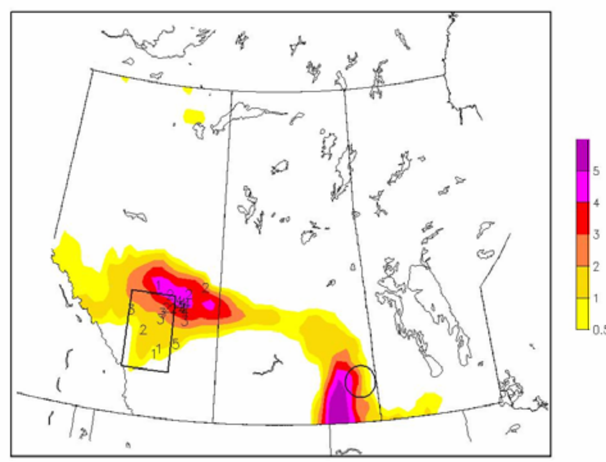

Fig. 4. Hail forecast map generated using prognostic GEM soundings and HAILCAST for 4 July 2000. Forecasted hail diameters are contoured in centimeters. The large rectangle encloses the ASA and the circle over southeastern Saskatchewan indicates where weather radar observed several strong thunderstorms. Surface hail reports are indicated by numbers, with "1" representing pea-sized hail (0.5 cm to 1.2 cm), "2" grape-sized hail (1.3 cm to 1.9 cm), "3' walnut-sized hail (2.0 cm to 3.2 cm), "4" golfball-sized hail (3.3 cm to 5.2 cm) and "5" greater than golfball-sized hail (> 5.2 cm).

6. VALIDATION OF HAILCAST

6.1 Forecasting of hail day or non-hail day within the ASA

In this section we quantify the forecast skill of HAILCAST to predict the occurrence of hail anywhere within the ASA. In other words, we test whether HAILCAST can predict accurately a hail day versus a non-hail day. Given the problems associated with the surface report database discussed in section 4.1, hail forecasts were verified using maximum VIL maps from the Olds-Didsbury radar.

Smith and Yau (1993) confirmed earlier studies which suggested that hailstorms typically develop over central Alberta when there is a "favourable" synoptic-scale height pattern at 500 mb. Specifically, Smith and Yau's study showed that hailstorm activity is usually associated with a 500 mb trough over or upstream of Alberta. Based on Smith and Yau's findings, we examined the 12 and 00 UTC 500-hPa upper-air analysis charts for each day to classify the upper-air flow over and upstream of the ASA (a region bounded by 45 ° N to 55 ° N and 110 ° W to 125 ° W). Particular attention was paid to the direction of the flow at 500 hPa and changes in 500-hPa heights between 12 UTC and 00 UTC. Each of the 92 days was then categorized as being either favourable for hailstorms (i.e., 500 mb trough or perturbation) or unfavourable for hailstorms (i.e., 500 mb ridge).

Careful analysis of upper-air charts revealed that the upper-air flow favoured the formation of hailstorms on 44 days during the study period. Comparison of the upper-air charts and maximum VIL maps for the summer of 2000 suggest that there was a positive correlation between hail activity over the ASA and the presence of an upper-air trough. In particular, of the 45 hail days observed by radar over the ASA, 71% were associated with an upper-air trough. By comparison, hail was observed by radar on almost 73% of the days when an upper-air trough was present. Radar detected severe hailswaths that were more than 50 km in length on 11 days, and on 8 (73%) of these days an upper-air trough was present. In contrast, only 10% of the hail days and 15% of the severe hail days were observed when the flow was deemed unfavourable for hailstorms.

Before discussing the performance statistics, it must be kept in mind that HAILCAST was not designed to explicitly predict the initiation of convection. Rather, the model provides an estimate of the maximum expected hail size given that convection is triggered. Notwithstanding this caveat, the model can still potentially improve lead times of warnings because forecasters can be proactive when issuing warnings on those occasions when storms that are moving into an area where the potential for large and damaging hail exists as indicated by HAILCAST.

The skill of HAILCAST to predict the occurrence of a hail day or a non-hail day within the ASA was quantified using the Probability of Detection (POD), False Alarm Ratio (FAR) and the Critical Success Index (CSI) and the Heidke Skill Score (HSS). These skill scores are commonly used (e.g., Marzban and Stumpf 1998). The disadvantage of using this approach to verify the forecast skill is that the model could forecast hail up to 250 km from the location where hail was observed and still score a hit.

Table 1 shows the performance statistics calculated for predicting hail anywhere over the ASA when the upper-air circulation was favourable for hailstorms. The hail forecasting technique scored a POD of 0.90, a FAR of 0.26, a CSI of 0.68 and a HSS of 0.16. Note that, by adopting this approach, the POD reflects the number of hits on days when the upper-air circulation pattern favoured organized storms and, consequently, hail days observed when the flow was unfavourable (13) would have been missed. The POD for forecasts of severe hail when an upper-air trough was present was significantly lower than that calculated for all hail days (0.75 versus 0.90), while the FAR was somewhat higher (0.32 versus 0.26). Consequently, the overall forecast skill for severe hail days was lower than for all hail days, with a CSI of 0.56. In contrast, the HSS for severe hail days was markedly higher (0.46 versus 0.16). The superior performance suggested by the HSS, could be attributed to the fact that the CSI does not give credit for null forecasts, and almost 40% of the forecasts for severe hail were null forecasts, compared to only 7% of those forecasts for hail of any size. This suggests a tendency for the hail forecasting technique to overpredict the occurrence of hail on trough days.

We also calculated the skill scores using all days (i.e. both synoptically favourable and unfavourable 500 mb flow patterns). This resulted in the FAR increasing between 13% and 19%, with only a slight increase (about 3%) in the POD.

Table 1: Skill scores calculated for predicting the occurrence of hail or severe hail within the ASA for the summer of 2000. Values under column "T" were calculated only using the model guidance on days when there was an upper-air trough upwind or over the ASA, scores under column "A" are the skill scores calculated for all days in the dataset, and scores under "L" were calculated using model guidance on days when lightning was observed over the ASA.

	All hail days (%)			Severe hail days (%)		
	T	**A**	**L**	**T**	**A**	**L**
POD	90	93	93	75	79	79
FAR	26	45	23	32	45	27
CSI	68	53	73	56	48	61
HSS	16	22	0	46	45	50

This was to be expected given that 71% of all the false alarms for hail were observed on days when the upper-air circulation was classified as unfavourable for hailstorms. Accordingly, the overall performance of the forecast technique (as quantified by the CSI) was reduced between 8% and 15% when the forecasts for all days were considered. By comparison, the HSS for forecasts on all hail days increased by 6%. This increase in the HSS could be ascribed to the fact that about 15% of the forecasts made for all days in the dataset were null forecasts, compared to only 7% of the days when an upper-air trough was present. The HSS of 0.45 obtained for the severe hail forecasts was almost the same as that achieved when forecasts were only considered for trough days.

6.2 Forecasting the spatial distribution of hail within the ASA

The second, more stringent, verification method was similar to that developed by Weiss et al. (1980) for the purpose of verifying SPC outlooks and watches. Weiss et al. addressed the problem of evaluating the skill of large spatial forecasts (approximately 100 000 km^2) using a low density event verification dataset. They noted that, while the POD is not affected by the size of the verification area, the FAR tends to increase as the verification area increases. For example, they found that FARs of 0.9 were typically achieved when they evaluated severe weather watches having an average area of about 100 000 km^2. The reason for this is that although the atmosphere can be unstable over very large areas, convection typically occurs only over a fraction of that area—this is especially true for severe convection. To try and address this problem, Weiss et al. recommended equating each surface report of severe weather with a much larger area. Their rationale for adopting this approach is that SPC forecasters do not anticipate severe weather over the entire severe weather watch area. Likewise, if HAILCAST predicts hail over the entire ASA, we do not expect that hailstorms will cover the entire area. The verification methodology developed by Weiss et al. was also designed to take into account the spatial distribution of the surface reports so as to mitigate biases that event clustering may have introduced into the verification statistics. We adopted the method of Weiss et al. as follows:

1. The ASA was divided into 25 blocks. Each block measured 0.5º by 0.5º (~1900 km^2). A block was counted as a forecast hail (severe hail) block if hail (severe hail) was forecast over at least half the area of the block. If the maximum VIL maps indicated hail/severe hail in a given block, then that block was tagged as a hail/severe hail block.
2. If hail was forecast within one block of an observed hail block the forecast was considered a hit. On most occasions, this criterion required the model to forecast hail within 35 km to 55 km of the observed hailswath; the furthest distance that the forecasted hail contour could be from a hailswath and still be considered a hit was approximately 90 km. The exact distance depended on the location of the hailswath relative to the forecast hail contour. The POD was calculated by dividing the number of observed hail blocks (along hailswaths) that occurred within one block of the forecast hail blocks by the total number of observed hail blocks.
3. The first step in calculating the FAR was to calculate the "good percentage area" (Weiss et al. 1980). This parameter represents the fraction of the area forecast to receive hail that was actually affected by hail. To this end, the sum of all those forecast hail blocks that bordered the observed hail blocks was calculated. The "good percentage area" was then calculated by dividing this total by the sum of all of the blocks predicted to receive hail. Finally, the overall skill of the forecasts was calculated using the Critical Success Index (CSI).

As mentioned previously, this verification technique applies spatial criteria on the forecast hail maps. In particular, the model is penalized if hail is forecast over a large area and hail occurred only over an isolated area.

Table 2: Same as for Table 1, except skill scores calculated for predicting the occurrence and location of hail/severe hail over the ASA for the summer of 2000.

	All hail days (%)			Severe hail days (%)		
	T	**A**	**L**	**T**	**A**	**L**
POD	84	87	87	60	66	66
FAR	41	54	47	53	57	49
CSI	54	43	49	36	35	41

Considering model forecasts only on days when the upper-air circulation was favourable, Table 2 shows that the POD over the ASA for all hail blocks was almost 0.84, with a FAR of 0.41 and a CSI of 0.54. A FAR of 0.41 indicates that 59% of the blocks forecast to receive hail were within one block of the observed hailswaths. On the severe hail days, the POD was significantly lower at 0.60, with a FAR of 0.53; the lower POD was primarily responsible for the inferior CSI of 0.36. A possible explanation for the higher FAR for the forecasts on severe hail days is that, on average, the severe hailswaths (observed by radar) covered a small area of the ASA (17%), while the model forecast severe hail over a relatively large area (29%). The small areal coverage of severe hailswaths would also make it more difficult to correctly forecast their position, which would in turn reduce the likelihood of a hit. A negative bias of almost 1 cm when predicting severe hail (see section 6.3) could also, in part, have contributed to a lower POD.

Table 2 indicates that including forecasts for all days in the calculation of verification statistics increased the FAR, while only resulting in a small increase in the POD. This was especially noticeable for the forecasts on all days, with the FAR increasing by 13% and the POD increasing by only 3%. In contrast, verifying forecasts on all days resulted in only a slight decrease in the skill of the severe hail forecasts, with the 4% increase in the FAR offset by a 6% increase in the POD.

In summary, over the ASA, HAILCAST has greater skill at forecasting the occurrence of a hail day or a no-hail day than at predicting the spatial distribution of the hailstorms. Generally, applying the spatial and areal constraints on the hail forecasts lowered the forecast skill by reducing the number of hits and by increasing the number of false alarms. Caution must be exercised in directly comparing the performance statistics calculated using the two verification techniques, as the methodologies used to calculate the skill scores are different.

6.3 Forecasting the maximum hail diameter

The forecast hail diameters were obtained by running HAILCAST using prognostic GRIB soundings interpolated to the latitude and longitude of the hail reports. To allow for uncertainty in the time and location of the hail reports, tight gradients in the forecast hail maps, and timing errors in the GEM model, HAILCAST was also run using a prognostic sounding valid for 21 UTC. The 21 UTC soundings were constructed by linearly interpolating the forecast data (at each level) valid for 18 and 00 UTC. The largest of the 21 and 00 UTC forecast hail sizes was then used to calculate the forecast hail-size category. Errors and hail-size category statistics were only calculated for those reports when the hail model forecast hail of at least 5 mm in diameter. In other words, we wished to establish how accurate the forecast hail sizes were, given the condition that a forecaster was expecting hail. By applying this criterion, 423 (almost 80%) of the 533 surface hail reports in the dataset were used to verify the accuracy of the forecast hail sizes. The mean errors and absolute errors were calculated by subtracting the observed hail diameter from the forecast hail diameter on the ground. Hail reports were allocated a representative diameter if no specific diameter measurement was provided; otherwise the reported diameter was used to calculate the hail size error.

Table 3: Summary of performance statistics calculated to quantify the accuracy of the hail-size category forecasts for hail/severe hail days over the ASA and for all hail reports on the Canadian prairies in 2000.

	ASA		All reports	
	All hail	Severe	All hail	Severe
Hail-size category (%)				
Correct	30	39	29	33
Within one	76	76	73	71
One too small	23	33	22	27
≥ two too small	8	24	12	26
One too large	23	4	22	11
≥ two too large	16	0	15	4
Hail-size error (cm)				
Mean error	0.3	-0.9	0.1	-0.7
Mean absolute error	1.2	1.2	1.3	1.5

Table 3 shows that, over the ASA, 30% of the categorical hail-size forecasts were correct and 76% were correct to within one size category. Very few of the forecasts were out by two or more categories. About 80% of the hail-size forecasts were within 2 cm of the observed diameters, while only approximately 5% of the errors were greater than 3 cm. The mean absolute error for the hail-size forecasts was 1.2 cm, with a positive bias of 0.3 cm. The accuracy of the hail-size forecasts for all the prairies was comparable, albeit slightly inferior.

The hail-size forecasts for severe hail days were comparable with those for all hail days, with 76% correct to within one hail-size category. Almost 40% of the hail-category forecasts were correct. There was a tendency for the forecasts to underestimate the hail category for severe events, with 57% of the forecasts one or more categories too small. About 25% of the forecasts were two or more categories too small. This tendency to underestimate the hail size category for severe events is reflected in the negative bias of -0.9 cm. Jewell and Brimelow (2004) noted a similar negative bias in hail-size forecasts produced using HAILCAST and proximity soundings for 392 severe hail events over the contiguous U.S. between 1997 and 2002.

Forecasts of hail size for severe events across all three provinces were slightly less accurate. The close correspondence between the hail-size category and error statistics for hail forecasts for events over the ASA and those for all three prairie provinces suggests that the hail forecasts did not display a marked regional bias.

7. FORECASTING HAIL USING HAILCAST COMBINED WITH LIGHTNING OBSERVATIONS

While the upper-air criterion is useful for excluding days when the likelihood of organized convection is low, hail (severe hail) was observed on 13 (9) days when the flow was considered unfavourable for organized convection according to the conceptual model of Smith and Yau (1993). Thus, an alternative approach is required to assist the forecaster in deciding whether thunderstorms will develop over the area where HAILCAST is predicting hail. One such approach would be to objectively identify those locations where there is a high probability of lightning. Burrows et al. (2005) found that forecast lightning algorithms using prognostic GEM model data showed skill at predicting areas of lightning activity over Alberta.

As a first step in determining the feasibility of utilizing such an approach to assist forecasters in interpreting the forecast hail maps, we investigated whether a correct forecast of thunderstorms anywhere over the ASA would improve the performance statistics for the summer of 2000. Of course, forecasters do not know *a priori* whether thunderstorms will develop or not, but for the purpose of this exercise, we will assume that forecasters could correctly predict the occurrence of thunderstorms over the ASA.

Daily lightning maps were used to identify where convection occurred within the ASA for each day between 1 June and 31 August 2000. A day was classified as a thunderstorm day if at least two lightning strikes were recorded within the ASA between 20 and 05 UTC. Tables 1 and 2 indicate that by evaluating the model output only on days when thunderstorms were observed over the ASA, we were able to achieve high CSI scores by maintaining a low FAR. In particular, considering the performance for predicting the occurrence of hail anywhere over the ASA, the FAR was reduced by 20% compared to forecasts evaluated for all days in the dataset. Similarly, for severe hail days the FAR was reduced by 18%. This marked lowering in the FAR led to a significant increase in the CSI scores on all hail days and severe hail days, with CSI scores of 0.73 and 0.61, respectively. Verification of the forecast locations using the lightning criterion for all hail and severe hail days improved the forecast skill by reducing the number of false alarms by 7% to 8%.

According to Tables 1 and 2, the highest CSI scores for predicting the occurrence of hail anywhere over the ASA were obtained using the lightning criterion. If one is also concerned about predicting the location of the hail, the trough criterion produced the highest CSI scores for this verification method. The highest CSI scores for severe hail forecasts were obtained using the lightning criterion for both verification methods. These statistics suggest that if one could accurately predict the occurrence of thunderstorms over a specified threat area, then the forecasts produced by HAILCAST over that region would generally provide useful guidance regarding the occurrence and size of the hail.

8. CONCLUSIONS

The objective of this paper was to determine the feasibility of producing spatial maps of the forecast maximum hail size over the Canadian prairie provinces using prognostic GEM model soundings as input for the HAILCAST model. Specifically, the hail forecasts were based on 12-h forecast soundings valid for 00 UTC each day between 1 June and 31 August 2000. The forecasts of maximum hail size were verified against radar reflectivity data and 533 surface reports of maximum hail size. The verification of the forecasts focused on the Alberta Study Area (ASA), a region of about 48 000 km^2 located over south-central Alberta.

This study has revealed that it is feasible to use prognostic GEM soundings as input data for HAILCAST, making this technique ideal for use in an operational setting. The comparison between radar-derived VIL data and HAILCAST forecasts showed the following:

1) HAILCAST is skillful in identifying a hail day versus a non-hail day over the ASA up to 12 hours in advance.
2) HAILCAST shows skill at distinguishing between non-severe and severe hail events.
3) HAILCAST has limited skill in predicting the distribution of hail on spatial scales less than approximately 60 km, but is skillful at predicting the main threat areas.
4) HAILCAST is skillful in predicting the maximum hail diameter.
5) Synoptic-scale constraints (such as the presence of a 500-mb trough) are needed to establish whether it is appropriate to run HAILCAST on a given day over Alberta.

Work is underway to correctly identify which GEM model field is most appropriate to identify in advance whether there will be deep convection and at which model grid points HAILCAST should be run. We also plan to determine whether information from the ensemble runs could offer further insight to forecasters. For example, a probabilistic approach may assist in predicting the location and size of hail, with maps showing the probability of hail diameter exceeding certain thresholds, as well as the standard deviation of the forecast hail sizes. This in turn could be useful for generating public forecasts of the probability of hail, a practice that is not currently done in Canada. There are potentially significant economic gains to be realized by any method that can improve the accuracy of public forecasts of severe hail. For example, farmers could move equipment and small livestock indoors, aircraft at aerodromes can be moved indoors, while the public could move cars and other valuable items located outdoors under cover.

ACKNOWLEDGEMENTS

This research was supported by grants from the Canadian Foundation for Climate and Atmospheric Sciences (CFCAS), the Natural Sciences and Engineering Research Council (NSERC) and Environment Canada. The authors wish to express their gratitude to the following individuals and organizations for their assistance in undertaking this research: Brad Shannon, Gary Burke, Neil Taylor, Patrick McCarthy, Alberta Severe Weather Management Society, Alberta Agriculture Financial Services Corporation.

REFERENCES

Baldwin, M.E., J.S. Kain, and M.P. Kay, 2002: Properties of the convection scheme in NCEP's Eta model that affect forecast sounding interpretation. *Wea. Forecasting*, **17**, 1063–1079.

Barge, B. L., 1974: Polarization measurements of precipitation backscatter in Alberta. *J. Rech. Atmos.,* **8,** 163-173.

Betts, A. K., 1986: A new convective adjustment scheme. Part I: Observational and theoretical basis. *Quart. J. Roy. Meteor. Soc.,* **112**, 677–692.

———, and M.J. Miller, 1986: A new convective adjustment scheme. Part II: Single column tests using GATE wave, BOMEX, ATEX and arctic air-mass data sets. *Quart. J. Roy. Meteor. Soc.,* **112**, 693–709.

Brimelow, J.C., 1999: Numerical modelling of hailstone growth in Alberta storms. M.S. thesis, Dept. of Earth and Atmospheric Sciences, University of Alberta, 153 pp.

———, G.W. Reuter, and E.P. Poolman, 2002a: Modeling maximum hail size in Alberta thunderstorms. *Wea. Forecasting*, **17**, 1048-1062.

———, T.W. Krauss, and G.W. Reuter, 2002b: Operational forecasts of maximum hailstone diameter in Mendoza, Argentina. *J. Wea. Mod*., **34**, 8-17.

———, G.W. Reuter, A. Bellon, and D. Hudak, 2004: A radar-based methodology for preparing a severe thunderstorm climatology in central Alberta. *Atmos.-Ocean*, **42**, 13-22.

Brooks, H.E., C.A. Doswell III, and R.A. Maddox, 1992: On the use of mesoscale and cloud-scale models in operational forecasting. *Wea. Forecasting*, **7**, 120-132.

———, ———, and L.J. Wicker, 1993: STORMTIPE: A forecasting experiment using a three-dimensional cloud model. *Wea. Forecasting*, **8**, 352-363.

———, ———, and J. Cooper, 1994: On the environments of tornadic and nontornadic mesocyclones. *Wea. Forecasting*, **9**, 606-618.

Burrows, W.R., C. Price, and L.J. Wilson, 2005: Warm season lightning probability prediction for Canada and the northern United States. *Wea. Forecasting* (in review).

Changnon, S.A., 1992: Temporal and spatial relations between hail and lightning**.** *J. Appl. Meteor.,* 31*,* 587–604

———, S.A., R.A. Pielke Jr., D. Changnon, R.T. Sylves, and R. Pulwarty, 2000: Human factors explain the increased losses from weather and climate extremes. *Bull. Amer. Meteor. Soc*., **81**, 437–442.

Charlton, R. B., B. M. Kachman, and L. Wojtiw, 1995: Urban hailstorms: A view from Alberta. *Natural Hazards*, **12**, 29-75.

Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998: The operational CMC–MRB global environmental multiscale (GEM) model. Part I: Design considerations and formulation. *Mon. Wea. Rev.,* **126**, 1373–1395.

Crook, N.A., 1996: Sensitivity of moist convection forced by boundary layer processes to low-level thermodynamic fields. *Mon. Wea.. Rev*., **124**, 1767-1785.

Dixon, M., and G. Wiener, 1993: TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting—A radar-based methodology. *J. Atmos. Oceanic Technol.*, **10**, 785–797.

Doswell, C.A. III, J.T. Schaefer, D.W. McCann, T.W. Schlatter, and H.B. Wobus, 1982: Thermodynamic analysis procedures at the National Severe Storms Forecast Center. Preprints, *9th Conf. Weather Forecasting and Analysis*, Seattle, WA, Amer. Meteor. Soc., 304-309.

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. Forecasting*, **13**, 277-285.

Etkin, D., and S.E. Brun, 1999: A note on Canada's hail climatology: 1977-1993. *Int. J. Climatol*., **19,** 1357-1373.

Foster, D.S., and F.C. Bates, 1956: A hail size forecasting technique. *Bull. Amer. Meteor. Soc.,* **35**, 135-140.

Grell, G.A, 1993: Prognostic evaluation of assumptions used by cumulus parameterizations**.** *Mon. Wea. Rev.,* **121**, 764–787.

Hamill, T.M., and A.T. Church, 2000: Conditional probabilities of significant tornadoes from RUC-2 forecasts. *Wea. Forecasting*, **15**, 461–475.

Hanna, S. R., and Yang R., 2001: Evaluations of mesoscale models' simulations of near-surface winds, temperature gradients, and mixing depths. *J. Appl. Meteor.,* **40**, 1095–1104.

Hart, R.E., G.S. Forbes, and R.H. Grumm, 1998: The use of hourly model-generated soundings to forecast mesoscale phenomena. Part I: Initial assessment in forecasting warm-season phenomena. *Wea. Forecasting*, **13**, 1165–1185.

Janjić, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.,* **122**, 927–945.

Jewell, R., and J.C. Brimelow, 2004: Evaluation of an Alberta hail growth model using severe hail proximity soundings in the United States. Preprints, *22nd Conf. on Severe Local Storms,* Hyannis, MA, Amer. Meteor. Soc., CD-ROM,

Joe, P., D. Burgess, R. Potts, T. Keenan, G. Stumpf, and A. Treloar, 2004: The S2K severe weather detection algorithms and their performance. *Wea. Forecasting*, **19**, 43–63.

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

Lenning, E., H.E. Fuelberg, and A.I. Watson, 1998: An evaluation of WSR-88D severe hail algorithms along the northeastern Gulf Coast. *Wea. Forecasting*, **13**, 1029–1045.

Marzban, C., and G.J. Stumpf, 1998: A neural network for damaging wind prediction. *Wea. Forecasting*, **13**, 151-163.

———, and A. Witt, 2001: A Bayesian neural network for severe-hail size prediction. *Wea. Forecasting,* **16**, 600–610.

Moore, J.T., and J.P. Pino, 1990: An interactive method for estimating maximum hailstone size from forecast soundings. *Wea. Forecasting*, **5**, 508-526.

Mueller, C.K., J.W. Wilson, and N.A. Crook, 1993: The utility of sounding and mesonet data to nowcast thunderstorm initiation. *Wea. Forecasting*, **8**, 132-146.

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

Renick, J.H., and J.B. Maxwell, 1977: Forecasting hailfall in Alberta. *Hail: A Review of Hail Science and Hail Suppression, Meteor. Monogr.,* No. 38, Amer. Meteor. Soc., 145-151.

Smith, S.B., and M.K. Yau, 1993: The causes of severe convective outbreaks in Alberta. Part II: Conceptual model and statistical analysis. *Mon. Wea. Rev*., **109**, 1126-1134.

Stensrud, D. J., J.-W Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensembles of mesoscale convective systems. *Mon. Wea. Rev.,* **128**, 2077–2107.

Thompson, R.L., 1998: Eta model storm-relative winds associated with tornadic and nontornadic supercells. *Wea. Forecasting*, **13**, 125–137.

———, R. Edwards, J.A. Hart, K.L. Elmore, and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the Rapid Update Cycle. *Wea. Forecasting*, **18**, 1243–1261.

Wagenmaker, R.B., 1992: Operational detection of hail by radar using heights of VIP-5 reflectivity echoes. *Natl. Wea. Dig*., **17**, 2-5.

Weiss, S.J., D.L. Kelly, and J.T. Schaefer, 1980: New objective verification techniques at the National Severe Storms Forecast Center. Preprints, *8th Conf. on Weather Forecasting and Analysis*, Denver, CO, Amer. Meteor. Soc., 412-419.

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. Forecasting*, **13**, 286-303.

Wojtiw, L., 1977: Climatology of hailstorms in central Alberta. *The Albertan Geographer,* **13**, 15-30.