Ryan Jewell *
Storm Prediction Center. Norman OK

Julian Brimelow University of Alberta, Edmonton

1. INTRODUCTION

Overall, hail poses a low threat to life compared to other types of severe convective weather like tornadoes. Although giant hailstones can cause injury, their greatest impact is the damage that they cause to property, crops and livestock.

Forecasters currently have very few tools at their disposal to aid them in predicting the maximum expected hail size. Moreover, the few tools that are available frequently forecast unrealistically large hail sizes in high CAPE environments. Perhaps the main reason for this is that the hail growth process is very complex and consequently, any ingredients-based forecasting methods that are typically used to forecast severe weather will usually result in poor hail size forecasts, with little if any skill (Doswell et al. 1992). The complexity of hail growth is handled best by three-dimensional (3D) cloud models (e.g., Xu 1983). However, 3D models are cumbersome and impractical to run in an operational setting (Brooks et al. 1992). A simple one-dimensional model, called HAILCAST, has been developed to predict the maximum expected hail diameter (D) at the surface (Brimelow et al. 2002a). HAILCAST has been implemented and tested at the Storm Prediction Center (SPC) over the past two years. simplicity and computational efficiency HAILCAST makes it a practical tool for predicting the maximum hail size.

This study was conducted from the perspective of an operational SPC forecaster. The forecast question that we addressed in this study is not whether hail will fall on a particular day but rather, given a positive forecast of severe hail (D > 0.75"), how large will the hail be? Specifically, should an outlook for significant hail (SIG; D \geq 2") be made?

Corresponding Author Address: Ryan Jewell, 1313 Halley Circle, Norman, OK 73069; email: Ryan.Jewell@noaa.gov

To this end we also tested the model's overall skill at forecasting hail size using a variety of proximity soundings, and tested the model's ability to successfully delineate between non-significant severe hail (NON-SIG; $0.75 \le D < 2$ ") and SIG hail environments. What follows is a brief description of the hail model and the database, followed by a description of the methodology used to verify the model forecasts. Finally, some preliminary test results are presented.

2. HAILCAST DESCRIPTION

HAILCAST is a one-dimensional coupled cloud and hail growth model, developed initially by Poolman (1992) and then improved upon by Brimelow et al. (2002a). Using an atmospheric sounding as input, the model produces an ensemble of updrafts based on permutations of the control surface temperature (T) and dewpoint (Td). A set of 25 ensemble members is produced by varying both T and Td between +1°C and -1 °C from their control values in increments of 0.5 °C. The model also uses the product of the surfacebased CAPE and 850 hPa to 6 km wind shear. referred to as the Energy Shear Index (ESI; Brimelow et al. 2002a), as a means to categorize storm type and to regulate the updraft properties. Depending upon the magnitude of the ESI, varying degrees of lateral and cloud top entrainment are applied to the updraft. Larger instability and shear will result in minimal entrainment and thus updraft speeds approaching the theoretical buoyancyderived values associated with supercells (Bluestein et al. 1988). ESI also regulates the updraft duration, with a maximum lifetime of 60 min. After a 300-μm embryo is introduced at cloud base, the hail model allows the embryo to ascend within the model-derived updraft and grow until either the updraft collapses or a hailstone reaches the surface. For a more detailed description of HAILCAST see Brimelow et al. (2002a).

3. THE HAIL REPORT DATABASE

HAILCAST has been evaluated in Alberta using a very high quality report database comprised of direct measurements of hail size from a high density hail observer network (Brimelow et al. 2002a). The quality of the severe report database in the U.S. is much less reliable, with the vast majority of hail sizes being estimated rather than measured. Therefore, caution must be exercised when using this database for verification purposes.

Analysis of hail databases indicates that large hail is observed much less frequently than small hail. This is because many strong thunderstorms can produce small hail, while it takes a much more particular set of mesoscale and stormscale conditions to produce large hail. Figure 1 shows the distribution of reported hail sizes in the U.S. between 1995 and 2002. Reports less than 1" in diameter (numbering 42,368) are not included. The most commonly reported diameter is 0.75", the threshold for severe hail as defined by the National Weather Service (NWS), averaging around 4000 reports per year. The combination of 0.75" to 1.00" hail accounts for over 7000 reports per year. In contrast, an average of only 183 reports of baseball-size hail (D = 2.75") were received each year.

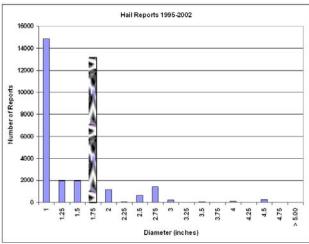


Figure 1. U.S. Hail report distribution of 1" or greater from 1995-2002 with golf ball-size hail reports highlighted. The combination of 0.75" and 0.88" reports totaled 42,368 (not shown).

The hail size distribution in Fig. 1 is not consistent with field observations of natural hail size distributions (Wong et al. 1988). Specifically, it shows that between 1995 and 2002 in the U.S. golf ball-size hail was reported almost as frequently as 1" hailstones. For the period 1955-

2002, reports of golf ball-size hail actually exceeded those of 1" hail and were the second most frequent size reported (Schaefer et al. 2004). Personal communication with storm chasers and people living near the climatological world maximum of giant hail frequency (central Oklahoma) reveal that golf ball-size hail is rarely observed when compared to the occurrence of quarter-size or smaller hail.

There is also reason to question 4.5" hail reports, the default size used by the NWS to represent softball-size hail. The first ever softball tournament at the 1933 Worlds Fair in Chicago used a 4.5" diameter softball (Amateur Softball Association However, the most common men's 2004). (women's) league softball in use today has a diameter of 3.8" (3.5"). Consequently, people are likely to associate a softball-size hailstone with a 3.8" sphere, such that reports of 4.5" softball-size database are hail in the likelv overestimations of the true size of the hail. Further, a 4.5" diameter stone has 63% more mass when compared to a stone the size of a common men's league softball (3.8").Consequently, the updraft required to produce a 4.5" stone must be significantly stronger than that required to support a 3.8" hailstone. To put this into perspective, the record 1970 Coffeyville, KS stone had a mass near 750 g, while a solid 4.5" sphere of ice" would have a mass of 720 g. It is unlikely near-world record stones fall with such frequency as the report database would have us believe. In order to increase the integrity of the hail report database, it can be argued that the NWS default softball hail diameter should be changed from 4.5" to 3.65", the average of modern men's and women's softballs.

4. DATA AND METHODOLOGY

A database of severe hail proximity soundings from the contiguous U.S. was constructed for the period January 1997 to August 2002 for the purpose of evaluating the HAILCAST model. It is important to note that soundings supporting surface-based convection were included in the dataset because, at the time of this study. HAILCAST was unable to simulate elevated convection (storms rooted above a surface stable layer). SVRPLOT (Hart 2003) was used to sort and plot hail reports. Hail reports were only included if they were observed within 100 nm of an upper-air site, and occurred between 21 and 02 UTC (+/- 2.5 hours from 2330Z). Reasons for using 2330Z as the representative sounding release time include: 1) It takes time for the

balloon to reach the mid and upper levels of the atmosphere, and 2) the possibility of late or multiple releases.

Hail reports were then categorized as either SIG or NON-SIG according to the size criteria specified in section 1. For all cases, the largest hail report for a given location was recorded. Care was taken when choosing the NON-SIG events, such that there were no SIG hail reports within at least 300 nm of the NON-SIG reports. The majority of NON-SIG hail events were on days when no SIG hail was reported, however.

Soundings were subjectively excluded if there was obvious convective contamination, including saturated profiles, suspect winds, or missing data (Fig. 2). Archived surface plots were used to verify whether or not the sounding was released in the same low-level air mass that was ingested by the hailstorm. In just a handful of SIG cases, the low levels of the soundings were impacted by either a change in air mass (e.g. behind a cold front) or rain-cooled outflow air, but were otherwise uncontaminated.

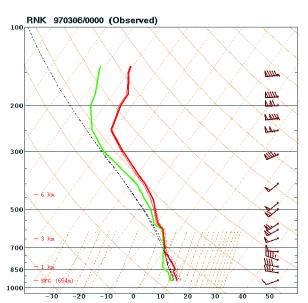


Figure 2. Example of a contaminated sounding. Degree of saturation suggests balloon likely entered convective cloud. Also note abnormal lapse rate between 700 and 500 hPa.

The sounding was included as long as it was uncontaminated above the modified LCL and the ESI was over 5 (highest category). HAILCAST does not use winds below 850 hPa, so erroneous or unrepresentative winds below this level are irrelevant to the model calculations. Because surface observations are used to modify the

boundary layer, rain-cooled air below the LCL is also irrelevant. An ESI over 5 ensures that small variations in low level shear caused by outflow contamination will not change updraft properties within the model. Figures 3 and 4 provide an example of how the soundings were modified by the model, based on input T and Td. After applying all the selection criteria, 382 proximity soundings were included in the dataset (Fig. 5). Of these, 203 are associated with SIG hail and 179 with NON-SIG hail.

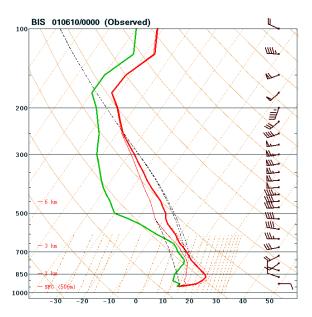


Figure 3. Example of low-level outflow contamination.

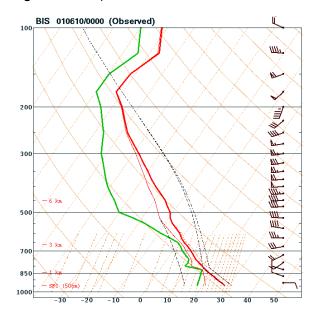


Figure 4. As in Figure 3, but after modification using input surface T and Td for hail model.



Figure 5. Locations and number of soundings used for both SIG and NON-SIG events. Database includes soundings from every month.

5. SELECTING THE CONTROL SURFACE T AND TD

As with any hail forecasting technique, the maximum hail size is highly dependent on the updraft properties, and thus the parcel properties used to calculate the updraft. Therefore, it was critical in this study to employ a robust method for selecting representative surface T and Td values.

Selecting a representative control surface T and Td was subjective, as is the case in an operational setting. While it is nearly impossible to determine the exact properties of the air feeding a storm, a realistic approach is to first estimate an upper limit. The rationale for establishing upper limits for T and Td was as follows. Given that the majority of hailstorms are observed at the time of day when the near surface lapse rate is superadiabatic, it is likely that the highest surface temperature observed between 21-02 UTC is an overestimate of the actual average potential temperature within the boundary layer. Therefore, using the warmest T will set an upper limit to the boundary layer potential temperature. The same logic applies to determining the dewpoint. Dewpoints typically decrease with height in the boundary layer, so the highest dewpoint found at the surface will often overestimate the average mixing ratio below cloud base.

Using surface observations, the following guidelines were used to find a representative T and Td.

- The highest T (T_{max}) and Td (Td_{max}) observed between hail report time and 2.5 hours prior to report time were recorded.
- These did not have to occur at the same observation point, but had to be within the same low-level air mass feeding the storm i.e., on the same side of any moisture or temperature gradients.
- Observations upwind of the report (with respect to the low-level flow) were preferred. An exception to this rule was if these observations were contaminated by precipitation, or if surface winds were light and variable.

HAILCAST takes the input (control) surface T and Td, and then varies them by 1 $^{\circ}$ C either side of the control values to construct an ensemble forecast. If one were to enter the maximum observed T and Td into the model, overly buoyant ensemble members would result. Updraft velocity, liquid water content, and thus hail growth rates might then be overestimated. Therefore, the control values were taken to be T_{max} and Td_{max} , *minus one degree Celsius*. In this way, the most unstable ensemble members will not exceed the probable observed upper limits.

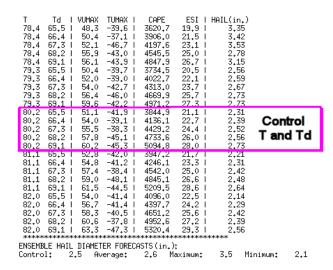


Table 1. Hailcast output of 25 ensemble members.

The above method of selecting the input T and Td, in combination with the HAILCAST ensemble process, effectively accounts for the effects of vertical mixing and inhomogeneous T and Td fields in the vicinity of the storms. It has been shown by Craven et al (2001) that using a mixed layer T and Td when computing the updraft properties is typically more appropriate than using

a surface-based parcel in which skin layers of moisture could be unrepresentative, and thus give overestimations of the realized instability. Therefore, using T_{max} minus 1 C and Td_{max} minus 1 C as control values will indirectly account for the effects of mixing.

6. MODEL TEST RESULTS

Figures 6 and 7 show scatter plots of the reported maximum hail size versus the maximum ensemble member size and ensemble average size, respectively. It is evident from the figures that the maximum ensemble member tends to slightly overestimate hail size, while the ensemble average has a strong tendency to underestimate the size. In the following scatter plots, note how the clusters of 4.5" reports appear to be outliers, as a result of the overestimation of the diameter of softball-size hail described previously.

Brimelow et al. (2002b) found that the average of all 25 ensemble members results in the best correlation between the forecasted and observed hail size. The current research corroborates this finding, with the correlation coefficient (r) for the maximum ensemble member equal to 0.61 (Fig. 6), versus 0.64 (Fig. 7) for the ensemble average. As expected, the single maximum member comparison provides less reliable results. This is probably caused by the use of only a single combination of T and Td, which increases the likelihood of selecting an unrepresentative updraft Also, since the ensemble members parcel. sometimes cross ESI category lines, this will cause the updraft properties of some members to be significantly changed. The average of all 25 members appears to reduce these effects.

Removing the suspect golf ball-size hail reports removes much of the variability and increases r for the ensemble average from 0.64 to 0.67, and from 0.61 to 0.63 for ensemble maximum. Substituting 3.65" for all 4.5" softball reports increases r further to 0.70 for ensemble average and to 0.66 for ensemble maximum size.

Overall, these skill numbers are quite respectable given the quality of the report database, the proximity of the soundings and complexity of the hail process. They indicate HAILCAST possesses promising skill at forecasting the maximum expected hail size.

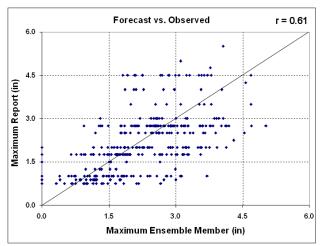


Figure 6. Observed diameter versus ensemble maximum diameter.

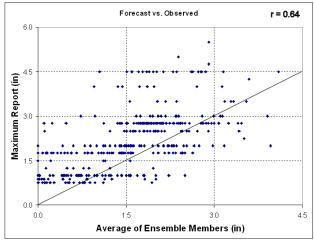


Figure 7. Observed diameter vs. ensemble average diameter.

7. SIG HAIL VERSUS NON-SIG HAIL PERFORMANCE STATISTICS

One of the duties of the SPC is to forecast the probability of SIG hail in the convective outlooks. In order to test the model's ability to discriminate between SIG and NON-SIG environments, the dataset was divided into two subsets: D ≥ 2" and $0.75 \le D \le 1.75$ " (there were no reports between 1.76-1.99" in the database). Figure 8 shows a box and whiskers plot of the hail model ensemble average output distribution for both subsets. The boxes denote the range of the 25th to 75th percentiles, and the tips of the whiskers denote the 10th and 90th percentiles. This plot shows substantial separation in hail model output when initialized with SIG and NON-SIG hail soundings. The 75th percentile of the NON-SIG hail cases barely overlaps the 10th percentile of SIG hail cases. These data suggest that the model is capable of successfully discriminating between

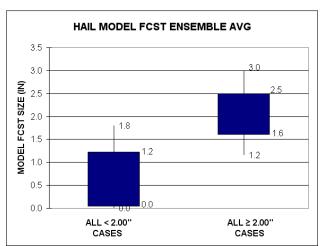


Figure 8. Box and whiskers plot of model runs for SIG and NON-SIG soundings.

SIG and NON-SIG events when initialized with data representative of the pre-storm environment.

Figures 9 and 10 show the ensemble average and maximum plotted in terms of a SIG hail contingency table with hits, misses, and false alarms shown by quadrants bounded by the thick 2" lines. Values of Critical Success Index (CSI) and True Skill Statistic (TSS, Doswell et al 1990) are higher for ensemble maximum than for ensemble average, and are plotted on Figures 9 and 10.

The ensemble average (Figure 9) tends to underestimate the hail size, despite having the best overall correlation between forecast and reported hail size. Although Figure 10 shows considerable skill in using maximum ensemble member to forecast maximum hail size (CSI = 0.71 for SIG HAIL), it is evident from Fig. 9 that shifting all forecast ensemble averages to the right will produce a significantly improved fit. The smaller variability (r = 0.64) contained within the ensemble average forecast makes bias correction an attractive option.

In order to determine the optimal correction, the dataset was divided into three independent subsets, each comprised of varying ratios of SIG and NON-SIG hail events.

The independent subsets selected were July 1997 to June 2000, July 2000 to August 2002, and all of 2000. The forecast sizes in each group were then adjusted uniformly until the CSI for SIG hail forecast was maximized. This analysis revealed that the CSI score was optimized by increasing the forecast diameters by an average of 0.6". This

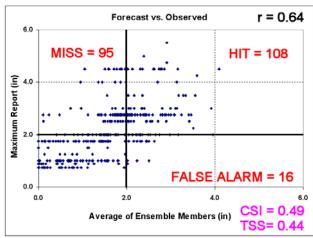


Figure 9. Ensemble average diameter versus observed hail diameter.

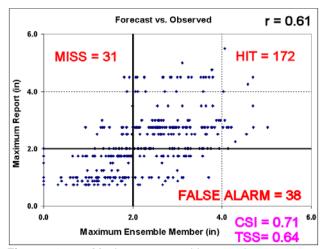


Figure 10. Maximum ensemble member versus observed hail diameter.

value was consistent for all three independent sets as well as the original set. In other words, the model ensemble average forecast has a negative bias, in terms of maximum skill score, of approximately 0.6".

Figure 11 shows that after the bias correction is applied, the POD of all SIG reports increases to 89% (181/203). The model produces 42 false alarms of SIG hail out of 179 NON-SIG cases, and the False Alarm Ratio (FAR) is 0.19. The CSI and TSS are 0.74 and 0.66, respectively, which is a notable improvement compared uncalibrated results. Note from Fig 11. there are twice as many false alarms as misses. This is preferable when issuing forecasts of potentially damaging hail. While a SIG report verifies the event, the absence of a SIG report does not necessarily mean that a SIG stone did not fall. Most SIG hail falls in sparsely populated areas of

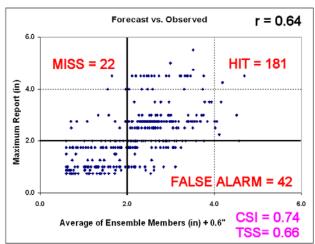


Figure 11. Bias corrected ensemble average diameter (AVG +0.6") versus observed hail diameter.

the country (see Fig. 5), thus the largest stone may not always fall where someone can report it.

In order to decrease the noise present in the data due to size reporting errors close to the 2" threshold, the database was filtered and all hail reports between 1.50" and 2.50" were removed. This left 159 SIG cases and 119 NON-SIG cases. Both of these subsets can be viewed as high confidence data categorically in terms of being markedly larger or smaller than the 2" threshold.

The improved performance of HAILCAST when using the high confidence data set for verification is clearly evident from Fig. 12. Specifically, the POD of the high confidence SIG hail is now 93% (148/159), while the number of false alarms is only 18 out of 119, with a FAR of 0.11. The CSI and TSS are now 0.84 and 0.83, respectively, and the correlation coefficient increases to 0.74.

Figure 13 shows the box and whiskers plot for the size-filtered dataset. As expected, the separation between the NON-SIG and SIG distributions increases compared to the unfiltered dataset shown in Fig. 8, with almost no overlap. These data again suggest that HAILCAST is a useful tool for assisting forecasters in delineating between NON-SIG and SIG hail events.

8. HAIL MODEL PERFORMANCE AND CAPE

A common method used to estimate maximum hail size is to calculate the maximum theoretical updraft strength due to buoyancy and to then calculate what size of hailstone can be supported by this velocity. This method is a very inaccurate

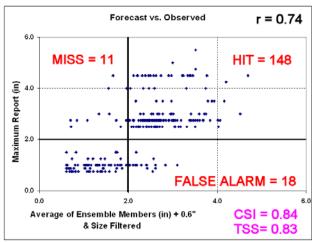


Figure 12. Bias corrected and filtered dataset with middle 1" removed centered on 2".

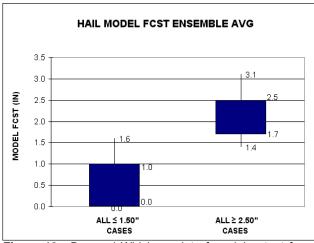


Figure 13. Box and Whiskers plot of model output for filtered datasets (Sizes are not bias corrected).

way of forecasting maximum hail size for several reasons:

- 1) Updraft strength is affected by water loading and entrainment, as well as wind shear on the updraft.
- 2) This method estimates maximum updraft velocity at the Equilibrium Level (EL), which is usually at a much higher height than where the most significant hail growth takes place (supercooled water vapor is unlikely to exist at the EL level).
- 3) Storm mode and updraft longevity are not taken into account.
- 4) Microphysical processes relevant for hail formation are not considered.

5) Using this method, large CAPE will almost always predict large hail (very high FAR).

Figure 14 shows CAPE distributions for the filtered NON-SIG and SIG subsets. Note that there is much overlap between the CAPE associated with NON-SIG and SIG hail events. Comparing Fig. 14 with Fig. 13, the ability of HAILCAST to discriminate between SIG and NON-SIG hail events in strongly unstable environments becomes apparent.

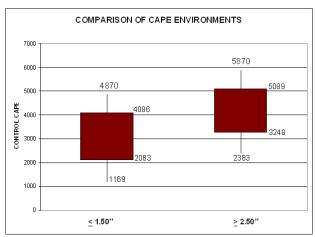


Figure 14. CAPE distribution between soundings for the filtered subsets

The only hail forecasting tool widely available to NWS field forecasters is through an AWIPS sounding algorithm. An example of output from this algorithm is shown in Figure 15. Note how the forecast maximum hail size for this particular sounding is nearly 10 inches in diameter. It is obvious that this output, meant to estimate "Max Hailsize", is worthless. Any sounding with moderate to high CAPE will produce giant, and many times, fictional hail sizes using the AWIPS algorithm.

Figure 16 shows the same sounding, but with the HAILCAST output. Note that the maximum forecast size is 2.2", with an ensemble average of 1.70". The verification for this sounding was a report of baseball-size hail (2.75"). Figures 17 and 18 are observed and RUC soundings respectively, for the 22 June 2003 Aurora record hail event. The maximum forecast size using the modified KOMA sounding is 4.10", with an ensemble average of 3.60". Of the 382 soundings in this database, only 5 had an ensemble average of 3.60" or larger. Note the 86 inch "Maximum Hailsize" off the AWIPS algorithm in Figure 18.

9. CONCLUSIONS

Thanks to increased computing power available today compared to just a few years ago, a model such as HAILCAST can be run operationally.

Testing of HAILCAST using severe hail proximity soundings yielded promising results. Despite using a relatively simplistic method of growing hail and the large degree of simplification in the cloud model, the basic microphysics are robust and appear to work well in various environments. The model is stable, and does not forecast unrealistic hail sizes.

HAILCAST is likely the best tool presently available to forecast hail size in an operational setting. Although HAILCAST will continue to be improved, it offers an objective hail size forecast that is scientifically sound and demonstrates considerable skill.

10. ACKNOWLEDGMENTS

Special thanks to Harold Brooks for introducing the SPC to HAILCAST. Many thanks to John Hart and David Bright for their extensive programming support. Thanks also to everyone at SPC for their support and assistance.

11. REFERENCES

Bluestein, H.B., E.W. McCaul, Jr., G.P. Byrd, and G.R. Woodall, 1988: Mobile sounding observations of a tornadic storm near the dryline: The Canadian Texas storm of 7 May 1986. *Mon. Wea. Rev.*, 116, 1790-1804.

Brimelow, J.C., G.W. Reuter, and E.R. Poolman, 2002a: Modeling Maximum Hail Size in Alberta Thunderstorms. *Wea. and Forecasting*, 17, 1048–1062.

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

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.

- Craven, J.P., Jewell, R.E., Brooks, HE. 2002: Comparison between Observed Convective Cloud-Base Heights and Lifting Condensation Level for Two Different Lifted Parcels. *Weather and Forecasting*, 17, No. 4, pp. 885–890.
- Craven, J.P., H.E. Brooks and J.A. Hart, 2002: Baseline Climatology of Sounding Derived Parameters Associated with Deep, Moist Convection. *Preprints*, 21st Conf. Severe Local Storms, San Antonio.
- Doswell, C.A. III, J.T. Schaefer, D.W. McCann, T.W. Schlatter, and Wobus, H.B., 1982: Thermodynamic analysis procedures at the National Severe Storms Forecast Centre. Preprints, 9th Conf. Weather Forecasting and Analysis, Seattle, Amer. Meteor. Soc., 304-309.
- Doswell, C.A., Davies Jones, R, Keller, D.L., 1990: On summary measures of skill in rare event forecasting based on contingency tables. *Wea. and Forecasting*, 5, 576-585.
- Hart, J. A., 1993: SVRPLOT: A new method of accessing and manipulating the NSSFC Severe Weather Database. Preprints, *17th Conf. on Severe Local Storms*, St. Louis, MO, Amer. Meteor. Soc., 40–41.
- Hart, J. A., J. Whistler, R. Lindsay, and M. Kay, 2003: NSHARP, version 3.90. Storm Prediction Center, National Centers for Environmental Prediction, Norman, OK,
- Moore, J. T., and J. P. Pino, 1990: An interactive method for estimating maximum hailstone size from forecast soundings. *Wea. Forecasting*, 5, 508–526.
- Poolman, E. R., 1992: Die voorspelling van haelkorrelgroei in Suid-Afrika (The forecasting of hail growth in South Africa). M.S. thesis, Faculty of Engineering, University of Pretoria, 113 pp.
- Schaefer, J.T., J.J. Levit, S.J. Weiss and D.W. McCarthy, 2004: The Frequency of Large Hail over the Contiguous United States. *Preprints*, 14th Conf. Applied Climatology, Seattle WA.
- Wong, R.K.W., N. Chidambaram, L. Cheng, M. English, 1988: The Sampling Variations of Hailstone Size Distributions. *Journal of Applied Meteorology*, 27, No. 3, pp. 254–260.

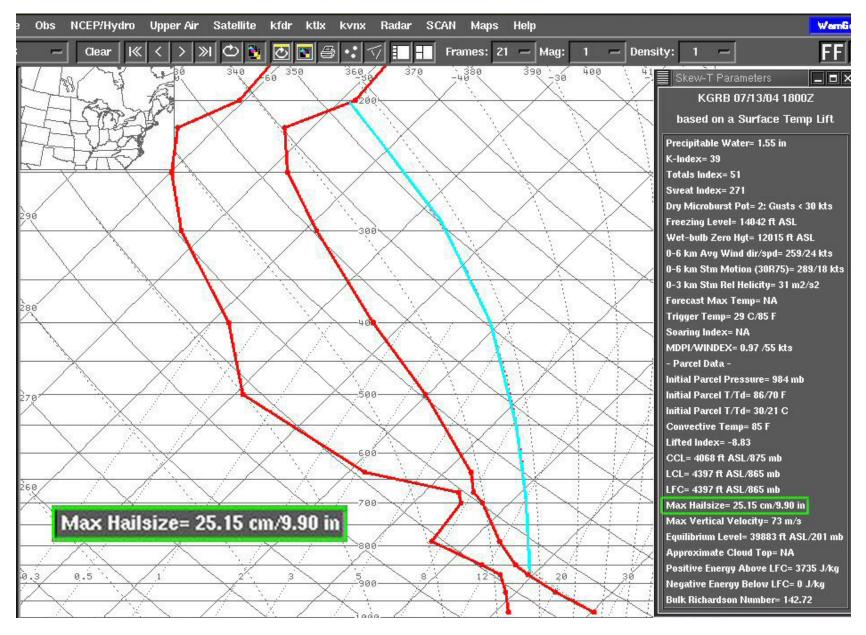


Figure 15 – Screenshot of AWIPS sounding interface for 07/13/04 18Z GRB. Note "Max Hailsize" highlighted on the right is nearly 10 inches.

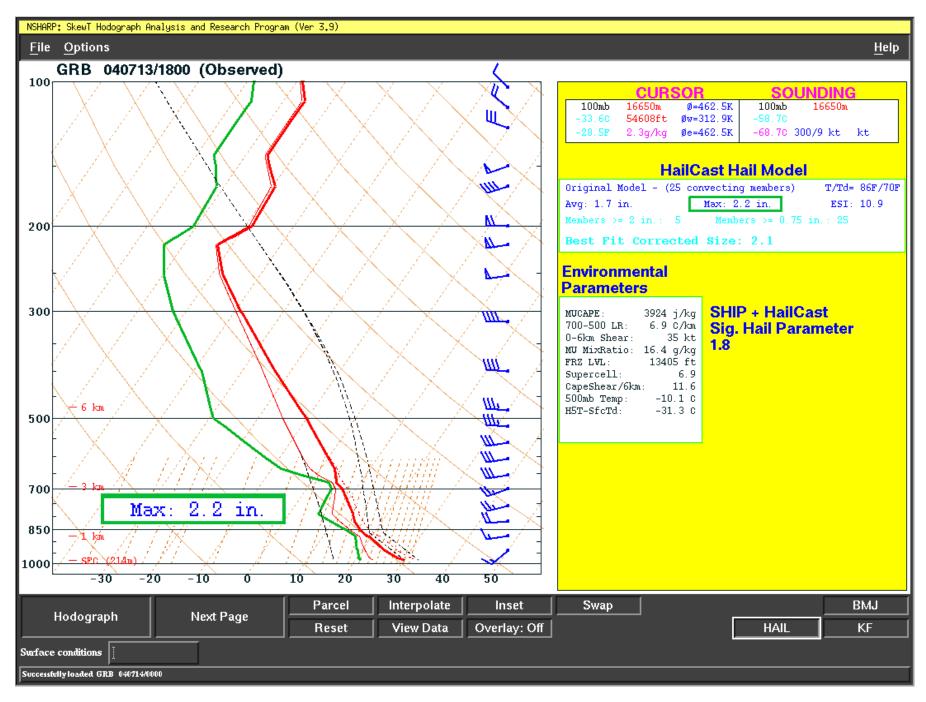


Figure 16 – NSHARP display of same sounding. Hail model output in upper right with Maximum Ensemble Member producing 2.2" diameter hail.

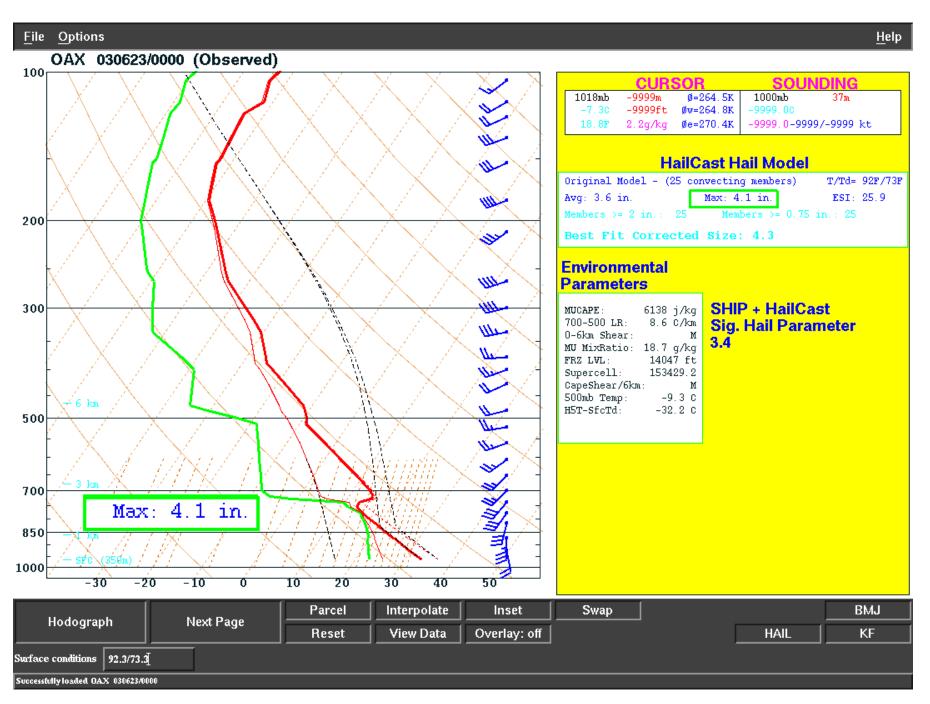


Figure 17 – NSHARP display of 00Z KOAX sounding for June 22 2003 Aurora NE record hail event. HAILCAST maximum forecast size of 4.1" diameter.

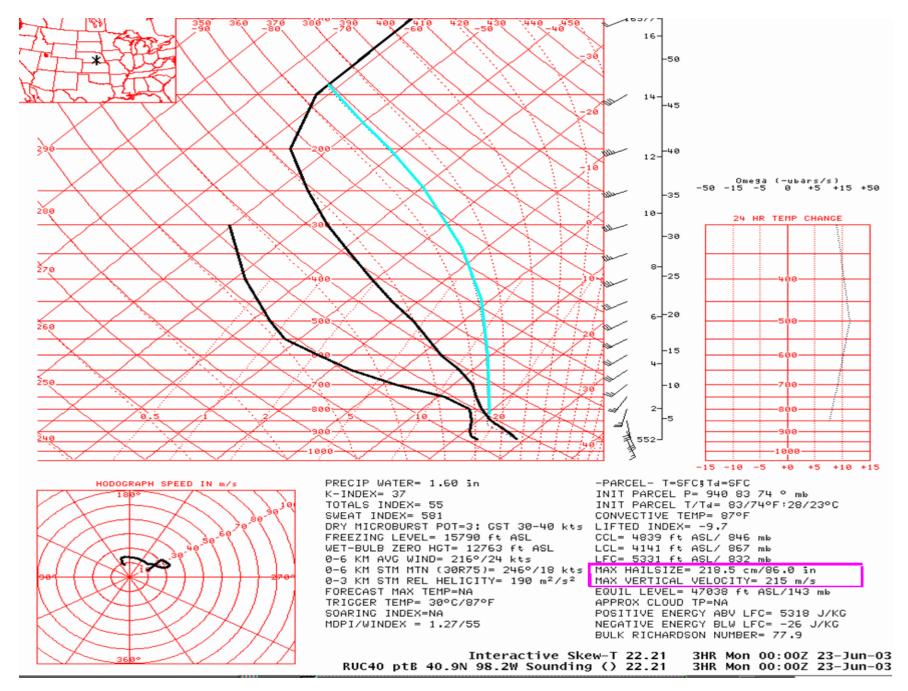


Figure 18 - AWIPS RUC Sounding for June 22 2003, near Aurora NE. Note "MAX HAILSIZE" of 86" and maximum updraft velocity of 215 m/s (0.65 MACH).