UPDATING THE USE OF HAILCAST IN NSHARP

Nathan A. Dahl*
University of Oklahoma/CIMMS and NOAA/NWS/Storm Prediction Center, Norman, OK

Rebecca Adams-Selin
Atmospheric and Environmental Research, Inc., Lexington, MA

Ryan Jewell and Israel Jirak
NOAA/NWS/Storm Prediction Center, Norman, OK

1 INTRODUCTION

The National Centers Sounding and Hodograph Analysis and Research Program (NSHARP) has been widely used in weather research and forecasting since its inception nearly 30 years ago, with continued maintenance and updates motivated by its use in the NOAA Storm Prediction Center (SPC) operations. (See Hart and Korotsky 1991, Doswell 1992, and Blumberg et al. 2017.) Among the applications installed in NSHARP is a 1-D coupled cloud and hail growth model calibrated to produce forecasts of maximum hail size, known as HAILCAST (Brimelow and Poolman 2002).

In the original form implemented in NSHARP (described in Jewell and Brimelow 2009, hereafter "JB09"), the model defines the environment using a single sounding and lifts a near-surface parcel to estimate in-cloud profiles of vertical velocity (based off a specified value at cloud base) and water content. The model then inserts a liquid hail embryo at cloud base and tracks its evolution until it either falls to the ground or remains liquid (or melts after freezing) with no prospect of being lifted into a freezing layer. To account for the fact that thunderstorm updrafts do not persist indefinitely in nature, HAILCAST specifies a maximum updraft duration, W_{dur}. If the model time reaches this value and the hailstone has not yet melted entirely or fallen to the ground, the updraft is turned off (i.e., the vertical velocity profile is set to zero at all levels) and the hailstone is allowed to fall unabated. W_{dur} is estimated using the product of the surface based CAPE (J/kg) and the 0-6 km bulk shear normalized by layer depth (s^{-1}), known as the Energy-Shear Index ("ESI"; Brimelow and Poolman 2002), with a maximum value of 3600 s.

The JB09 implementation of HAILCAST generates a forecast ensemble of 25 "member" embryos by varying the parcel base temperature and dewpoint in 0.5 C increments over a $\pm$1 C range around the observed values. There are two versions of the model (each producing its own ensemble), defined by the method of specifying initial embryo size, cloud-base updraft speed, and ESI thresholds governing storm type, as well as the method used to calculate the final forecast size. In the first version ("JB09v1"), the listed parameters are set to calibrated values based on the mixing ratio of the most unstable environmental

* Corresponding author address: Dr. Nathan Dahl, Cooperative Institute for Mesoscale Meteorological Studies, 120 David L. Boren Blvd., Suite 2100, Norman, OK 73072. email: *nathan.dahl@noaa.gov*

parcel, and the bias-corrected ensemble mean is selected as the forecast size. In the second version ("JB09v2"), the parameters are always set to the calibrated values that performed best overall in JB09 and the bias-corrected ensemble maximum is selected as the forecast size.

JB09 gives a complete description of the settings and calibrations for the original HAILCAST versions and details the results of tests using a set of 914 proximity soundings for hail reports during the period from 1989 to 2004. In those tests, JB09v1 and JB09v2 both showed skill, although the former tended to underpredict large hail sizes due to the use of an ensemble average while the latter's use of an ensemble maximum correlated better with the maximum reported size on average. There was also a large spread in the forecasts for a given observed size, indicating large uncertainty and the potential for considerable error in the forecast for any given event. (See Fig. 1.)

Adams-Selin and Ziegler (2016) noted possible error sources in the JB09 method. For example, the use of a 1-D cloud model and ESI to estimate the characteristics of the thunderstorm updraft (e.g. intensity and W_{dur}), and the assumption that the hail embryo remains within the updraft throughout its trajectory, may be too simplistic. Accordingly, an updated version of HAILCAST (hereafter "AER HAILCAST") was developed to work in conjunction with convection-allowing models, viz. WRF.

Along with the switch from the 1-D cloud model to the WRF output for estimation of updraft characteristics, AER HAILCAST uses a sinusoidal time function to parameterize the motion of hail embryos across the updraft. Furthermore, instead of a single liquid embryo inserted at cloud base, AER HAILCAST inserts five frozen embryos of varying sizes into the midlevels of the cloud. Additional refinements include the implementation of variable density in hailstone growth and modifications to the model treatment of collection efficiency and melting and excess water shedding from the hailstone surface. For a complete description of AER HAILCAST, see Adams-Selin and Zeigler (2016) and Adams-Selin et al. (2019).

AER HAILCAST has shown promise when run in conjunction with WRF. (See Fig. 2.) Those results suggest that the updates implemented in AER HAILCAST may also be beneficial if used in NSHARP to the extent possible. For example, since NSHARP only operates on single soundings, it is currently limited to using the 1-D cloud model and ESI (rather than 3-D model output) to estimate W_{dur}; however, the hailstone motion parameterization, initial embryo specifications, and updates to the hail growth model have been installed in new NSHARP versions of HAILCAST with the same cloud-base updraft and ESI threshold settings as those employed in JB09v1 and JB09v2 (designated "AERv1" and "AERv2," respectively). The accuracy and reliability of all four NSHARP versions of HAILCAST is examined below.

2 METHODS

As an extension of JB09, the various HAILCAST versions were first tested on the full set of >= 1-inch "ground truth" hail reports from the continental United States for the 2015 calendar year. Since HAILCAST is intended to forecast *maximum* hail size, the hail reports were first passed through a spatiotemporal filter; only the maximum hail

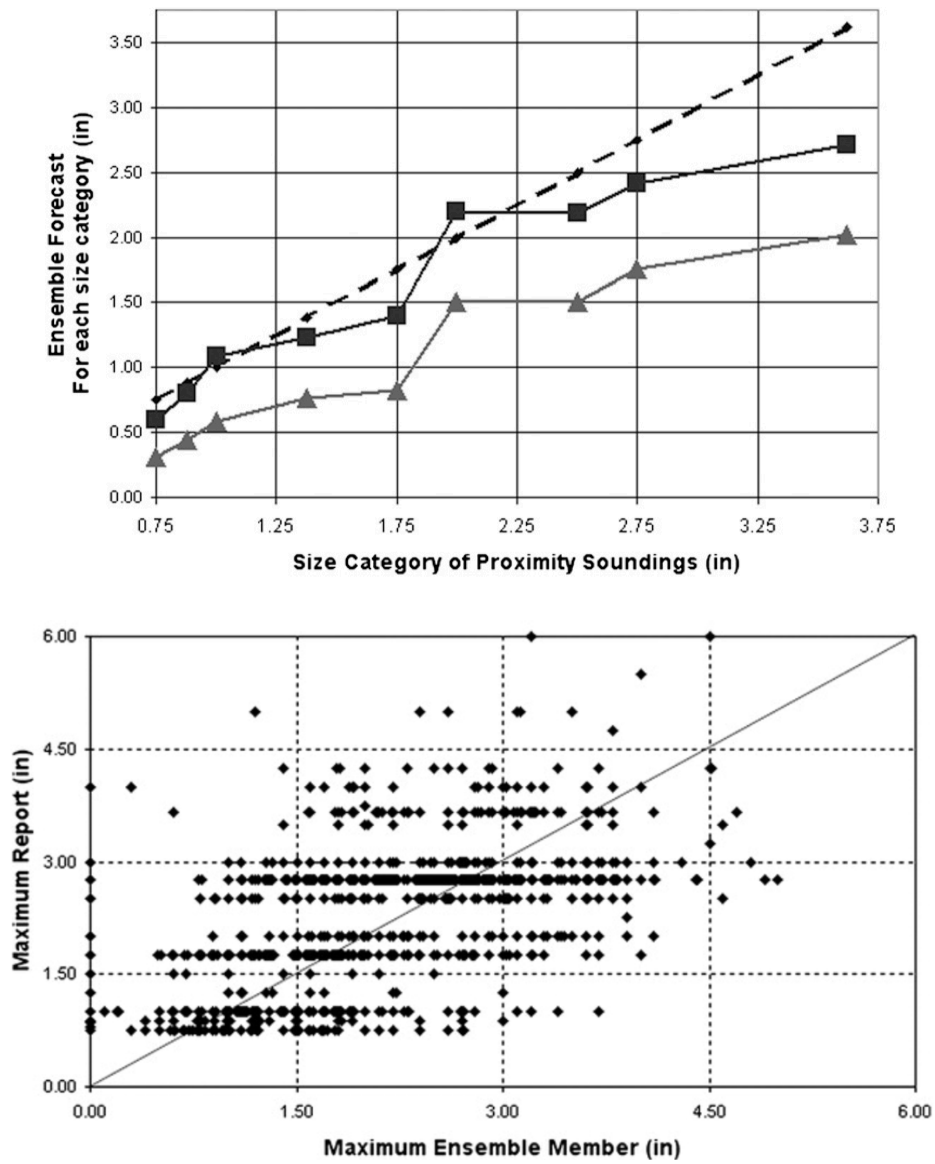

Figure 1 (top) Observed size category vs. mean of binned JB09v2 forecast sizes from selected 1989-2004 cases. (bottom) Corresponding scatter plot of forecasted vs. reported hail size for JB09v2. (Figures copied from Jewell and Brimelow 2009).

size reported within a 100 km radius over a ±2 hr time window was retained, leaving 829 cases in the final sample. Instead of proximity soundings, vertical profiles from the corresponding SPC mesonalyses were plugged into NSHARP; for each hail report, the profile with the greatest MUCAPE within a 40 km radius of the report and the two hour period up to and including the report time was selected as the most representative depiction of the environment prior to convective contamination.

Due to concerns regarding relative scarcity and possible errors (e.g. subjective clustering around certain thresholds as described in JB09) in "ground truth" hail reports, further testing was performed using archived Maximum Expected Size of Hail (MESH; Cintineo et al. 2012) tracks derived

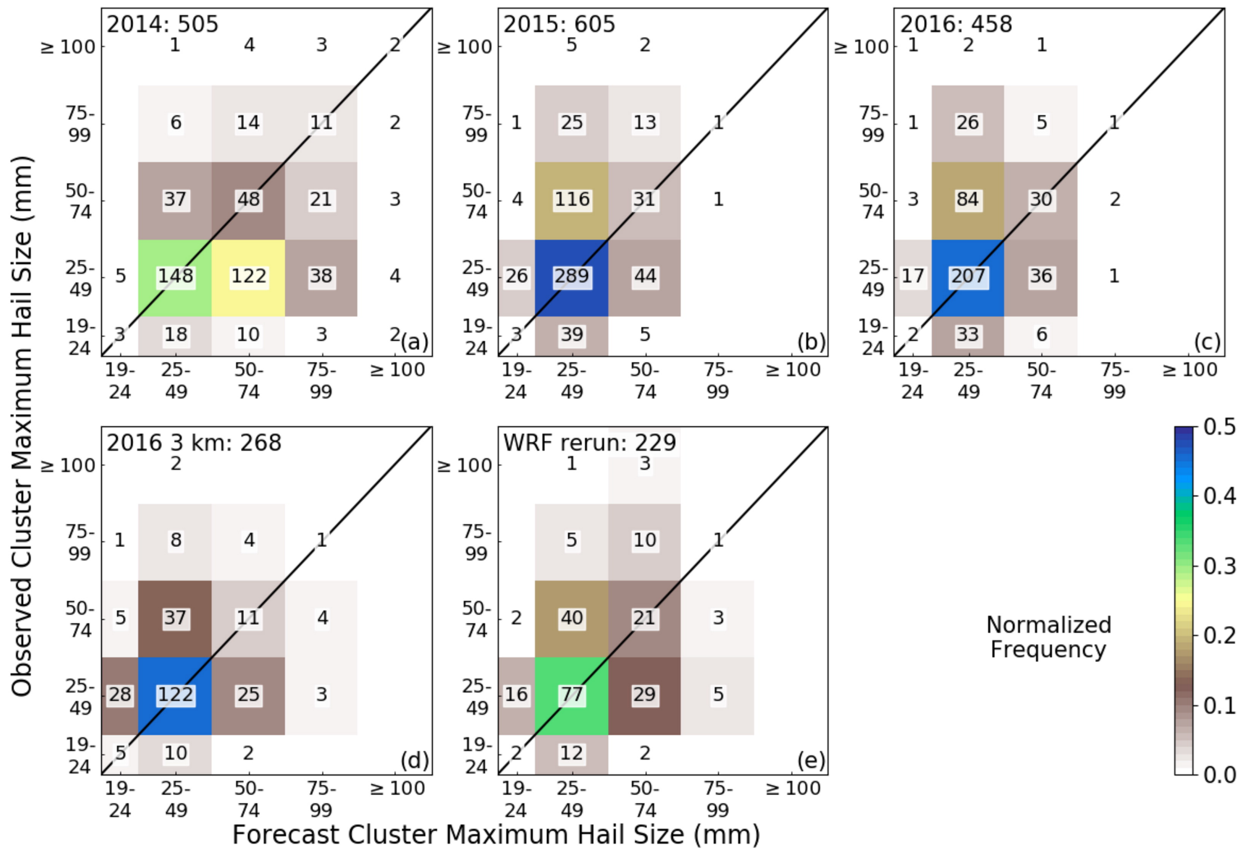

Figure 2 Two-dimensional histograms of frequency of maximum observed and forecasted hail sizes for matched clusters in object-based verification of WRF-HAILCAST for cases during the 2014-2016 NOAA/Hazardous Weather Testbed Spring Forecasting Experiments. (Figure copied from Adams-Selin et al. 2019.)

from radar observations from the period from June 2011 to May 2017. The tracks are archived in 1 x 1 km pixels over the CONUS. Pixel values are binned by size (using thresholds of 0.25, 0.5, 0.75, 1.0, 1.5, 1.75, 2.0, 2.5, 2.75, 3.0, and 4.0+ inches) for each track; to reduce noise, the maximum size for a given track was defined as the size bin containing five or more pixels.

The resulting maximum sizes were then spatiotemporally filtered, and the corresponding environmental sounding profiles were selected, in the same manner as described above for the "ground truth" hail reports sample. This MESH dataset is much larger (55019 cases) and is not prone to the subjective errors noted previously for "ground truth" reports; however, prior studies (e.g. Cintineo et al. 2012) have reported a small positive bias in MESH values, which the current analysis neglects. All four HAILCAST versions (JB09v1, JB09v2, AERv1, and AERv2) were then run on the full "ground truth" and MESH samples.

It should be noted that the AERv1 and AERv2 implementations produce an ensemble of 125 members (five embryos, 25 parcel base T/Td perturbations per embryo) and the method for determining the final forecast size has not been previously established (whereas the method for JB09v1 and JB09v2 is described in the previous section). Here, both the ensemble maximum (MAX) and the maximum value of the means for the individual embryos (MME) were tested. (It should be noted that the full ensemble mean and the mean of the maximum values for the individual embryos were also tested, but did not improve substantially on the results and are therefore not shown here.)

3 RESULTS

The scatter plots in Fig. 3 compare the archived hail size reports and corresponding forecast sizes for the various versions of HAILCAST. The bias and RMSE for each forecast method are shown in Fig. 4. There is a clear high bias in the forecasts, mainly for observed sizes <2 inches. For AER HAILCAST, using MME instead of MAX reduces this bias as well as the overall error; however, MME also tends to substantially underforecast hail size for observations >2 inches. As a result, there is little, if any, significant trend in observed size as MME forecast size increases (see Fig. 5), limiting the value of the MME forecasts in distinguishing between "non-severe," "severe," and "significant severe" hail cases.

Figure 6 shows the scatter plots for the MESH archive. Even though the sample is much larger and the data source used for verification is different (i.e. maximum MESH instead of "ground truth" observations), the results are qualitatively similar: JB09v1, JB09v2, AERv1 MAX, and AERv2 MAX show a strong tendency to overforecast size for

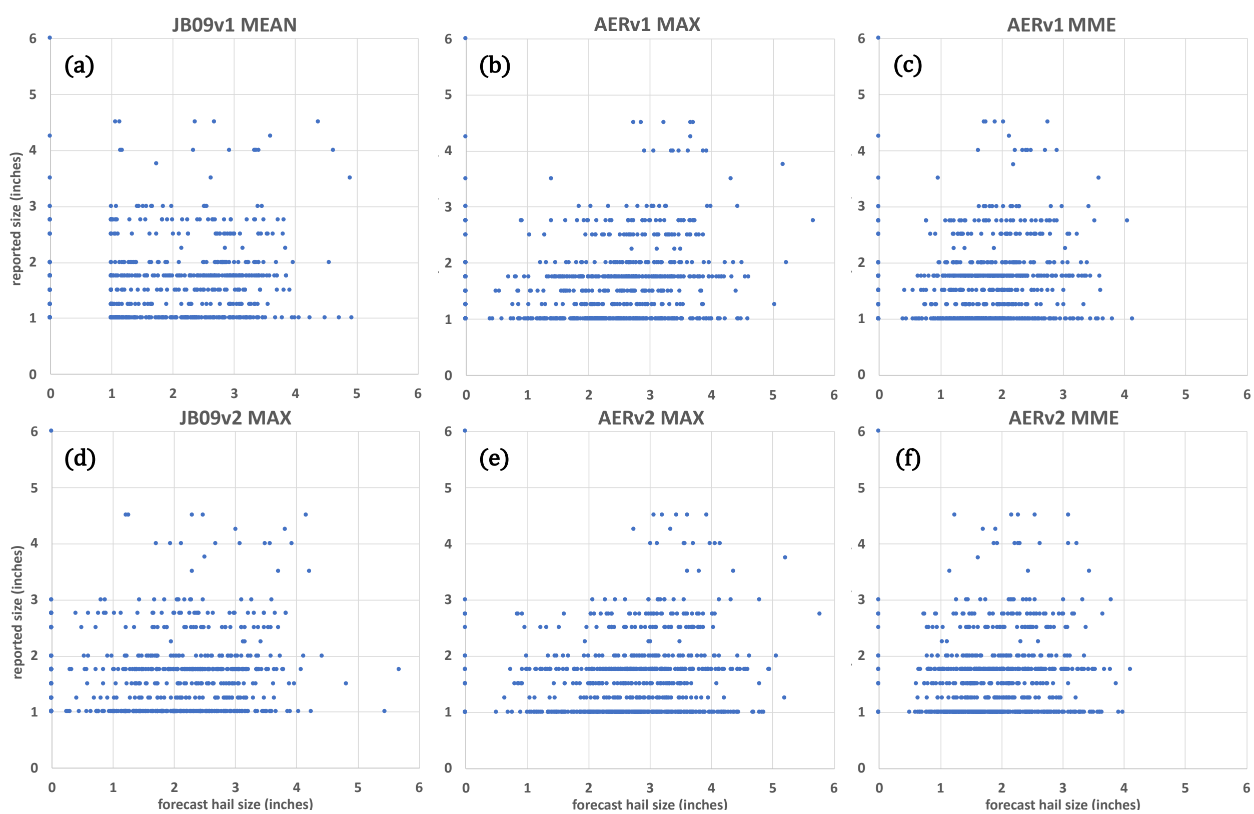

Figure 3 Scatter plots of observed vs. forecasted maximum hail size for (a) JB09v1, (b) AERv1 MAX, (c) AERv1 MME, (d) JB09v2, (e) AERv2 MAX, and (f) AERv2 MME HAILCAST versions.

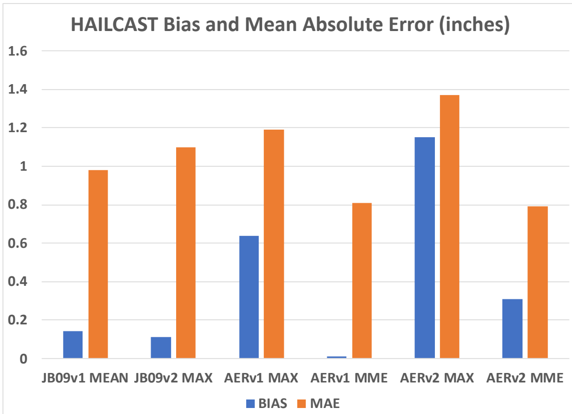

Figure 4 Bias (blue) and RMSE (orange) relative to "ground truth" hail size reports for different versions of HAILCAST

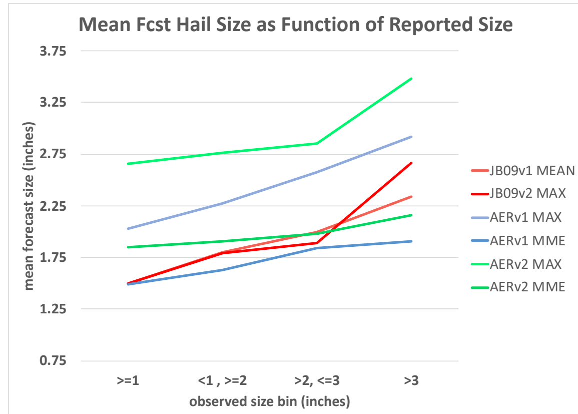

Figure 5 Mean forecast size as a function of "ground truth" hail size bin for different versions of HAILCAST

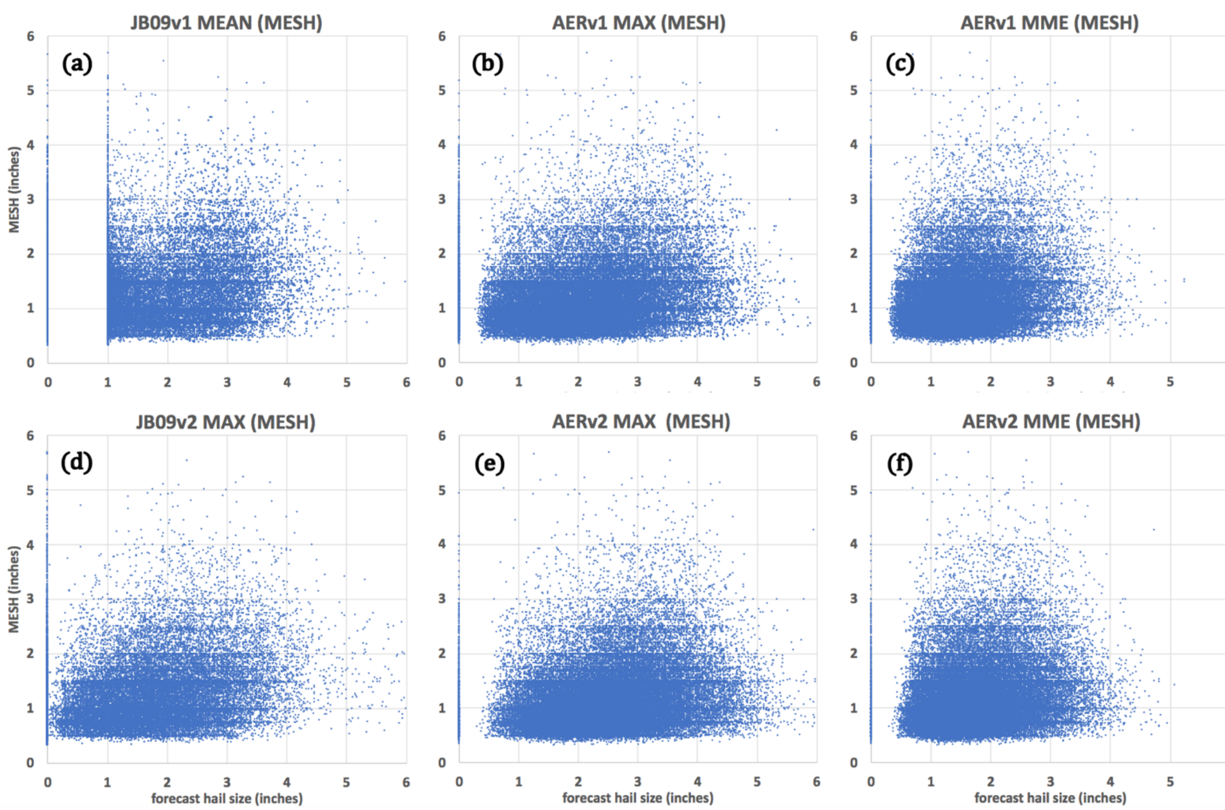

Figure 6 Same as Fig. 3, but for MESH vs. forecasted maximum hail size.

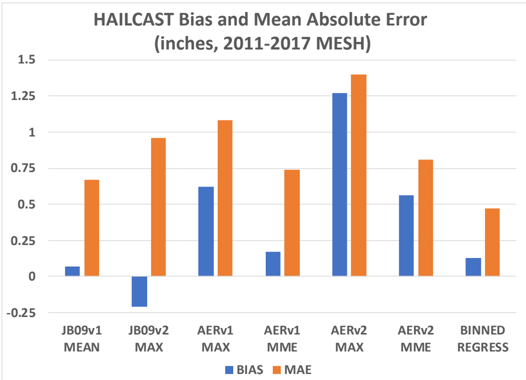

Figure 7 Same as Fig. 4, but relative to MESH. Results from binned regression based on CAPE and 0-6 km shear are included at right.

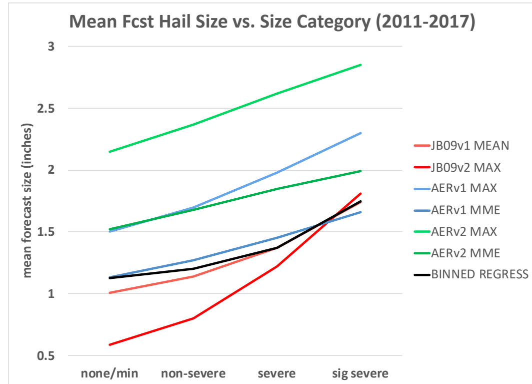

Figure 8 Same as Fig. 5, but relative to MESH. Result from binned regression based on CAPE and 0-6 km shear is shown in black.

smaller MESH, while AERv1 MME and AERv2 MME are less biased for smaller MESH but tend to underforecast larger MESH values. Thus, the error statistics are improved by using MME instead of MAX (see Fig. 7), but the ability to distinguish between "non-severe," "severe," and "significant severe" cases suffers (see (Fig. 8).

4 DISCUSSION AND FUTURE WORK

The deficiencies noted here for the various versions of HAILCAST are attributable at least in part to the simplistic treatment of the environment in and around the storm in current implementation of the 1-D cloud model. In this framework, nothing is done to account for local modification of conditions (e.g., due to nearby convection) in the vicinity of the hail occurrence. Furthermore, the estimated embryo updraft residence time W_{dur} depends entirely on ESI, which depends solely on CAPE and bulk shear.

With these limitations, the value added by the hail growth model is questionable. A separate set of hail size forecasts was produced for the MESH dataset using simple linear regressions of maximum MESH as a function of CAPE within bins spaced every 500 J/kg for CAPE and every 15 m/s for shear, with a final bulk calibration performed on the aggregated results. A scatter plot of the forecasts is shown in Fig. 9, while the error statistics and trends are appended to Figs. 7 and 8. Even though this method makes no use of HAILCAST, ot compares favorably to the best predictions from the HAILCAST ensembles. This lends support to the use of historical analogs (e.g., the Sounding Analog Retrieval System; see Jewell 2010) as an alternative to explicit hail growth modeling when data are limited to 1-D, pre-convection atmospheric profiles.

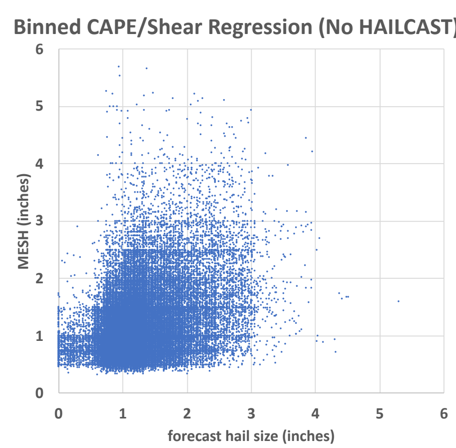

Figure 9 Scatter plot of MESH vs. forecasted hail size using simple CAPE/shear bin regression of MESH values from the 2011-2017 period

On the other hand, subsequent tests demonstrate that HAILCAST is capable of producing highly skilled forecasts based on single environmental profiles if key parameters are appropriately specified. One of the main error sources in the current implementation is the estimation of W_{dur} based solely on the profile-derived ESI. To evaluate the potential for model improvement if this error is reduced, a "superensemble" was created for every case in the MESH dataset by running HAILCAST with W_{dur} explicitly specified, using values ranging from 600 s to 3600 s in 600 s increments.

The member from each superensemble that most closely matched the maximum MESH value) for that case was selected as the "best forecast." The results are shown in the scatter plots in Figure 10. For each version of HAILCAST, the "best forecast" values are generally quite close to the maximum MESH values. This is particularly true for AERv2, suggesting that there is the potential for substantial benefit from implementing AER HAILCAST in NSHARP. However, in order to realize this benefit, a more reliable method of estimating W_{dur} from a single profile must be found. Greater insight is also needed to help define a subset of ensemble members (embryo size and T/Td perturbation) that are most representative of the hail size potential for a given environment. Because the number of combinations of sounding-derived parameters that may be relevant to these questions is too large to examine manually, future work will attempt to clarify these matters using machine learning.

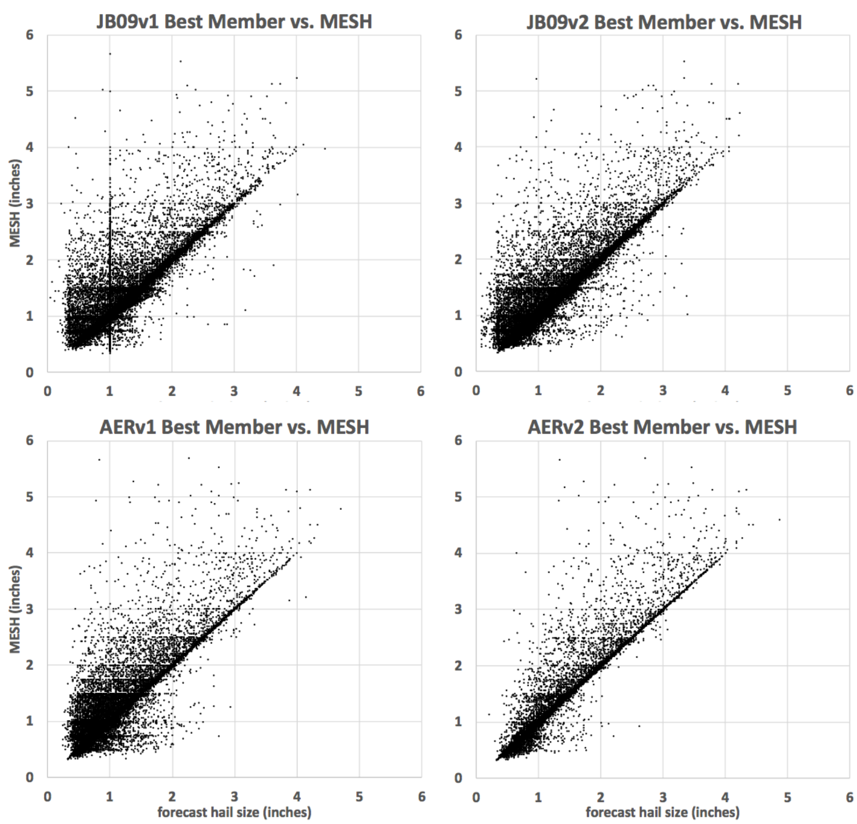

Figure 10 Scatter plots of MESH vs. forecasted hail size using best member from specified-W_{dur} "superensemble" for all cases in the MESH dataset for all versions of HAILCAST.

ACKNOWLEDGEMENTS

This work was partially funded by NOAA JTTI grant number NA18OAR4590388. This extended abstract was prepared by Nathan Dahl with funding provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA16OAR4320115, U.S. Department of Commerce. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of NOAA or the U.S. Department of Commerce.

REFERENCES

Adams-Selin, R. D., and C. L. Ziegler, 2016: Forecasting hail using a one-dimensional hail growth model within WRF. *Mon. Wea. Rev.*, **144**, 4919-4939.

Adams-Selin, R. D., A. J. Clark, C. J. Melick, S. R. Dembek, I. L. Jirak, and C. L. Zeigler, 2019: Evolution of WRF-HAILCAST during the 2014-16 NOAA/Hazardous Weather Testbed Spring Forecasting Experiments. *Wea. Forecasting*, **34**, 61-79.

Blumberg, W. G., K. T Halbert, T. A. Supinie, P. T. Marsh, R. L. Thompson, and J. A. Hart, 2017: SHARPpy: An open-source sounding analysis toolkit for the atmospheric sciences. *Bull. Amer. Meteor. Soc.*, **98**, 1625, 1636.

Brimelow, J. C., and E. R. Poolman, 2002: Modeling maximum hail size in Alberta thunderstorms. *Wea. Forecasting*, **17**, 1048-1062.

Cintineo, J. L., T. M. Smith, V. Lakshmanan, H. E. Brooks, and K. L. Ortega, 2012: An Objective High-Resolution Hail Climatology of the Contiguous United States. *Wea. Forecasting*, **27**, 1235-1248.

Doswell, C. A., III, 1992: Forecaster workstation design: Concepts and issues. *Wea. Forecasting*, **7**, 398-407.

Hart, J. A., and W. Korotsky, 1991: The SHARP workstation v1.50 users guide. NOAA/National Weather Service, 30 pp.

Jewell, R., 2010: A Sounding Analog System to Produce Probabilitic Forecasts of Maximum Hail Size. *25th Conf. on Severe Local Storms*, Denver, CO, Amer. Meteor. Soc., 4B.3.

Jewell, R., and J. C. Brimelow, 2009: Evaluation of Alberta hail growth model using severe hail proximity soundings from the United States. *Wea. Forecasting*, **24**, 1592-1609.