

A LOCAL LARGE HAIL PROBABILITY EQUATION FOR COLUMBIA, SC

Mark DeLisi
NOAA/National Weather Service Forecast Office
West Columbia, SC

Editor's Note: The author's current affiliation is NWSFO Mt. Holly, NJ.

1. INTRODUCTION

Billet et al. (1997) derived a successful probability of large hail (diameter greater than or equal to 0.75 inch) equation as an aid in National Weather Service (NWS) severe thunderstorm warning operations. Hail of that diameter or larger requires a severe thunderstorm warning be issued by NWS offices. Shortly thereafter, a Weather Surveillance Radar-1988 Doppler (WSR-88D) radar probability of severe hail (POSH) became available for warning operations (Witt et al. 1998). This study recreates the steps taken in Billet et al. (1997) to derive a local probability of large hail equation (LPLH) for the Columbia, SC National Weather Service Forecast Office (CAE) warning area, and it assesses the utility of the LPLH relative to the POSH.

A logistic regression approach was used to develop a LPLH for the CAE warning area. Independent variables included in the LPLH were vertically integrated liquid (VIL) and the ratio of VIL to echo top height (ET), both computed by the WSR-88D; and 500-hPa temperature at Peachtree City, Georgia (FFC) and Charleston, South Carolina (CHS) extracted from upper air soundings using the

Skew-T/Hodograph Analysis and Research Program, or SHARP (Hart and Korotky 1991). The dependent variable, hail diameter or the absence of hail, was derived from spotter reports for the CAE warning area from May, 1995 through September, 1996. There were 136 cases used to develop the regression equation.

The LPLH was verified using 69 spotter reports from the period September, 1996 through September, 1997. The POSH was verified as well. Verification statistics included the Brier score and the chi-square (χ^2) statistic.

The goal of this study was to develop an objective method to estimate the probability of large hail for use in forecast operations. Since the WSR-88D algorithms also produce a POSH, the locally produced method was verified in comparison to determine if the local regression equation was more accurate.

Since Billet et al. (1997) used data from a WSR-88D and an upper air sounding that were co-located while this study did not, similarities and differences in the results from Billet et al. (1997) are discussed.

2. DATA

The dependent variable in the study was the categorical (binary) representation of the occurrence of large hail. Dependent variable data were obtained from reports from local fire, police, and postal workers, trained spotters, NWS employees, and others. Independent variable data were obtained from the WSR-88D and upper air soundings from Peachtree City, Georgia (FFC) and Charleston, South Carolina (CHS) archived locally and extracted via SHARP.

Potential independent variables include the following:

- 1) vertically integrated liquid
- 2) vertically integrated liquid density
- 3) lifted index
- 4) convective available potential energy
- 5) sweat index
- 6) total totals index
- 7) bulk Richardson number
- 8) 700-500-hPa lapse rate
- 9) 850-hPa temperature
- 10) 700-hPa temperature
- 11) 500-hPa temperature
- 12) 300-hPa temperature
- 13) precipitable water
- 14) 700-hPa dewpoint depression
- 15) freezing level
- 16) wet-bulb zero level
- 17) equilibrium level
- 18) echo top height
- 19) storm-relative helicity
- 20) storm-relative directional shear
- 21) positive shear
- 22) mean storm relative inflow (sfc to 2 km)
- 23) 500-200-hPa speed shear

The list of potential independent variables was similar to Billet et al. (1997). Variables not utilized by Billet et al. (1997) were: sweat

index, total totals index, bulk Richardson number, equilibrium level, echo top height (ET), speed shear in the layer between 200 h-Pa and 500 h-Pa, and vertically integrated liquid (VIL) density. VIL density, which is the ratio of VIL to ET, has been shown to be useful in discriminating between severe and non-severe storms (Paxton and Shepherd 1993).

WSR-88D independent variable values came from storm interrogation within 20 minutes of the time of hail, or in the cases when hail was reported not to have occurred, from storm interrogation during the same period at the time of maximum VIL. This information came from local Archive III and Archive IV optical disks. Most of the time, upper air variable values came from the most recent FFC soundings. The soundings were modified at the surface to match CAE temperature and dew point at the time of hail occurrence.

The FFC sounding is launched approximately 210 miles from CAE, but it was determined to be the most representative sounding of the CAE warning area in most cases. All other soundings except the WSO Greensboro (GSO) and CHS soundings are launched farther away from CAE than the FFC sounding. The GSO sounding is launched too far north to be as representative as the FFC sounding. The fairly frequent marine influence on the CHS sounding usually rendered it less representative than the FFC sounding. However, in four of the 136 cases, there was an air-mass boundary between CAE and FFC that made the most recent FFC sounding less representative of the atmosphere over CAE than the CHS sounding. In those cases, the CHS sounding was used. The CHS sounding is launched approximately 95 miles from CAE.

In order to maximize the number of cases for the logistic regression technique, virtually all hail reports and all verifiable reports of no hail were counted as cases. There were only two limiting criteria. The first was that only one report per thunderstorm cell was retained. For a given cell, when there were more than one report of hail, one hail report was chosen at random.

When there were one report of hail and one or more reports of no hail, the hail report was chosen. When there were no reports of hail and more than one report of no hail, one no-hail report was chosen at random. The second limiting criterion was that all cases were between 14 and 105 miles from the radar site. The minimum distance was necessary to avoid problems with the radar's "cone of silence." The maximum distance was the shortest distance that encompassed the entire CAE warning area. There were 136 cases used to derive the equation. Of those 136 cases, 44 were of no hail, 37 were of small hail (diameter less than 0.75 inch), and 55 were of large hail.

Two potential sources of error in the independent data (sounding based) were that the radar site and the sounding sites were not collocated, and that soundings were not coincident in time with storms. Another possible source of error in the independent data (radar based) was that in most cases only a data interval was available for VIL. A VIL of 40 up to but not including 45 kg m^{-2} was designated as 40 kg m^{-2} , a VIL of 45 up to but not including 50 kg m^{-2} was designated as 45 kg m^{-2} , and so on.

A source of error in the dependent data was the estimation of hail sizes by spotters. Very few spotters actually measure hail diameter. Another source of error in the dependent data

that is common to locations with sparse populations was that the occurrence of hail may be unreported and/or undetected. The CAE warning area has many parts that are sparsely populated. This study went to considerable lengths to minimize these sources of error.

3. EQUATION

A commercially available statistical software package was used to derive a probability of large hail equation using logistic regression analysis. The logistic regression approach results in an equation that expresses the probability of some categorical event (in this case, the occurrence of large hail) as the sum of weighted values of quantifiable variables. A complete explanation of logistic regression can be found in Freeman (1987).

The statistic used to select independent variables in this approach is the likelihood ratio chi-square statistic (G^2). G^2 relates directly to the discrepancies between (a) the predicted number of categorical event occurrences and non-occurrences from a model and (b) the actual number of event occurrences and non-occurrences. A model with a high G^2 value has little or no predictive value; so, the more model G^2 is reduced by inclusion of a given independent variable, the better the predictive value of that variable. A p-value is associated with the reduction in G^2 by a given variable, and in this case the p-value is the chance of obtaining such a reduction randomly. Therefore, low p-values are associated with variables that have high predictive value.

The first step of this regression approach is to run each potential independent variable in a one-independent variable equation, and the

variable that results in the greatest reduction in G^2 is kept (provided the associated p-value is less than or equal to .05). The next step is to run each remaining variable together with the one kept from the previous step in a two-independent variable equation. The one that results in the greatest reduction in G^2 is also kept. This process is continued until the reduction in G^2 resulting from keeping the “best” remaining independent variable has an associated p-value greater than .05. That “best” remaining independent variable is not kept. Then, the two-way interactions between the retained independent variables ($X_a X_b$) should be tested as potential independent variables in a similar manner.

The resulting logistic regression equation is of the form:

$$P = 1/[1 + \exp(B_0 + B_1 X_1 + \dots + B_n X_n)], \quad (1)$$

where:

P = probability of large hail

B_0 = constant

B_1 through B_n = coefficients

X_1 through X_n = independent variables.

In this study, the independent variables kept were VIL density, VIL, and 500-hPa temperature. The respective p-values associated with the resultant reductions in G^2 were zero, .0293, and .0088. No two-way interaction resulted in a reduction of G^2 great enough to yield a p-value less than or equal to .05.

Collinearity between independent variables is not as detrimental to the results of logistic regression as it is to those of multiple linear regression. Freeman (1987) notes that, although great care should be used when interpreting the results of a logistic regression when some independent variables are highly

correlated, the results themselves are not contaminated. Among the three independent variables retained in this study, only VIL and VIL density were significantly correlated. Pearson’s product-moment correlation coefficient between the two values was .7578. The p-value of the correlation was zero with 134 degrees of freedom. In this case, the p-value meant that the probability of obtaining such a high correlation coefficient between two unrelated variables was zero.

The final logistic regression model (LPLH) was:

$$P = 1/[1 + \exp(-8.463 + 2.309X_1 + 0.08680X_2 - 0.2089X_3)], \quad (2)$$

where:

P = probability of large hail

X_1 = VIL density (.003281 kg m⁻³)

X_2 = VIL (kg m⁻²)

X_3 = 500-hPa temperature (°C).

The unusual units associated with VIL density occurred because the radar expresses ET in thousands of feet, while VIL is expressed in kg m⁻². Feet were converted to meters to keep the units consistent.

The ranges of the three independent variables in the dependent data set were 0.05 to 2.17 for VIL density, 1 to 65 for VIL, and -19.9 to -4.3 for 500-hPa temperature. A statistically-derived equation should not be used when any of the independent variable values falls outside of its range in the dependent data set.

4. VERIFICATION

The LPLH and the POSH were verified by using an independent data set of 69 cases from September, 1996 through September, 1997.

Whether or not hail had occurred, the highest LPLH and POSH probability values from between 20 minutes prior to a reliable report and the time of that report were used in the verification process. Also, considerable attention was paid to updating the parameters required as input to the POSH algorithm (the 0°C and -20°C levels).

Two tests were used to verify each method of prediction. The first, the chi-square (χ^2) test, provides a test of the hypothesis that a method's predicted frequency of large hail and the actual frequency of large hail are the same. The strength of this approach is that the test yields a probability of obtaining the method's results given the hypothesis. This is because χ^2 has a known distribution, which means that values of the statistic have known p-values.

The value of χ^2 can range from zero to infinity, with higher values being associated with lower p-values. If the test yields a low p-value, it means that the probability that one should accept the tested hypothesis is low. In this study, a sufficiently low p-value (no greater than .1, and preferably no greater than .05) meant that the method's predicted frequency of large hail likely was not the same as the actual frequency of large hail, hence the method was likely unreliable. The weakness of chi-square analysis is that it provides no direct way to compare two different methods.

The chi-square test was used first to test the hypothesis that the LPLH predicted frequency of large hail and the actual frequency of large hail were the same over the 69 cases. The value of χ^2 was 0.5020 with four degrees of freedom. The associated p-value was .9733. The indication not to reject the hypothesis was very strong. The test was then used to test the

hypothesis that the POSH predicted frequency of large hail and the actual frequency of large hail were the same over the 69 cases. The value of χ^2 was 6.680 with four degrees of freedom. The associated p-value was .1538. Although the indication not to reject the hypothesis was not nearly as strong, there was no indication to reject the hypothesis at any credible probability level. The chi-square test was therefore unable to reject, for either method, the hypothesis that the predicted and actual frequencies of large hail were the same.

The second test, the Brier score, provides a method of comparing two different methods of forecasting the probability of a categorical event. It is widely used by the NWS to assess probability of precipitation forecasts. The Brier score can range between zero and one; lower Brier scores are associated with better forecasts. A complete explanation of the Brier score can be found in Wilks (1995). The Brier score allows one to compare different methods of forecasting probabilities; however, there is no way to quantitatively assess significance, because the Brier score has no known distribution.

The Brier score associated with the LPLH was .1205. The Brier score associated with the POSH was .1395. The reduction in Brier score that resulted from using the LPLH in lieu of the POSH was 14 percent.

The equation derived by Billet et al. (1997) had the apparent advantage of a radar that was co-located with a sounding. That equation was verified using an independent data set of 88 cases.

Its χ^2 value was 4.478 with nine degrees of freedom, which yielded an associated p-value of .8772. The Brier score associated with the equation from Billet et al. (1997) was .1207.

It is possible that the use of VIL density as a predictor variable in this study compensated for such an advantage.

5. PRACTICAL APPLICATION

Subsequent to verification, radar operators at CAE have used the LPLH in lieu of the POSH in their warning operations. They have developed strategies independent of this study, but those strategies are presented here.

A program was written that uses the LPLH to compute the probability of large hail. After a sounding run and when hail is considered a possibility, the 500-hPa temperature and an anticipated ET value (based on the equilibrium level of the sounding) are input. Then, various VIL values are input. The lowest VIL value that results in a computed probability of at least 70 percent is considered a first guess at the “VIL of the day.”

When convection is underway, actual ET and VIL values from the WSR-88D are input to compute LPLH probabilities. Again, radar operators at CAE generally warn once the probability reaches or exceeds 70 percent.

6. CONCLUSION

Insofar as one can say a Brier score of .1205 is acceptable, the chi-square test and the Brier score support the LPLH as a legitimate tool for predicting large hail in a thunderstorm cell.

Because of the nature of the two tests and because of the small size of the independent data set, one cannot say with certainty that the LPLH is superior to the POSH based on the tests' results. However, in both cases the

LPLH compares favorably to the POSH. This is true despite the sources of error outlined previously. In practice, the radar operators at CAE prefer the LPLH to the POSH.

Despite the lack of a co-located sounding, the results from this study are quite similar to those from Billet et al. (1997). This could be due to the inclusion of the predictor variable VIL density.

Subsequent to initiation of this study, WSR-88D software loads have included a new parameter, cell-based VIL, that is expressed as a discrete value rather than as a data interval.

This might make cell-based VIL a more attractive potential independent variable than VIL. This possibility may be investigated in subsequent studies. Finally, the belief that local predictive tools should be developed despite the existence of centralized guidance is supported.

ACKNOWLEDGMENTS

My thanks to all those whose input influenced the course of this study, especially Gary Beeley, Science Operations Officer at WSFO ATL. Thanks also to Warren Snyder, Science Operations Officer at WSFO ALB, for historical sounding data. Thanks finally to Michael Cammarata, Science Operations Officer at WSFO CAE, and Laurie Hermes, Eastern Region Headquarters Scientific Services Division, for their valuable comments on this manuscript.

REFERENCES

- Billet, J., B. G. Smith, M. DeLisi and C. Gates, 1997: Use of regression techniques to predict hail size and the probability of large hail. *Wea. Forecasting*, 12, 154-164.
- Freeman, D. H., 1987: *Applied Categorical Data Analysis*. Marcel Decker, 318 pp.
- Hart, J. A., and W. D. Korotky, 1991. The SHARP Workstation v1.50. A skew T/hodograph analysis and research program for the IBM and compatible PC. User's Manual. NOAA/NWS Forecast Office, Charleston, WV, 62 pp.
- Paxton, C. H., and J. M. Shepherd, 1993: Radar diagnostic parameters as indicators of severe weather in Central Florida. *NOAA Technical Memorandum NWS SR-149*, 11 pp.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.
- Witt, A., M. D. Eilts, J. T. Johnson, G. J. Stumpf, E. D. Mitchell, and K. W. Thomas, 1998: An enhanced hail detection algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 286-303.