LETTERS TO THE EDITOR

Personal opinions expressed by authors are not to be considered as approved by their offices of employment or the NWA.

Dear Editors,

I would like to offer some comments on the recent paper by M. A. Rose entitled "Downbursts" (*Nat'l. Wea. Dig.*, 21 (1), pp. 11-17). These comments relate to the statistical methods used to analyze the data in the study.

The author suggests an equation,

$$V\ max = (V_d\ V_{avg}\ gz)^{1/4} \quad (1)$$

to estimate the maximum potential downburst velocity (V max) in m s^{-1}. The author then calculates estimates of V max for a set of downburst cases using data from the last atmospheric sounding prior to the event. The author then computes the difference between the estimated V max and "observed" V max expressed to a tenth of a m s^{-1}. The values used for observed V max were obtained from a nearby observation site, or were inferred from damage reports. Since observed V max was not measured at the time and location of occurrence or was inferred from damage reports, it is likely that there are substantial differences between the true V max and those used in the calculation of the differences, especially when considered on the scale to which the differences were calculated. Thus, the error statistics may not be very meaningful when viewed in this context.

Given the difficulty in accurately measuring the predictand data, attempting to specifically estimate V max may not be the best approach. Perhaps trying to predict whether or not V max exceeds a given threshold would provide the same utility while lending itself to more accurate measurement. This would involve using a larger sample of cases, including those in which V max did not exceed the threshold speed. Where the author later develops a linear regression equation to predict V max, a logistical regression equation (Freeman 1987) could be developed to express the probability that V max exceeds the predetermined threshold speed. A threshold *probability* delineating the occurrence or non-occurrence of the event can then be selected so as to maximize the threat score and produce a favorable bias on the dependent data (Bermowitz and Best 1978).

The author states that the majority of the cases examined exhibit a definite correlation between calculated and observed Vmax. The existence and strength of the correlation is not obvious to the reader since only a series of forecasts and observations are provided for examination. In order to support this assertion, statistics which quantify the correlation and the significance of the correlation should be presented. Further, the author only includes cases in this study where the *observed* downburst speeds were equal to or exceeded 25.8 m s^{-1}. To better test the merits of the predictive equation, a sample of cases comprised of forecast downburst speeds can be compared with observed values. Such a sample would likely include cases where observed values were less than 25.8 m s^{-1}.

The author then develops a linear regression equation to predict V max using equation (1) as a predictor. The author states that the significance of the regression equation is in greatly reduced errors. This is not surprising since regression analysis produces an equation which minimizes the sum of the squared differences between predicted and observed values over the data sample. It would be interesting to test the author's hypothesis regarding the physical mechanisms responsible for downbursts by allowing not only the combination of variables which comprise equation (1) to be included in the regression analysis, but each of the variables independently as well as other variables to be included as well. If equation (1) is truly the best predictor over the sample, it should explain the greatest fraction of the variance in the predictand.

In addition, in order to test the predictive value of the regression equation, the equation should ideally be verified on an independent sample. This, of course is not always practical due to limitations on the availability of data. An independent verification can be simulated using a resampling technique known as *cross-validation* (Wilks 1995). Given n cases in the data sample, n regression equations are formulated using n-1 cases in the dependent sample and alternating the case left out of the dependent sample for verification. While not as appealing as verifying a single equation developed on dependent data against a sufficient sample of independent data, this approach can serve as a surrogate where data is limited.

The author develops an interesting hypothesis regarding the physical mechanisms contributing to the production of downbursts in convective storms. The author attempts to support this hypothesis through the use of statistical techniques imposed upon a sample of data. The statistical design employed in this study, and the resultant statistics as presented, do not provide meaningful evidence in support of the authors hypothesis. A different approach in the statistical design of this study could have provided much more substantive evidence in favor of or against the hypothesis.

References

Bermowitz, R. J., and D. L. Best, 1978: An Objective Method for Maximizing the Threat Score. TDL Office Note 78-15. National Weather Service, NOAA, U. S. Department of Commerce, 11 pp.

Freeman, D. H. Jr., 1987: *Applied Categorical Data Analysis*. Marcel Dekker, Inc., 318 pp.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.

Michael W. Cammarata
Science and Operations Officer
NOAA/National Weather Service Forecast Office
West Columbia, South Carolina