J4.6 DEVELOPMENT OF A DISCRIMINANT ANALYSIS MIXED PRECIPITATION (DAMP) FORECAST MODEL FOR MID-ATLANTIC WINTER STORMS.

J. Dustin Hux *, Paul C. Knappenberger, Patrick J. Michaels, Philip J. Stenger
State Climatology Office, University of Virginia, Charlottesville, Virginia

Hugh D. Cobb III
Tropical Prediction Center, Miami, Florida

Michael P. Rusnak
National Weather Service Forecast Office, Wakefield, Virginia

1. Introduction and Background

The ability to produce accurate winter precipitation type and duration forecasts is critical in regions with high frequencies of ice pellets and freezing rain. While all types of winter precipitation events are potentially disruptive, extended and widespread periods of freezing rain or freezing rain mix are by far the most destructive and hardest to predict. Because it coats exposed surfaces with a layer of ice that in some instances can exceed two inches (Lemon 1961; Michaels et al. 1991), freezing rain can damage structures, down power lines, break trees, and make surface and air travel virtually impossible. That, in turn, produces extremely dangerous and costly conditions for the transportation and utilities sectors, as well as the general public. As a result, predicting precipitation duration within type is exceedingly important to the user community.

In the mid-Atlantic, the frequency of the "wintry mix" of precipitation—freezing rain and ice pellets—is considerable. Portions of the western piedmont of Virginia average more than 50 hours per year (Figure 1) (Michaels et al. 1991; Gay and Davis 1993; Keeter et al. 1995) and ice pellet frequencies in North Carolina indicate that this maximum, though somewhat reduced, continues along the western piedmont to the South Carolina border (Musick 1991). Indeed, Keeter et al. (1995) have written that "perhaps the most challenging winter weather forecast problem in the Southeast is forecasting precipitation type."

Our research objective was to develop an improved methodology for the operational forecasting of winter mixed-precipitation type in the Mid-Atlantic. Specifically, the requirements were to develop a model that is easy for the forecaster to use and understand, has the highest possible vertical resolution, incorporates climatological data from a long period of record, and provides the forecaster with probability guidance in situations that are difficult to resolve.

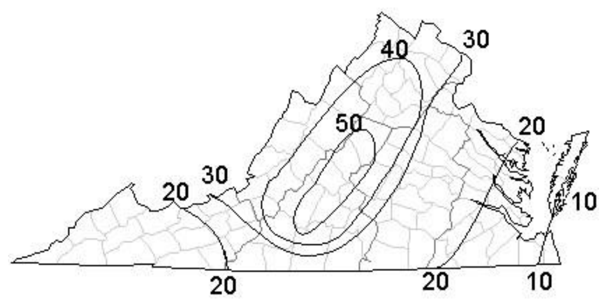

Figure 1. Average number of freezing rain hours per year in Virginia (after Michaels et al., 1991).

To that end, empirical relationships between the number of freezing levels, the temperatures from mandatory and significant levels, and other variables derived from routine NWS rawinsondes were examined with regard to precipitation type using discriminant analysis. Discriminant analysis provided the ability to separate and classify precipitation type (categorical data) based on continuous meteorological data and to produce probability estimates for group membership. The equations describing these relationships were used to develop the Discriminant Analysis Mixed Precipitation (DAMP) Model. Essentially the DAMP model addresses what type of precipitation will occur during a winter precipitation event.

2. Data and Methodology

Data from two distinct sources were merged to investigate the meteorological conditions necessary to support the various winter precipitation types. The first included the National Oceanic and Atmospheric Administration's (NOAA) hourly surface airways observations of precipitation type and surface temperature from Greensboro, NC (1948–1995) and Washington Dulles Airport, VA (1962–1995). All hours reporting precipitation from the months November through March were extracted for each station. Precipitation type was defined as rain (liquid rain or

* *Corresponding author address:* J. Dustin Hux, Univ. of Virginia, Dept. of Environmental Sciences, Charlottesville, VA 22904; email: jdh2r@virginia.edu.

drizzle), freezing mix (freezing rain/drizzle or freezing rain/drizzle mixed with any other precipitation type), ice pellets, and snow (or snow/rain mix). The second data set was the record of twice-a-day (0000 and 1200 UTC) upper-air observations for Greensboro (1948–1995) and Washington Dulles (1962–1995) extracted from NOAA's Radiosonde Data of North America. These were the only two stations in the region of interest that simultaneously collected precipitation type and sounding data over an extended period of record—one that ensures that enough observations are available within and between precipitation types to establish stable distributional relationships.

For each station, these data sets were combined to ensure that the most representative upper-air observations were available for each precipitation occurrence. To accomplish that, only precipitation occurring within two hours of an upper-air observation time (a five-hour window) was used for analysis. Within this five-hour window, each hour's precipitation type was assigned to the upper-air observation at the window's midpoint. Though this could introduce errors into the analysis in the form of different precipitation types within a five-hour window being associated with a single upper-air observation, the occurrence of this situation was rare, and precipitation type generally remained the same within each five-hour window. Additionally, in instances when any one meteorological variable was missing or not available, the entire case was excluded from analysis. Although that eliminated a large amount of data, the number of hourly observations including precipitation was sufficiently large that many hundreds of observations for each station remained—2,475 and 5,845 total observations for Washington Dulles and Greensboro, respectively.

Through this method, the data set constructed for each station contained the observed precipitation type, surface temperature, and the upper-air observations of height (m) and temperature (°C) at 50 hPa increments between 1000 hPa and 500 hPa. In addition to the height and temperature data, we calculated and included the height(s) (m) of the 0°C isotherm (referred to as "freezing levels" even though 0°C is actually the melting temperature of ice, not the freezing temperature of water), and a temperature index (summation of temperatures) between each freezing level.

Using some basic meteorology, the upper-air data were combined to create vertical temperature profiles for each observation of precipitation type. These profiles were then separated depending on the number of freezing levels (i.e., the number of times the sounding crossed the 0°C isotherm) and examined to determine if they could be used to isolate specific precipitation types and as a result, segregate winter precipitation scenarios by forecasting difficulty. Discriminant analysis was used to further separate and classify observations within each scenario if categorizing the profiles failed to isolate precipitation types.

Discriminant analysis (Fisher 1936), a multidimensional discrimination and classification scheme, has two major goals: 1) Description of group separation and 2) Prediction of group membership. To quantify group separation, linear functions of the variables were used to elucidate group differences by identifying each variables relative contribution to group separation. In the next step, linear or quadratic functions were used to classify individual observations into one of the groups.

To reduce bias during the development of the DAMP model while using all of the available data, the robustness of discriminant functions was checked using cross-validation. Cross-validation treats n–1 out of n observations as a training set, develops the discriminant functions based on these n–1 observations, and applies them to classify the remaining observation for n iterations. This method should achieve a nearly unbiased estimate. The model's performance was evaluated by examining the resulting classification probabilities (e.g., probability of detection and false alarm ratios). Probability of detection (POD) is the ratio of the number of correct forecasts to the total number of observations, by precipitation type in this research. The false alarm ratio is the number of correct forecasts divided by the total number of forecasts.

3. Results and Discussion of the DAMP Model Development

There are four basic temperature profiles during the winter for both Greensboro (GSO), NC, and Washington Dulles Airport (IAD), VA—zero, one, two, and three or more freezing level(s) (Figure 2). Each of

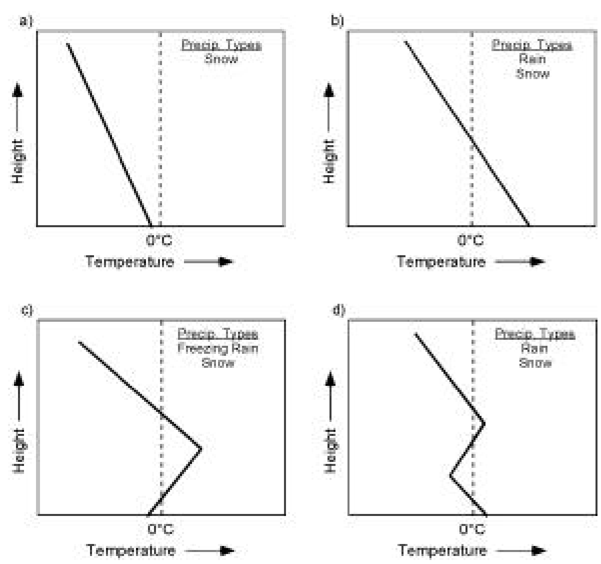

Figure 2. The four vertical temperature profiles associated with winter precipitation types.

these profiles produces characteristic types of precipitation (snow, freezing rain, freezing rain mix, and rain). Note that pure ice pellets were not included as a precipitation type, and, as a result, the models were not trained on pure ice pellets, since they occurred infrequently and were largely non-disruptive when compared with freezing rain in the region. From an operational perspective, it is more important to correctly forecast any precipitation, whether pure or a mix, that contained freezing rain or pure snow than to correctly forecast pure ice pellets. Unlike other precipitation types, ice pellets prove difficult to isolate (Keeter and Cline 1991; Bocchieri 1980; Czys et al. 1996; Zerr 1997). Whereas the majority of observations of other precipitation types could be isolated to one or two sounding types, the observations of ice pellets were evenly distributed across the scenarios. Additionally, the statistical analyses could rarely predict the occurrence of ice pellets and tended to include ice pellets with the precipitation type that had the most observations within each profile. That, in turn, weakened the models' abilities to correctly predict the other precipitation types.

Site-specific models (decision trees) were then built by separating the vertical temperature profiles into the four different freezing level categories and performing statistical analyses within each category (Figure 3). The trees showed the forecaster the decisions that are made during execution of the model. A complete discussion of the DAMP model for Washington Dulles Airport (IAD), VA, follows; Greensboro (GSO), NC, is mentioned when the models were different.

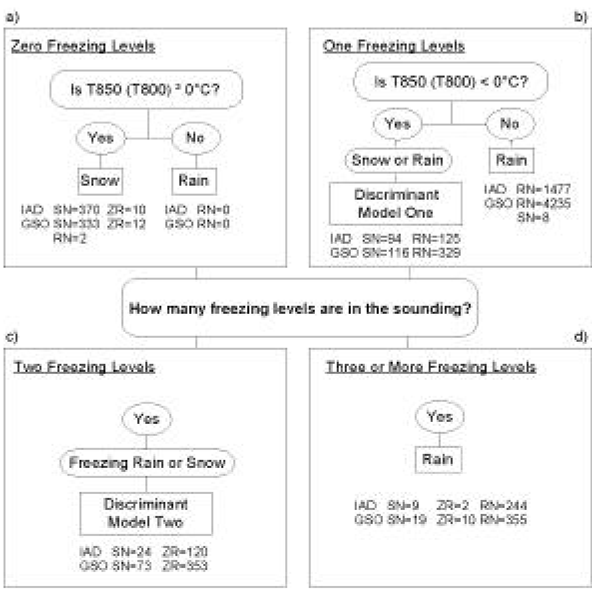

Figure 3. The DAMP model decision tree for Washington Dulles Airport (IAD) and Greensboro (GSO). The trees are identical for both stations with the exception of the observations of each precipitation type (RN = rain, ZR = freezing rain, and SN = snow) isolated by each scenario, and that the critical temperature (denoted in parentheses) is different for Greensboro.

3.1. Zero Freezing Levels

The decisions made by the DAMP model in the zero freezing levels category (Figure 2a) do not require the use of discriminant analysis. To forecast precipitation type in this section of the model, the only variable required was a temperature from somewhere in the sounding to determine whether the entire atmospheric column was warmer or colder than 0°C. The 850 hPa temperature (IAD) and 800 hPa temperature (GSO) were selected for this purpose because these temperatures (subsequently referred to as the critical temperatures) were the most important for discrimination in the one crossing and two crossing parts of each model and were, for simplicity, used in this situation.

In the zero freezing level category, there were no winter precipitation cases where the entire sounding (1000–500 hPa) was warmer than 0°C for either station (Figure 3a). Conversely, when the entire column was 0°C or colder at IAD, two precipitation types were observed—snow (n=370) and freezing rain (n=10). The results for GSO were similar (333 observations of snow and 12 observations of freezing rain) except that two observations of rain entered this grouping. These observations of snow represented 74.4% and 60.0% of all the snow cases observed at IAD and GSO, respectively. As a result, the criteria for membership in this category were very effective in the isolation of snow. While other types of precipitation do occur in this scenario, the DAMP model was trained to forecast snow that "climatologically speaking," occurred 97.4% (IAD) and 96.0% (GSO) of the time.

While the two rain observations at GSO were likely due to observational or measurement errors, there were several possible explanations for the occurrence of freezing rain when the entire column is below freezing. The most likely culprits are a shallow moist layer that does not extend to the level where dendritic growth is possible (temperatures of –12°C to –15°C), supercooled water making it to the surface, or a combination of both. Czys et al. (1996) and Zerr (1997) both noted that if the hydrometeors are relatively small, then freezing drizzle could occur even if the entire sounding was colder than 0°C because cloud and fog droplets (diameter < 100µm) have very low freezing probabilities. The methodology incorporated by Huffman and Norman (1988) successfully segregated these "Supercooled Warm Rain Process" (SWRP) cases using sounding moisture and temperature variables. However, SWRP occurred more often (30%–40%) at the stations in their research. Freezing precipitation in an entirely subfreezing atmosphere accounted for only 0.4% of all cases and 8.7% of freezing rain cases for IAD (0.2% and 3.5% for GSO). Therefore, since SWRP requires the use of moisture variables not incorporated into other nodes of the DAMP models and the number of SWRP cases was few, that distinction was not made in the current research.

3.2. One Freezing Level

The one freezing level grouping included soundings where the surface temperature was greater than 0°C, became sub-freezing at some point above the surface, and remained sub-freezing (Figure 2b). For IAD and GSO, 1,697 and 4,688 observations enter this section of the model. This is approximately 69% and 80%, respectively, of all observations. Again, in this part of the DAMP model the first decision to make was whether the critical temperature was greater than or less than 0°C. Over the 33–years of IAD observations, when the critical temperature was higher than 0°C (Figure 3b) the only type of precipitation observed was rain (n=1477). This was nearly 80% of all the rain cases for IAD. Likewise, for GSO a critical temperature greater than 0°C isolated 4235 rain cases (86%) and only eight snow observations (Figure 3b). The likelihood of receiving rain in this scenario was 99.8%. As a result, the IAD and GSO DAMP models were both programmed to forecast rain in this situation—a decision made using climatology rather than discriminant analysis.

However, if the critical temperatures were subfreezing then 125 (329) rain and 94 (116) snow cases resulted at IAD (GSO) (Figure 3b). Since both rain and snow were common in this instance, discriminant analysis was used to resolve the precipitation type. The variables that the models required for the analyses were as follows: 1) the critical temperature and 2) the height of the freezing level. By implementing discriminant analysis, precipitation type PODs within this section of the model were 81.8% and 87.4% for IAD and GSO (Tables 1 and 2).

3.3. Two Freezing Levels

The two freezing level category (Figure 3c) isolated nearly all the occurrences of freezing rain during the period of record for IAD (120 observations or 91%) and GSO (353 observations or 94%). Along with the freezing rain, there were 24 observations of snow that accounted for about 5% of all the snow hours for IAD. The 73 observations at GSO represented 13% of the total snow observations. The DAMP models incorporated discriminant analysis in this situation to distinguish between precipitation types. The variables determined to be the best predictors were the critical temperature and the heights of the first and second freezing levels. The critical temperature at this level (i.e. 850 hPa for IAD and 800 hPa for GSO) tended to fall between the two freezing levels and, as a result, this critical temperature was a good indicator of the amount of warm air available for melting.

Without the aid of a model, this is a very difficult forecast, because the solution cannot be resolved in two dimensions (i.e., in each bivariate pairing, there are considerable regions of uncertainty). But, by incorporating discriminant analysis to examine the data

Table 1. Classification results for Washington Dulles Airport DAMP model for soundings with one freezing level cases when the 850-mb temperature is below 0°C. Percentage correct/incorrect, in parentheses, are POD/FAR, respectively.

	Forecast		
Observed	Snow	Rain	Total
Snow	76 (78.3%)	18 (19.1%)	94
Rain	21 (16.8%)	104 (85.2%)	125
Total	97	122	219

Table 2. Classification results for Greensboro DAMP model for soundings with one freezing level cases when the 800-mb temperature is below 0°C. Percentage correct/incorrect, in parentheses, are POD/FAR, respectively.

	Forecast		
Observed	Snow	Rain	Total
Snow	89 (75.4%)	27 (23.3%)	116
Rain	29 (8.8%)	300 (91.7%)	329
Total	118	327	445

Table 3. Classification results for Washington Dulles Airport DAMP model for soundings with two freezing levels. Percentage correct/incorrect, in parentheses, are POD/FAR, respectively.

	Forecast		
Observed	Snow	Freezing Rain	Total
Snow	5 (50.0%)	19 (79.2%)	24
Freezing Rain	5 (4.2%)	115 (85.8%)	120
Total	10	134	144

Table 4. Classification results for Greensboro DAMP model for soundings with two freezing levels (Percents = POD/FAR).

	Forecast		
Observed	Snow	Freezing Rain	Total
Snow	22 (73.3%)	51 (69.8%)	73
Freezing Rain	8 (2.3%)	345 (87.1%)	353
Total	30	396	426

in multi-dimensional space, the separation between precipitation types increased. The results of the discriminant analyses at this node showed that freezing rain was forecast correctly 85.8% and 87.1% of the time for IAD and GSO, respectively (Tables 3 and 4). The GSO DAMP model correctly predicted snow in 73.3% of the cases (Table 4). Although, snowfall forecasts within this part of DAMP were correct only 50% of the time for IAD (Table 3), very little snow occurred under these conditions and the increased ability to predict freezing rain far outweighed the ramifications of missing snow.

3.4. Three or More Freezing Levels

The final category of the DAMP models was the three or more freezing level cases. Only 10.3% of the IAD precipitation observations (6.1% of the GSO observations) occurred when there were three or more points in the vertical temperature profile that crossed 0°C (Figure 3d). Again, climatology influenced the decisions the DAMP models were programmed to make under this scenario. In this three–crossing case the DAMP models were forced to predict rain because it occurred 96% and 93% of the time for IAD and GSO, respectively.

3.5. Overall DAMP Model Performance

The overall classification performance results for both DAMP models indicated that a high percentage of correct precipitation type forecasts may be achieved by dividing the observations into the various profiles and as a result isolating distinct precipitation types (Tables 5 and 6). Rain forecasts were approximately 98% correct for both models. The IAD DAMP model snow forecasts were 92.6% correct and freezing rain forecasts were 85.8% correct (Table 5). Snow and freezing rain forecasts were 89.7% and 87.1% correct, respectively, for the GSO DAMP model. Neither model ever forecasts freezing rain when rain actually occurred. Further, the results indicated that when freezing rain was actually observed, the models missed the forecast by predicting rain in only 0.1% and 0.2% of the cases for IAD (Table 5) and GSO (Table 6), respectively. From an operational and public safety point of view, these results are important because the occurrence of freezing rain when rain is the forecast is very dangerous. Fortunately, the DAMP model performs very well in such disruptive and hard-to-predict situations.

Table 5. The overall classification results for the Washington Dulles Airport DAMP Model. Percentage correct/incorrect, in parentheses, are POD/FAR, respectively.

Observed	Forecast: Rain	Forecast: Freezing Rain	Forecast: Snow	Total
Rain	1825 (98.4%)	0 (0.0%)	21 (1.1%)	1846
Freezing Rain	2 (1.5%)	115 (85.5%)	15 (11.1%)	132
Snow	27 (5.5%)	19 (3.8%)	451 (92.6%)	497
Total	1854	134	487	2475

Table 6. The overall classification results for the Greensboro DAMP Model. Percentage correct/incorrect, in parentheses, are POD/FAR, respectively.

Observed	Forecast: Rain	Forecast: Freezing Rain	Forecast: Snow	Total
Rain	4890 (98.7%)	0 (0.0%)	31 (0.6%)	4921
Freezing Rain	10 (2.7%)	345 (87.1%)	20 (5.3%)	375
Snow	54 (9.8%)	51 (9.3%)	444 (89.7%)	549
Total	4954	396	495	5845

5. Conclusions

The methodology incorporated into the Discriminant Analysis Mixed Precipitation (DAMP) model development used a combination of climatology and discriminant analysis to create a very effective winter precipitation type forecasting tool, providing winter precipitation type guidance in a region where such forecasts are very difficult. Freezing level height(s), vertical temperature profiles, and basic meteorology are entered to resolve the forecasts of snow, rain and freezing rain.

The DAMP models made better use of the available data and provided the ability to isolate and provide the forecaster with guidance in the form of probabilities of detection (POD) for each precipitation type in scenarios where predicting precipitation type was normally a very difficult decision. That represents a major advantage over other current techniques. These multidimensional, physically based models proved easy for the forecaster to use: the model-decision-making process was easy to follow in virtually every case. Most importantly, the models did an outstanding job of predicting precipitation type and duration in mixed-precipitation events. Besides achieving very high correct prediction rates for the precipitation types modeled, the most noteworthy advantage of the DAMP models' was that they rarely missed the forecast of freezing rain—of particular importance for the credibility of the operational forecast.

Although the DAMP models were developed for locations (GSO and IAD) where upper air data and surface data were co-located, very preliminary results indicated that they may retain a high degree of accuracy when applied at other locations in similar climatological regions. The IAD DAMP model has been successfully applied to events occurring in Piedmont and Tidewater locations in Virginia and North Carolina. The GSO DAMP model was suitable for higher elevation stations in the Blue Ridge Mountains and foothills. However, because the site-specific nature of the models could present problems when applied at remote locations, future research must examine and attempt to correct this complication.

The duration of each precipitation category can be determined by successively applying the DAMP model to the output of each time step of any forecast model. The DAMP model's dependence on the model output in an operational forecasting mode is known as a "perfect prog" approach—one that assumes that the model output/prognosis is perfect. An inherent problem with this approach is that errors in the numerical models enter DAMP. If the modeled heights and temperatures are identical to the observed heights and temperatures, then DAMP's performance should be unaffected. Alternatively, if the numerical forecast models are erroneous, several outcomes are possible depending on the degree to which the models depart from the actual observations and whether they occur in the borderline regions (in multidimensional space) between precipitation types. The DAMP precipitation type prognosis can be correct if the error in the numerical forecast model is within the resolution of DAMP. DAMP will be wrong if the modeled heights and temperatures erroneously output heights and temperatures that are associated with another precipitation type or if the error is outside of DAMP's resolution. The degree to which these errors effect the DAMP models' results will vary from model run to model run and must be assessed by the operational forecaster.

Another complication with the DAMP model that is also common to other current generation models was its inability to isolate and therefore forecast pure ice pellets. Future iterations of this model will address this problem by attempting to identify the parameter(s) that isolate ice pellets from other types of winter precipitation.

In addition to correcting problems, more refinements are planned for future iterations of the DAMP model. The first of these refinements is to determine if there are some unexplored variables that may help improve the model. Incorporation of upper-level winds and moisture variables are being considered as a possibility. Another change that may help improve the DAMP model is to identify a critical temperature (e.g., maximum or minimum temperature in certain layers) that is no longer tied to a pressure surface but perhaps to some other dynamic feature. Additionally, further automation of the DAMP models that allows for direct incorporation of numerical weather model output will provide for easier use by the operational forecaster.

6. Acknowledgements

This research was conducted under the Cooperative Program for Operational Meteorology, Education, and Training (COMET) grant UCAR SS96-73653. Special thanks are extended to the Wakefield, VA, National Weather Service Forecast Office staff for their comments regarding the project and for numerous interactions that allowed us to become more familiar with the winter forecast process from their perspective, and to Robert E. Davis for his statistical advice, insight concerning winter weather, and suggestions regarding the direction of the project.

7. References

Bocchieri, J.R., 1980: The objective use of upper-air soundings to specify precipitation type. *Mon. Wea. Rev.*, **108,** 596–603.

Czys, R.R., R.W. Scott, K.C. Tang, R.W. Przybylinski, and M.E. Sabones, 1996: A physically based, non-dimensional parameter for discriminating between locations of freezing rain and sleet. *Wea. Forecasting,* **11,** 591–598.

Fisher, R.A., 1936: The use of multiple measurement in taxonomic problems. *Annals of Eugenics,* **7,** 179–188.

Gay, D.A., and R.E. Davis, 1993: Freezing rain and sleet climatology of the southeastern USA. *Clim. Research,* **3,** 209–220.

Keeter, K.K., and J.W. Cline, 1991: The objective use of observed and forecast thickness values to predict precipitation type in North Carolina. *Wea. Forecasting,* **6,** 456–469.

——, S. Businger, L.G. Lee, and J.S. Waldstreicher, 1995: Winter weather forecasting throughout the Eastern United States. Part III: The effects of topography and the variability of winter weather in the Carolinas and Virginia. *Wea. Forecasting,* **0,** 1–19.

Lemon, P.C., 1961: Forest ecology of ice storms. *Bull. Torrey Botanical Club,* **88,** 21–29.

Michaels P.J., P.J. Stenger, P.C. Knappenberger, and D.E. Stooksbury, 1991: Virginia ice storms. *Southeastern Climate Review.,* **2,** 11–17.

Zerr, R.J., 1997: Freezing rain: an observational and theoretical study. *J. Appl. Meteor.,* **36,** 1647–1661.