U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
SYSTEMS DEVELOPMENT OFFICE
TECHNIQUES DEVELOPMENT LABORATORY

TDL OFFICE NOTE 82-2

RECENT EXPERIMENTS IN THE USE OF MODEL OUTPUT STATISTICS FOR FORECASTING SNOW AMOUNTS

Joseph R. Bocchieri

February 1982

RECENT EXPERIMENTS IN THE USE OF MODEL OUTPUT STATISTICS FOR FORECASTING SNOW AMOUNTS

Joseph R. Bocchieri

1. INTRODUCTION

A Model Output Statistics (MOS) system for forecasting heavy snow has been operational within the National Weather Service since October 1977 (Bocchieri, 1979a); National Weather Service, 1978). Heavy snow is defined as a fall of $\geq$4 inches during a 12-h period at a station. In MOS (Glahn and Lowry, 1972), a statistical relationship is determined between the forecast output of a numerical prediction model (or models) and observed occurrences of a particular weather element. For the operational heavy snow system, we used output from both the Limited-area Fine Mesh (LFM) model (National Weather Service, 1971; Gerrity, 1977) and a finer mesh version of the LFM, called LFM-II (National Weather Service, 1977; Newell and Deaven, 1981), to develop prediction equations that give the conditional probability of heavy snow for 12-24 h periods after 0000 GMT and 1200 GMT. An estimate of the unconditional probability of heavy snow and a categorical forecast are also provided. The unconditional probability is derived by taking the product of the conditional probability, the probability of precipitation (PoP) (Lowry and Glahn, 1976; National Weather Service, 1980), and the conditional probability of frozen precipitation (PoF) (Bocchieri, 1979b).

We've recently derived a set of experimental snow amount forecast equations with the goal of developing a new operational system which would provide forecasts for other snow amount categories, not just heavy snow, and other forecast projections other than 12-24 hours. In the process of deriving these equations, we performed a number of experiments. In one experiment, we determined the effect of removing LFM boundary-layer type predictors from the forecast equations, since future operational numerical models at the National Meteorological Center may not include a boundary layer in the form presently used in the LFM. The results indicate that forecast equations that didn't include boundary-layer predictors were only slightly less accurate than equations that included boundary-layer predictors. In another experiment, we determined the optimum number of predictors to include in the forecast equations and found that 12 predictors was about right.

Conditional probability of heavy snow forecasts made by the experimental equations were then compared to similar forecasts made by the operational system; the results indicate that there was little difference in the accuracy of the two systems. This was encouraging because the experimental equations didn't include boundary-layer type predictors.

After performing these experiments for the 12-24 h projection, we developed conditional probability equations for several snow amount categories for the 12-18, 18-24, and 24-36 h projections from 0000 GMT. We then compared two methods for estimating the unconditional probability of occurrence of the snow amount categories, given the conditional probabilities from the experimental system. One method, called PRODUCT, consisted simply of multiplying the conditional probabilities for each snow amount category by the PoP for the 12-h period and the average PoF for the same 12-h period. This is quite similar to the method used in the operational system. The other method, called MOSSQR, involved the derivation of regression equations to predict the unconditional probabilities. Potential predictors for these equations were conditional snow amount probabilities, PoP and PoF forecasts, and various products thereof. A comparative verification showed there was little difference between the two methods. Hence, the PRODUCT method was then used to estimate the unconditional probabilities of the various snow amount categories and probability threshold values were determined on the developmental sample so that categorical snow amount forecasts could be made. We verified the categorical forecasts for each projection on both developmental and independent data. The results indicate the scores generally deteriorated on the independent sample, especially for the 18-24 and 24-36 h projections. We also compared the experimental forecasts with operational forecasts for the ≥4 inch category and found that the two systems were of comparable accuracy.

Based on these results, we decided to continue the development and testing of forecast equations for the 12-24 h projection. We concluded that the accuracy of the forecasts for the other projections was not good enough to warrant further development efforts at this time.

2. DEVELOPMENT AND TESTING OF CONDITIONAL SNOW AMOUNT FORECAST EQUATIONS

The potential predictors from the LFM and LFM-II models used in the development of experimental, conditional probability of snow amount [PoSA(S)] equations are listed in Table 1. Model output variables valid for 6-, 12-, 18-, 24-, 36-, 42-, and 48-h projections were included in addition to station elevation and the sine and cosine of the day of year. The table also gives the acronyms by which the various predictors will be referred in this paper. The predictors were used in both unsmoothed and 9-point, space-smoothed form.

To form the predictand for the PoSA(S) equations, we included only "pure snow" events in our developmental sample. A pure snow event is defined as the occurrence at a station of ≥0.1 inches of snow and/or sleet, and no other type of precipitation during a 6- or 12-h period depending on the projection. Therefore, the PoSA(S) equations produce probability forecasts for a site given that a pure snow event occurs.

Actually, we could isolate only quasi-pure snow events in the developmental sample. The basic data were 6-h snowfall amounts at 195 stations in the conterminous United States. In order to isolate the pure snow events we examined the "weather" observations to determine the type of precipitation falling within the period. However, weather observations were available only every third hour. Therefore, since we couldn't be certain that only pure snow fell within the 12-h period, a number of snow events may have been only quasi-pure.

We used the Regression Estimation of Event Probability (REEP) screening technique (Miller, 1964) to develop PoSA(S) forecast equations. This technique objectively selects a subset of effective predictors from a large set of potential predictors to use in multiple linear regression equations. The equations give estimates of the probabilities of occurrence of a given set of binary-type predictands. As discussed later, snow amount was categorized into cumulative, binary-type predictands, ≥1, ≥2, ≥3, ≥4, and ≥6 inches. The predictand is called binary because in the developmental phase it was assigned a value of 1 or 0 in a given case depending, respectively, upon whether or not that particular snow amount category occurred. The potential predictors in Table 1 were included in both binary and continuous form. The use of binary predictors helps to account for non-linear relationships which may exist between the predictors and predictands. A good description of the screening procedure can be found in Glahn and Lowry (1972).

a. Preliminary Experiments

In the process of developing the experimental PoSA(S) equations, we did a number of experiments. We first determined the effect of removing LFM boundary-layer predictors from the PoSA(S) equations. Two PoSA(S) forecast systems were developed with data combined from 195 conterminous United States stations and seven winter seasons, September through April, 1972-73 through 1978-79. One system included boundary-layer predictors, among others, and the other system didn't. The REEP technique was used in conjunction with the potential predictors in Table 1 to develop 12-term equations for each system. The two PoSA(S) systems forecast the probability of ≥ 4 inches of snow, given that snow occurs, in the 12-24 h period after 0000 GMT.

We did a comparative verification between the two PoSA(S) systems with independent data from the winter season of 1979-80. The results show that the Brier score (Brier, 1950) for the PoSA(S) system which included boundary-layer predictors was about 1% better than the Brier score of the system that didn't have these predictors. We judged that this improvement was not large enough to include boundary-layer predictors in further development of PoSA(S) equations, since these fields may not be available from future numerical models run at the National Meteorological Center.

In another experiment, we determined the optimum number of predictors to include in PoSA(S) equations. First, PoSA(S) equations were derived for the 12-24 h period after 0000 GMT for each of several geographic regions. The developmental sample consisted of data from 195 stations and eight winter seasons (1972-73 through 1979-80). The regions were determined in the following manner. We derived PoSA(S) equations for the ≥ 1 and ≥ 4 inch categories for the 12-24 h period from both 0000 GMT and 1200 GMT with data combined from 195 stations for the developmental sample--the so-called generalized operator approach. We then evaluated the equations to obtain forecasts for each station on the developmental sample. A statistic called the relative probability bias was computed for each station and for each snow amount category by:

$$\text{Relative Probability Bias} = \frac{\overline{\text{PoSA(S)}} - \text{RF}}{\text{RF}}, \qquad (1)$$

where $\overline{\text{PoSA(S)}}$ is the average conditional probability forecast for a particular snow amount category for each station and RF is the relative frequency of that category for each station for the developmental sample. We subjectively determined the regions shown in Fig. 1 by grouping stations having similar relative probability bias characteristics; other factors we considered were the density of stations and the climatic frequency of snow amount. We needed to make the regions large to insure that a sufficient number of snow amount cases would be available for equation development.

After determining the regions, separate PoSA(S) equations sets containing 1, 2, 4 ..., 20 predictors were derived for ≥1, ≥2, ≥3, ≥4, and ≥6 inch categories for each region. We then computed Brier scores for forecasts from each equation set on independent data combined for 195 stations from the winter seasons of 1980-81. The results in Table 2 indicate that the improvement (decrease) in the Brier score as the number of predictors increases depends on the snow amount category. For instance, for the ≥1 inch category, the 16-predictor equation improved upon the 1-predictor equation by 5.7%. For the ≥3 inch category, the 10-predictor equation improved upon the 1-predictor equation by 1.5%. However, for the ≥6 inch category, the number of predictors in the equations made little difference in the Brier score. Overall, there was little improvement in the Brier scores beyond about 12 predictors and we decided this number was optimum. These results generally agree with those found by other investigators for other forecast variables (Annett et al., 1972; Bocchieri and Glahn, 1972; Zurndorfer and Bermowitz, 1976).

We then compared forecasts from 12-term PoSA(S) equations for the ≥4 inch category with forecasts from the operational conditional probability of heavy snow system. The Brier score was computed for both systems for an independent sample consisting of about 1400 cases from the winter season of 1980-81. The results indicate the experimental forecasts were better than the operational heavy snow forecasts by 1.4% in the Brier score. Although the level of improvement is small, it should be noted that boundary-layer type predictors from the LFM were not included in the experimental equations, but they are used in the operational system. Experiments described previously in this paper indicate that equations without boundary-layer predictors were about 1% worse than equations with boundary-layer predictors. Based on this comparison, we decided to proceed with development of experimental PoSA(S) equations for other forecast projections.

b. Further Development of Experimental PoSA(S) Equations

In addition to deriving experimental equations for the 12-24 h projection, we also developed PoSA(S) equations for the 12-18 h, 18-24 h, and 24-36 h projections from 0000 GMT. For the two 6-h periods, three categories of snow amount were were: ≥1, ≥2, and ≥3 inches. For the 24-36 h projection, we used the same snow amount categories as for the 12-24 h period; ≥1, ≥2, ≥3, ≥4, and ≥6 inches. The potential predictors in Table 1 were used to develop PoSA(S) equations for these additional projections for each of the regions in Fig. 1. The developmental sample for the 12-18 h and 18-24 h periods consisted of data from eight winter seasons, 1972-73 through 1979-80. For the 24-36 h projection, we had about 4 1/2 seasons of data available, February 1976 through April 1980.

Table 3 shows the total reduction of variance for 12-term PoSA(S) equations for each region for the 12-24 h projection from 0000 GMT. The relative frequency for each snow amount category and the total number of pure snow cases are also given. As expected, the reduction of variance was generally lower over the western, more mountainous portion of the United States (regions 1 and 2); it's generally more difficult to predict snow amounts in mountainous areas because of local, orographic effects. Also, as expected, the reduction of variance was generally lower for the higher, less frequent snow amount categories. However, in regions 3, 4, and 5 the reduction of variance for the ≥2 inch category was higher than for the ≥1 inch category even though the later category occurred much more frequently. A possible reason is that those weather systems producing at least 2 inches of snow are more organized, occur on a larger scale, and are easier to predict than weather systems producing lesser snow amounts.

In Table 4, the predictor types are ranked as determined by the REEP screening procedure for the 12-24 h and 24-36 h projections from 0000 GMT. This list was determined by both frequency and order of selection; for the purpose of this ranking, all predictor prjections, smoothings, and binary limits were combined for each type of variable. The results indicate the LFM forecasts of P AMT and MEAN R HUM were the most important predictors for both projections. Various parameters derived at 850-mb, such as 850 DIV, 850 VORT, and 850 M CONV, ranked next in importance at 12-24 hours, and 850 U and 850 DIV ranked next in importance at 24-36 hours. Other investigators such as Browne and Younkin (1970), Brandes and Spar (1971), and Spiegler and Fisher (1971) also found the 850-mb level to be useful for snow amount prodiction. Variables at 700-mb, such as 700 U and 700 VV followed the 850-mb parameters for 12-24 hours. For the 24-36 h projection, the 500 VORT ranked next followed by 700 U. Goree and Younkin (1966) and Weber (1978) also found various parameters at 500-mb to be useful. 10-5 THICK ADV ranked relatively high for the 12-24 h projection; Younkin (1968) found strong warm advection to be associated with heavy snow in the western United States.

3. EXPERIMENTS IN ESTIMATING THE UNCONDITIONAL PROBABILITY OF SNOW AMOUNT

We experimented with two methods for obtaining unconditional probability of snow amount, PoSA, forecasts for the 12-24 h period after 0000 GMT. One method, called PRODUCT, consists simply of multiplying the conditional probability forecast, PoSA(S), for each snow amount category, by the PoP for the corresponding 12-h period and the average PoF ($\overline{\text{PoF}}$) for the same 12-h period. This is expressed mathematically[1], for instance for the ≥ 1 inch category, by:

$$\text{PoSA}(\geq 1 \text{ inch}) = \text{PoSA(S)}(\geq 1 \text{ inch}) \times \text{PoP} \times \overline{\text{PoF}}. \qquad (2)$$

This method is essentially the one used in the operational system to obtain estimates of the unconditional probability of heavy snow.

The other method, called MOSSQR, involved derivation of REEP equations to predict PoSA. Potential predictors for these equations were PoSA(S), PoP, and $\overline{\text{PoF}}$ forecasts, and various products thereof. The developmental sample for this derivation consisted of the months October through March, 1972-73 through 1979-80. We first developed generalized-operator PoSA equations with data combined from 192 stations for the ≥ 1, ≥ 2, ≥ 3, ≥ 4, and ≥ 6 inch categories. We then attempted to determine regions in a manner similar to that used for the PoSA(S) equations in Section 2. That is, we computed the relative probability bias for each station by Eq. (1) using the developmental sample and then tried to combine stations with similar bias characteristics to determine the regions

---

[1] In a true mathematical sense, this method doesn't give the unconditional probability of snow amount for 12-h periods in which mixed precipitation occurs, since PoSA(S) equations were developed with pure snow cases only.

However, the bias values were rather erratic and we had trouble grouping stations into regions. A possible reason for this is that the predictors used in the PoSA equations, i.e., PoSA(S), PoP, and $\overline{\text{PoF}}$ forecasts, were obtained from systems that had already been regionalized so that much of the bias was implicitly accounted for.

Table 5 shows the 12-predictors included in the MOSSQR prediction equations in the order determined by the REEP procedure. The additional reduction of variance accounted for by each predictor for each snow amount category is also shown. The first predictor chosen, for example, is an estimate of PoSA for the ≥ 1 inch snow amount category and is given by the product of PoSA(S) for ≥ 1 inch, and the PoP and $\overline{\text{PoF}}$ for the 12-24 h period. This predictor was chosen because of its contribution to the reduction of variance of the ≥ 1 inch predictand category (27.4%). Some of the predictors chosen were binary in form. For instance the third predictor takes on the value 1 if the estimate of PoSA for ≥ 2 inches is $\leq 70\%$; otherwise, this predictor is given the value 0.

We did a comparative verification between the PRODUCT and MOSSQR methods for both the developmental and independent data samples. The independent sample consisted of data from October 1980 through March 1981. Table 6 shows the Brier scores for both samples for 192 stations combined and for each snow amount category. The results indicate the following: 1) for the developmental sample, there was little difference between the methods, except for the ≥ 4 and ≥ 6 inch categories where MOSSQR improved upon the PRODUCT method by 1.5% and 4.2%, respectively; and 2) for the independent sample, there was little difference between the PRODUCT and MOSSQR methods for all snow amount categories. Based on these results, we decided to use the PRODUCT method for further development and testing of experimental PoSA equations, since it will be easier to apply this method in operations.

4. VERIFICATION OF CATEGORICAL SNOW AMOUNT FORECASTS

To help us determine whether the experimental snow amount forecast system would be suitable for operational implementation, we examined two factors on independent data: 1) the stability of verification scores computed for categorical snow amount forecasts, and 2) a comparative verification between the experimental and operational system for the categorical heavy snow forecasts.

In order to make categorical snow amount forecasts from the probability forecasts, we developed threshold probability values for each snow amount category, each region, and each forecast projection. This was done by maximizing the threat score[2] for each category while restricting the categorical bias[3] to ≤ 1.30. The threat score is used at the Heavy Precipitation Branch of the National Meteorological Center to verify operational heavy snow forecasts.

---

[2] Threat score = $H/(F + O - H)$ where H is the number of correct forecasts of a category and F and O are the number of forecasts and observations of that category, respectively.

[3] Bias is the number of forecasts of a category divided by the number of observations of that category. A categorical bias equal to 1.00 means unbiased forecasts of that category.

As described by Bermowitz and Zurndorfer (1979), maximizing the threat score requires calculation of a threshold probability that will maximize the threat score for dichotomous forecasts of a category. The threshold probability for a category, say ≥1 inch of snow, is a value that if exceeded by a probability forecast for that category would result in a categorical forecast of ≥1 inch. If the threshold value is not exceeded, the categorical forecast would be <1 inch.

Threshold probabilities were computed with the use of an empirical iterative technique. On each iteration, threat scores were computed for categorical forecasts made by comparing probability forecasts against threshold probabilities. For the initial iteration, a first guess threshold probability was provided. For subsequent iteration, the threshold probability was incremented by a preselected value. Threshold values were chosen so that the categorical bias was ≤1.30, even if the maximum threat score was associated with a bias >1.30.

As as example, assume that the probability forecasts for the 12-24 h period for the categories ≥1, ≥2, ≥3, ≥4, and ≥6 inches are 0.40, 0.30, 0.20, 0.15, and 0.05, respectively. Furthermore, assume that the threshold probabilities that maximize the threat score for each of the five categories are 0.38, 0.32, 0.23, 0.16, and 0.08. The procedure starts at the category ≥1 inch and compares the probability forecast (0.40) with the threshold value for that category (0.38). Since the threshold value is exceeded, at least 1 inch is predicted. The next step is to proceed to the category ≥2 inch and see if the probability forecast (0.30) exceeds the threshold value (0.32) for that category. Since it doesn't, the procedure is terminated and the forecast amount is 1 to 2 inches. Of course, if the probability forecasts had exceeded the threshold values for all categories, the forecast amount would have been ≥6 inches. Similarly, if the probability forecast was less than the threshold value for the category ≥1 inch, the forecast amount would have been <1 inch.

After determining the threshold probability values, we produced and verified categorical snow amount forecasts for both the developmental and independent data samples, which were the same as those used in Section 3. Table 7 shows the threat score and bias for the 12-24, 12-18, 18-24, and 24-36 h forecast projections from 0000 GMT. For the two 12-h periods, we verified five categories of snow amount, ≥1, ≥2, ≥3, ≥4, and ≥6 inches; for the two 6-h periods, we verified three categories, ≥1, ≥2, and ≥3 inches. The results indicate the following:

1. For the developmental sample, the biases were generally between 1.00 and 1.30. The bias was forced to be such when the threshold probability values were derived. For these bias values, the threat scores were between .20 and .30 for the ≥1, >2, and ≥3 inch categories for the 12-24 and 24-36 h projections and for the ≥1 and ≥2 inch categories for the two 6-h periods. The threat scores were generally between .15 and .20 for the remaining categories for each projection.

2. For the independent sample, both the threat score and bias deteriorated as compared to the developmental sample. The deterioration was worse for the 18-24 h and 24-36 h periods than for the 12-18 h and 12-24 h periods. Also, the deterioration was generally worse

for the upper snow amount categories as compared to the lower categories for each forecast period. For instance, for the 12-24 h period, the bias decreased to .83 for ≥1 inch and decreased to between .60 and .65 for ≥2, ≥3, and ≥4; the threat score remained about the same for ≥1 inch, decreased to .19 for ≥2 inch, and decreased to between .10 and .15 for the other categories. However, for the 24-36 h projection, the decrease in the bias was more extreme for the ≥3, ≥4, and ≥6 inch categories as compared to the 12-24 h period; the decrease in the threat scores was also more extreme, and, in fact, the threat scores were near 0 for the ≥3, ≥4, and ≥6 inch categories. A similar scenario can be noted when results for the 18-24 h period are compared to results for the 12-18 h period.

We also compared the experimental system to the operational system for the ≥4 inch category for the 12-24 h period after 0000 GMT. The bias and threat score for the operational system were .69 and .15, respectively, while, as shown in Table 7, the scores for the experimental system were .64 and .15, respectively. The two systems, therefore, were of comparable accuracy. Recall that when we compared the experimental and operational systems for conditional probability of heavy snow forecasts in Section 2, we also found the results to be similar.

5. SUMMARY AND CONCLUSIONS

A MOS system for forecasting heavy snow (≥4 inch in a 12-h period) has been operational within the National Weather Service since October 1977. For that system, we used LFM model output to develop prediction equations for the conditional probability of heavy snow during the 12-24 h period after 0000 GMT and 1200 GMT. Estimates of the unconditional probabilities of heavy snow and categorical forecasts are also provided.

We've recently done several experiments with the goal of developing an improved operational system which would provide forecasts for other snow amount categories, not just heavy snow, and other forecast projections, in addition to the 12-24 h projection. First, we determined the effect of removing LFM boundary-layer predictors from experimental conditional probability of snow amount [PoSA(S)] equations. Two PoSA(S) forecast systems were developed with the REEP screening procedure. One system included boundary layer predictors and the other didn't. A comparative verification between the two systems on independent data shows the system which included boundary-layer predictors was about 1% more accurate than the system that didn't have these predictors. We judged that this improvement was not large enough to include boundary-layer fields in further development of PoSA(S) equations, since these predictors may not be available from future numerical models run operationally at the National Meteorological Center.

In another experiment, we determined the optimum number of predictors to include in PoSA(S) equations. Regionalized equations were developed for ≥1, ≥2, ≥3, ≥4, and ≥6 inch snow amount categories for the 12-24 h period after 0000 GMT. Separate equation sets were derived containing 1, 2, 4, ..., 20 predictors. We then verified probability forecasts from each equation set on independent data. The improvement in the Brier score as the number of predictors increased depended on the snow amount category. Overall, there was little improvement in the Brier scores beyond about 12 predictors.

We then compared forecasts from 12-term PoSA(S) equations for the ≥4 inch category with forecasts from the operational conditional probability of heavy snow system. The Brier score computed for forecasts from both systems for an independent sample indicate that the experimental forecasts were about as accurate as the operational forecasts, in spite of the fact that the experimental equations didn't include boundary-layer type predictors while the operational equations did. We therefore proceeded with the development of experimental PoSA(S) equations for other projections.

PoSA(S) equations were developed for the 12-18 h, 18-24 h, and 24-36 h projections from 0000 GMT. An analysis of the predictor types chosen by the REEP screening procedure for the 12-24 h and 24-36 h projections showed that LFM forecasts of precipitation amount and surface to 500-mb mean relative humidity were the most important predictors for both projections. Various parameters derived at the 850-mb level, such as divergence, vorticity, moisture convergence, and wind components ranked next in importance. Similar variables at the 700-mb and 500-mb levels followed.

We also expermented with two methods for obtaining unconditional probability of snow amount, PoSA, forecasts for the 12-24 h period after 0000 GMT. One method, called PRODUCT, consists simply of multiplying PoSA(S) forecasts for each snow amount category by the PoP for the corresponding 12-h period and the average conditional probability of frozen precipitation, $\overline{\text{PoF}}$, for the same 12-h period. This is quite similar to the method used in the present operational heavy snow system. Another method, called MOSSQR, involved derivation of regression equations to predict PoSA. Potential predictors for these equations were PoSA(S), PoP, and $\overline{\text{PoF}}$ forecasts, and various products thereof. A comparative verification between PRODUCT and MOSSQR on independent data indicates there was little difference in Brier scores for the two systems. Hence, we decided to use the PRODUCT method to estimate PoSA, since it will be easier to put into operation.

To help us determine whether the experimental snow amount forecast system would be suitable for possible operational implementation, we verified categorical snow amount forecasts on independent data. First, we used the PRODUCT method to estimate PoSA for each projection, each region, and for both the developmental and independent data samples. The PoSA forecasts were then transformed into categorical snow amount forecasts; to do this, we derived threshold probability values for each snow amount category, each region, and each forecast projection using the developmental sample. The threshold values were chosen such that the threat score was maximized while restricting the bias to ≤1.30. We then computed the threat score and bias for categorical snow amount forecasts for the independent data sample. The results indicate that both the threat score and bias deteriorated as compared to these scores for the developmental sample. The deterioration was worse for the 18-24 h and 24-36 h periods than for the 12-18 h and 12-24 h periods. Also, the deterioration in the scores was generally worse for the higher snow amount categories compared to the lower categories for each forecast period. We also compared the threat score and bias for the ≥4 inch category for the experimental system to the same scores computed for the operational heavy snow system. The results indicated that the two systems were of comparable accuracy.

Based on the verification of the categorical forecasts, we concluded that the experimental forecasts for the 12-18, 18-24, and 24-36 h periods were not good enough to consider operational implementation of snow amount forecasts for these periods at this time. On the other hand, we think that the results for the 12-24 h period are promising enough to warrant further development of snow amount forecast equations for this period for possible operational implementation, especially since test results show the experimental forecasts for the heavy snow category are of comparable accuracy to the present operational system. We plan to combine the developmental and independent data samples to redevelop the snow amount forecast equations for the 12-24 h projection and then retest the equations on a different sample. The results of this new development and testing will be reported later.

6. ACKNOWLEDGEMENT

The author wishes to thank the many members of the Techniques Development Laboratory who developed and maintain the MOS system.

7. REFERENCES

Annett, J. R., H. R. Glahn, and D. A. Lowry, 1972: The use of model output statistics (MOS) to estimate daily maximum temperatures. NOAA Tech. Memo. NWS TDL-45, NOAA, U.S. Department of Commerce, 14 pp.

Bermowitz, R. J., and E. A. Zurndorfer, 1979: Automated guidance for predicting quantitative precipitation. Mon. Wea. Rev., 107, 122-128.

Bocchieri, J. R., and H. R. Glahn, 1972: Use of model output statistics for predicting ceiling height. Mon. Wea. Rev., 100, 869-879.

_____, 1979a: The use of LFM output for automated prediction of heavy snow. Preprints Fourth Conference on Numerical Weather Predication, Silver Spring, Amer. Meteor. Soc., 77-81.

_____, 1979b: A new operational system for forecasting precipitation type. Mon. Wea. Rev., 107, 637-648.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. Mon. Wea. Rev., 78, 1-3.

Brandes, E. A., and J. Spar, 1971: A search for necessary conditions for heavy snow on the east coast. J. Appl. Meteor., 10, 397-409.

Browne, R. F., and R. J. Younkin, 1970: Some relationships between 850 millibar lows and heavy snow occurrences over the central and eastern United States. Mon. Wea. Rev., 98, 399-401.

Gerrity, J. F., Jr., 1977: The LFM model-1976: a documentation. NOAA Tech. Memo. NWS NMC-60, NOAA, U.S. Department of Commerce, 68 pp.

Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. J. Appl. Meteor., 11, 1203-1211.

Goree, P. A., and R. J. Younkin, 1966: Synoptic climatology of heavy snowfall over the central and eastern United States. Mon. Wea. Rev., 94, 663-668.

Lowry, D. A., and H. R. Glahn, 1976: An operational model for forecasting probability of precipitation--PEATMOS PoP. Mon. Wea. Rev., 104, 221-231.

Miller, R. G., 1964: Regression estimation of event probabilities. Tech. Rep. No. 1, Contract CWB-10704, The Travelers Research Center, Inc. Hartford, Conn., 153 pp. [NTIS AD 602037].

National Weather Service, 1971: The limited-area fine mesh (LFM) model. NWS Tech. Proc. Bull., No. 67, NOAA, U.S. Department of Commerce, 11 pp.

_____, 1977: High resolusion LFM (LFM-II). NWS Tech. Proc. Bull., No. 206, NOAA, U.S. Department of Commerce, 6 pp.

_____, 1978: The use of model output statistics for predicting the probability of heavy snow. NWS Tech. Proc. Bull., No. 246, NOAA, U.S. Department of Commerce, 10 pp.

_____, 1980: The use of model output statistics for predicting probability of precipitation. NWS Tech. Proc. Bull. No. 289, NOAA, U.S. Department of Commerce, 13 pp.

Newell and Deaven, 1981: The LFM-II model--1980. NOAA Tech. Memo. NWS NMC-66, NOAA, U.S. Department of Commerce, 20 pp.

Spiegler, D. B., and G. F. Fisher, 1971: A snowfall prediction method for the atlantic seaboard. Mon. Wea. Rev., 99, 311-325.

Weber, E. M., 1978: Major midwestern snowstorms. Preprints Conference on Weather Forecasting and Analysis and Aviation Meteorology, American Meteorological Society, Silver Spring, 30-37.

Younkin, R. J., 1968: Circulation patterns associated with heavy snowfall over the western United States. Mon. Wea. Rev., 96, 851-853.

Zurndorfer, E. A., and R. J. Bermowitz, 1976: Determination of an optimum number of predictors for probability of precipitation amount forecasting. TDL Office Note 76-17, National Weather Service, NOAA, U.S. Department of Commerce, 7 pp.

Table 1. The potential predictors used to develop experimental, conditional probability of snow amount equations.

Acronym	Description
B L REL HUM	Boundary-layer relative humidity
B L DIV	Boundary-layer divergence
B L VORT	Boundary-layer relative vorticity
B L M CONV	Boundary-layer moisture convergence
B L U	Boundary-layer east-west wind component
B L V	Boundary-layer north-south wind component
S L P TEND	Sea-level pressure tendency
850 HGT	850-mb height
850 HGT TEND	850-mb height tendency
850 VORT	850-mb relative vorticity
850 T ADV	850-mb temperature advection
850 DIV	850-mb divergence
850 M CONV	850-mb moisture convergence
850 U	850-mb east-west wind component
850 V	850-mb north-south wind component
700 VORT	700-mb relative vorticity
700 VORT ADV	700-mb vorticity advection
700 DIV	700-mb divergence
700 M CONV	700-mb moisture convergence
700 T ADV	700-mb temperature advection
700 VV	700-mb vertical velocity
700 U	700-mb east-west wind component
700 V	700-mb north-south wind component
500 HGT TEND	500-mb height tendency
500 VORT	500-mb relative vorticity
500 VORT ADV	500-mb vorticity advection
10-5 THICK ADV	Advection of 1000-500 mb thickness with 700-mb geostrophic wind
500 U	500-mb east-west wind component
500 V	500-mb north-south wind component
MEAN R H	Mean relative humidity (sfc.-500 mb)
P AMT	Precipitation amount
P WATER	Precipitable water (sfc.-500 mb)
STA ELEV	Station elevation
SIN DOY	Sine of the day of year
COS DOY	Cosine of the day of year

Table 2. Brier scores computed for snow amount forecasts from regionalized PoSA(S) equation sets containing 1, 2, 4, ..., 20 preditors. The sample consists of independent data combined from 195 stations for the winter season of 1980-81. The forecast projection is the 12-24 h period from 0000 GMT.

Number of Predictors	Brier Scores				
	Snow Amount Category (inches)				
	≥1	≥2	≥3	≥4	≥6
1	.406	.232	.133	.072	.023
2	.395	.232	.133	.073	.023
4	.389	.232	.133	.073	.022
6	.388	.233	.132	.071	.022
8	.387	.230	.132	.071	.023
10	.386	.229	.131	.071	.023
12	.385	.230	.132	.071	.023
16	.383	.229	.131	.072	.023
20	.384	.229	.131	.071	.023

Table 3. The reduction of variance for REEP conditional probability of snow amount equations for the 12-24 h projection from 0000 GMT for each of the five regions shown in Fig. 1. The developmental data were from 8 winter seasons (1972-73 through 1979-80). The relative frequency (%) of each snow amount category is also shown in parentheses.

Region	Total Reduction of Variance (%)					Total Number of Snow Cases
	Snow Amount Category (inches)					
	≥1	≥2	≥3	≥4	≥6	
1	10.9 (37.3)	9.9 (16.9)	7.3 (7.4)	8.6 (3.1)	7.4 (0.9)	1307
2	19.2 (49.3)	15.8 (27.9)	14.5 (17.8)	11.6 (10.3)	6.8 (3.4)	1100
3	17.8 (36.7)	17.9 (17.9)	17.9 (9.2)	18.3 (4.6)	13.2 (1.4)	4170
4	18.0 (39.4)	24.9 (19.6)	20.6 (11.1)	18.9 (6.4)	18.2 (2.3)	3785
5	22.8 (36.4)	24.1 (17.8)	22.5 (9.8)	20.8 (5.9)	17.3 (2.5)	3799

Table 4. The ranking of predictors in the conditional probability of snow amount equations for the 12-24 h and 24-36 h projections from 0000 GMT. Ranking is based both on the order and frequency of selection as determined by the REEP screening procedure. Predictor acronyms are defined in Table 1.

12-24 h Projection	24-36 h Projection
P AMT	P AMT
MEAN R HUM	MEAN R HUM
850 DIV	850 U
850 VORT	850 DIV
850 M CONV	500 VORT
700 U	700 U
700 VV	850 M CONV
10-5 THICK ADV	700 VV
700 V	P WATER
850 U	500 VORT ADV
P WATER	850 HGT TEND
500 VORT	700 V
700 VORT	700 M CONV
700 M CONV	10-5 TH ADV

Table 5. The predictors included in the MOSSQR equations (as determined by the REEP screening procedure) to predict PoSA for the 12-24 h projection after 0000 GMT. The developmental sample consisted of data combined from 192 stations for October through March, 1972-73 through 1979-80. The additional reduction of variance given by each predictor is shown. The number of cases for each snow amount category is given in parentheses.

Predictors	Additional Reduction of Variance (%)				
	Snow Amount Category (inches)				
	≥1	≥2	≥3	≥4	≥6
	(6599)	(3170)	(1679)	(899)	(307)
PoSA(S) (≥1 in) x PoP x $\overline{\text{PoF}}$	27.4	19.8	13.7	9.3	4.5
PoSA(S) (≥6 in) x PoP x $\overline{\text{PoF}}$	0.1	1.5	2.8	4.5	7.2
PoSA(S) (≥2 in) x PoP x $\overline{\text{Pof}}$ ≤70%	0.0	0.0	0.1	0.6	1.4
PoSA(S) (≥3 in) x PoP x $\overline{\text{PoF}}$	0.0	0.5	0.8	0.5	0.0
PoSA(S) (≥4 in) x PoP x $\overline{\text{PoF}}$ ≤50%	0.0	0.0	0.0	0.0	0.5
PoSA(S) (≥2 in) x PoP x $\overline{\text{PoF}}$ ≤50%	0.0	0.0	0.1	0.3	0.1
PoSA(S) (≥6 in) x PoP x $\overline{\text{PoF}}$ ≤4.5%	0.0	0.0	0.0	0.0	0.2
PoSA(S) (≥2 in) x PoP x $\overline{\text{PoF}}$ ≤18%	0.1	0.2	0.1	0.1	0.1
PoSA(S) (≥4 in) x PoP x $\overline{\text{PoF}}$	0.0	0.0	0.0	0.2	0.0
Climatic Frequency ≥6 in	0.0	0.0	0.0	0.1	0.1
PoSA(S) (≥6 in) x PoP x $\overline{\text{PoF}}$ ≤1.5%	0.0	0.0	0.0	0.0	0.1
PoSA(S) (≥2 in) x PoP x $\overline{\text{PoF}}$ ≤80%	0.0	0.0	0.0	0.0	0.1

Table 6. The Brier scores for the PRODUCT and MOSSQR methods for obtaining unconditional probability of snow amount forecasts. The developmental (independent) sample consisted of data from 8 winter seasons (1 winter season) 1972-73 through 1979-80 (1980-81). Data were combined from 192 stations. The percent improvement of the MOSSQR method over the PRODUCT method is also shown.

System	Snow Amount Category (inches)				
	≥1	≥2	≥3	≥4	≥6
	Developmental Data (236,535 cases)				
PRODUCT	.0401	.0212	.0119	.0068	.0024
MOSSQR	.0400	.0212	.0118	.0067	.0023
% Improvement (MOSSQR/PRODUCT)	+0.2	+0.0	+0.8	+1.5	+4.2
	Independent Data (32,053 cases)				
PRODUCT	.0296	.0152	.0085	.0046	.0016
MOSSQR	.0294	.0152	.0086	.0046	.0016
% Improvement (MOSSQR/PRODUCT)	+0.7	+0.0	-1.1	+0.0	+0.0

Table 7. Verification scores for categorical snow amount forecasts made from experimental PoSA equations for both developmental (D) and independent (I) samples. The samples are the same as those used for Table 6. Data were combined from 192 stations.

Forecast Projection from 0000 GMT	Verification Score	Snow Amount Category (inches)									
		≥1		≥2		≥3		≥4		≥6	
		D	I	D	I	D	I	D	I	D	I
12-24 h	Threat score	.30	.28	.25	.19	.21	.14	.19	.15	.18	.12
	Bias	1.08	.83	1.10	.65	1.07	.61	1.09	.64	1.03	1.00
	Number of snow cases	6599	629	3174	298	1671	150	902	83	310	27
12-18 h	Threat score	.27	.26	.21	.16	.17	.10	--	--	--	--
	Bias	1.10	.77	1.13	.71	1.02	.56	--	--	--	--
	Number of snow cases	3470	306	1353	126	584	55	--	--	--	--
18-24 h	Threat score	.26	.19	.21	.11	.18	.12	--	--	--	--
	Bias	1.22	.63	1.05	.39	1.20	.40	--	--	--	--
	Number of snow cases	2980	305	1212	119	528	47	--	--	--	--
24-36 h	Threat score	.28	.22	.22	.13	.20	.05	.17	.00	.12	.00
	Bias	1.29	1.09	1.14	.75	1.19	.57	1.16	.25	1.11	.14
	Number of snow cases	4387	655	2082	278	1108	133	564	69	185	21

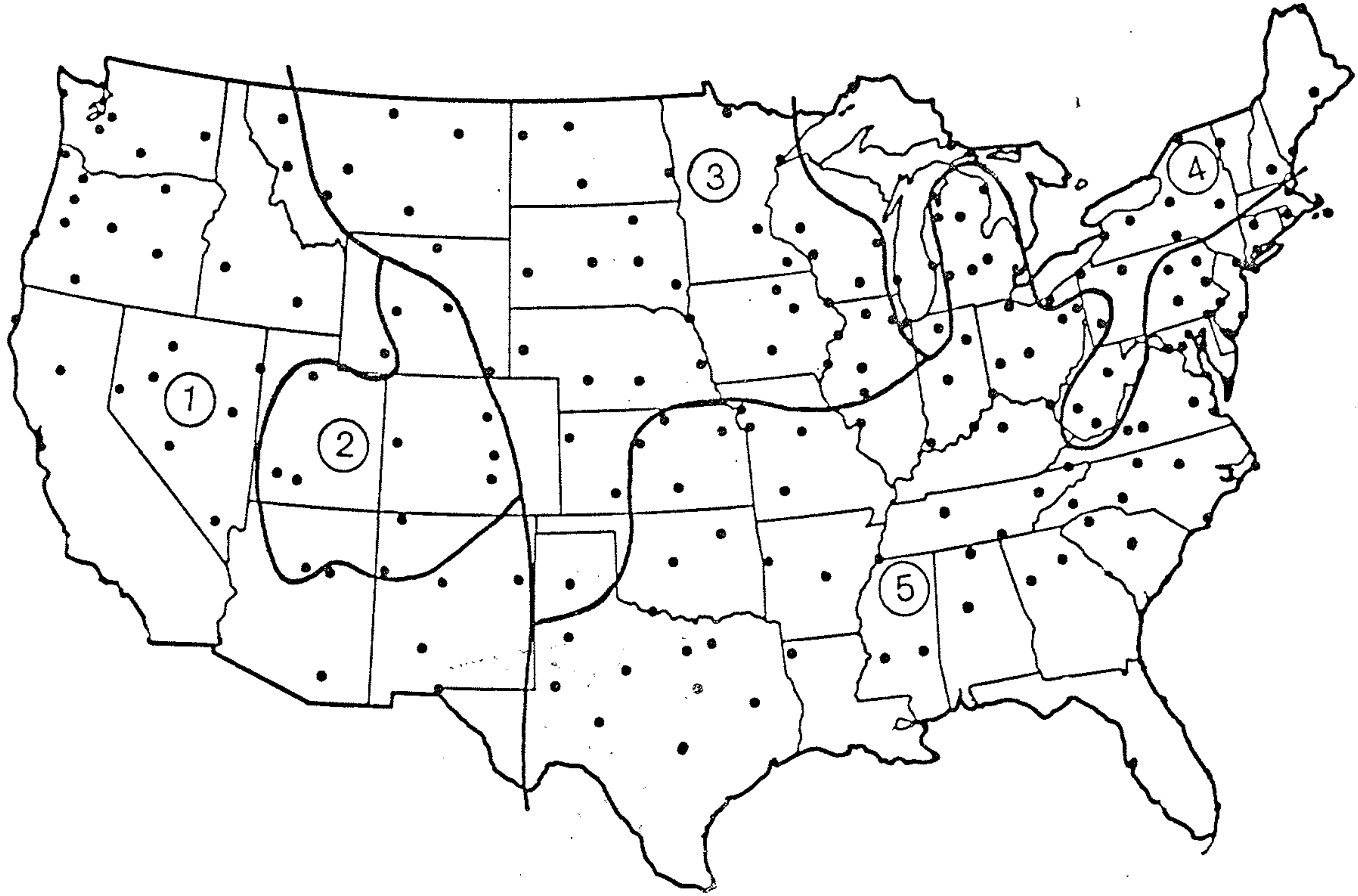

Figure 1. The five regions used in the development of the experimental PoSA(S) equations. The dots show the stations for which snow amount data were available in the developmental archive.