file

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

TDL OFFICE NOTE 82-3

RECENT IMPROVEMENTS IN AN AUTOMATED SYSTEM FOR FORECASTING PRECIPITATION TYPE

Joseph R. Bocchieri and George J. Maglaras

March 1982

RECENT IMPROVEMENTS IN AN AUTOMATED SYSTEM FOR FORECASTING PRECIPITATION TYPE

Joseph R. Bocchieri and George J. Maglaras

1. INTRODUCTION

A new system for forecasting the conditional probability of precipitation type (PoPT) (Bocchieri, 1979a) became operational within the National Weather Service in September 1978. To develop the forecast equations for PoPT, we used the Model Output Statistics (MOS) technique (Glahn and Lowry, 1972) with 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). The PoPT system evolved from an earlier operational system for forecasting the probability of frozen precipitation (PoF) (Glahn and Bocchieri, 1975; Bocchieri and Glahn, 1976).

In PoPT, precipitation type is defined as three categories: snow or ice pellets (SNOW), freezing rain or drizzle (ZR), and rain or mixed precipitation types (RAIN). The probability forecasts are conditional because the system assumes precipitation will occur; i.e., only precipitation cases were included in the developmental sample. Also, the forecasts are valid at specific times, i.e., every sixth hour from 12 through 48 hours in advance.

In an effort to improve the PoPT system, we developed a new, experimental set of PoPT forecast equations called EXP. EXP differs from the operational PoPT system, OPER, in several ways. First, EXP was developed with eight winter seasons (September through April) of LFM and LFM-II model output for the 12- and 24-h projections and about four winter seasons of data for the 36- and 48-h projections; the developmental sample for OPER consisted of about five seasons for the $\leq$ 24-h projections and two to three seasons for the longer-range projections. Second, in EXP, the "50% values" of several LFM predictors for each station were rederived using the larger data sample; a new constant, called "spread," also was developed. Both of the constants were obtained by fitting an S-shaped logit curve (Brelsford and Jones, 1967; Jones, 1968) to the data. The 50% value is that value which indicates a 50-50 chance of frozen precipitation for a station, provided precipitation occurs. The spread constant defines the shape of the logit curve; that is, for a given predictor, some curves are quite steep while others are quite shallow depending on the station. Most of the predictors used in the EXP system were transformed from their original values through application of these constants. Third, as compared to OPER, EXP includes improved interactive predictors, additional LFM u- and v-wind component predictors, additional space-smoothing of predictors, and climatic frequencies of ZR and SNOW. Finally, we redefined the precipitation type categories for EXP in the sense that freezing precipitation mixed with any other type was included in the ZR category; in OPER, this type of mixed precipitation is defined as RAIN.

We performed two experiments in the process of developing the EXP system. In the first experiment, we derived one set of EXP forecast equations using the Regression Estimation of Event Probability (REEP) statistical model (Miller, 1964) and another set using the logit model. REEP is essentially a linear regression model while logit, which was used to develop the OPER system, is non-linear. We then did a comparative verification on independent data between the logit-based set and the REEP set. The results indicate that the logit-based EXP was better. In the other experiment, on independent data we compared forecasts made from EXP logit equations to OPER system forecasts. The overall results indicate that EXP was better than OPER, especially for short-range ZR forecasting. Based on these experimental results, we decided to use the EXP approach in deriving new PoPT forecast equations for operational use.

2. DEVELOPMENT OF THE EXPERIMENTAL EQUATIONS

For the purposes of comparing REEP and logit EXP-type equations and comparing the EXP system with the OPER system, we developed three sets of EXP forecast equations for each of the 12-, 24-, 36-, and 48-h projections from 0000 GMT. Two of the sets were linear regression equations developed with the REEP screening procedure, while the other set was developed with the logit model. For the 12- and 24-h projections, the developmental sample consisted of data from eight winter seasons, 1972-73 through 1979-80; for the 36- and 48-h projections, data were available for four and one-half seasons, February 1976 through April 1980. Data from September 1980 through March 1981 were set aside for use as an independent sample.

In the REEP screening procedure, a subset of effective predictors for use in linear-regression equations is objectively selected from a larger set of potential predictors. The equations developed give estimates of the probabilities of occurrence for a given set of binary-type predictands. We divided precipitation type into three binary-type predictands: RAIN, ZR, and SNOW. The predictands are called binary because in the developmental phase each predictand was assigned a value of either 1 or 0 in a given case depending on whether or not that particular precipitation type occurred. The potential predictors were either in binary or continuous form. The use of binary predictors helps to account for possible non-linear relationships between the predictors and predictand. A good description of the REEP screening procedure can be found in Glahn and Lowry (1972).

Our present logit computer program doesn't have a screening option; therefore, the REEP screening procedure was used to determine the best set of predictors to include in the logit model. Predictors are included in the logit equations in continuous form only. As described in Bocchieri (1979a) the logit model was used to develop the OPER system.

a. The Potential Predictors

Table 1 shows the potential predictor variables we used to develop the EXP equations. Model output variables valid for the 6-, 12-, 18-, 24-, 36-, 42-, and 48-h projections were included. The observed surface variables were valid at 0300 or 1500 GMT. The table also gives the acronyms by which the various predictors will be referred in this paper. There are several differences between this set of potential predictors and the set used to develop the OPER system. For EXP, unsmoothed and 5-point space-smoothed variables were screened for ≤ 24-h forecasts; for > 24-h forecasts, we used unsmoothed, 5-point and 9-point space-smoothed predictors. For OPER, only unsmoothed variables were used for ≤ 24-h projections, and both unsmoothed and 9-point space-smoothed predictors were used for projections > 24 hours. Also, for EXP we used LFM u- and v- wind component forecasts at various levels for most of the forecast projections; whereas, in OPER, only boundary-layer and 850-mb u- and v-wind components were included for projections > 24 hours. We also included climatic frequencies of the ZR and SNOW categories as potential predictors in EXP. These frequencies weren't used to develop OPER. We derived the frequencies for each of about 230 stations for each of the months September through April by combining the 3-hourly surface observations of weather for the period 1972-73 through 1979-80. Only those observations which contained reports of precipitation were used; the frequencies are therefore conditional on the event that precipitation occurs. This sample was stratified by month but not by time of day.

To some extent, transforming predictors allows us to combine data from different stations in developing forecast equations. For EXP, we improved the predictor transformation procedure. In the OPER and PoF systems, we transformed several of the LFM model output variables into deviations from 50% values (see Glahn and Bocchieri, 1975 and Bocchieri, 1979a for details). Briefly, the 50% value of a variable is that value which indicates a 50-50 chance of SNOW at a station, provided precipitation occurs. We determined the 50% value for a model output predictor for each station by using the logit model to fit the data. The logit model provides a means for fitting a sigmoid or S-shaped curve when the dependent variable is binary and the independent variable is continuous. As discussed by Glahn and Bocchieri, the 50% value of a variable can vary quite a bit from station to station depending on local factors, especially station elevation. Our assumption was that a given deviation of a predictor from its 50% value should produce the same probability of SNOW at different stations. This assumption would be exact if the logit curve for a given predictor had the same shape for each station. Actually, this isn`t true; that is, for a given predictor, some curves are quite steep while others are quite shallow. For example, for a steep logit curve, the difference in the 850 T between the 5% and 95% points of the curve might be 4 K; however, for a shallow logit curve, this difference might be 8 K. Bocchieri (1979b) explained and illustrated these concepts in more detail and described experiments which showed the accuracy of probability of SNOW forecasts were improved by transforming predictors to account not only for the difference in 50% values between stations, but also to account for the difference in steepness or spread of the logit curves. For EXP therefore, the BLPT, 850 T, 850 WBT, 10-8.5 Th, 10-5 Th, and 8.5-5 Th predictors were transformed by,

$$X_T = \frac{X - (50\%\ \text{value})}{\text{spread}} , \qquad (1)$$

where X_T is the transformed predictor, X is the original value of the predictor, and spread is the difference in X units between the 95% point and the 50% point of the logit curve. We determined the 50% value and spread constants for each of the above predictors using the developmental sample in a manner similar to that described in Bocchieri (1979a and 1979b). Throughout the remainder of the paper, predictors tranformed according to Eq. (1) will be referred to as "standardized" predictors.

The joint predictors in Table 1 also were included to help capture first-order interactive effects between model output predictors. This concept is explained in more detail in Bocchieri (1979a). These joint predictors were developed in a manner similar to that used in the OPER system. First, using the developmental sample, graphs were constructed to show the relative frequency of ZR or of SNOW as a function of various pairs of LFM predictors. The 850 T + BLPT and 8.5-5 Th + 10-8.5 Th pairs were used for both the ZR and SNOW categories, while the 10.5 Th + BLPT pair was used for only the ZR category. Similar pairs were used in the OPER system except that the 8.5-5 Th + 10-8.5 Th pair is not included for the SNOW category; as shown later, this predictor is relatively important for SNOW for the longer range projections in EXP. Also, for the EXP equations, the BLPT was used instead of the boundary-layer wet-bulb temperature in the 850 T + BLPT pair for SNOW. This was done because a change in the earth`s terrain used in the LFM model affected the computation of the wet-bulb temperature. Before constructing the graphs for EXP, predictors were standardized.

Figs. 1, 2, and 3 show the empirical probability of ZR, given that precipitation occurs, as a function of the 850 T + BLPT, 10-5 Th + BLPT and 8.5-5 Th + 10-8.5 Th predictor pairs, respectively; in data sparse regions of the graphs the analysis is shown by dashed lines. Figs. 4 and 5 show the empirical probability of SNOW, given that precipitation occurs, as a function of the 850 T + BLPT and 8.5-5 Th + 10-8.5 Th predictor pairs, respectively. These graphs are generally similar to the ones used in the OPER system (see Figs. 1, 2, 3, and 4 in Bocchieri, 1979a) except that, as would be expected, there are differences in the probability surfaces within the data sparse regions of the graphs. A meteorological interpretation of these graphs for ZR and SNOW generally follows that given in Bocchieri (1979a). For example, Fig. 1 can be interpreted as follows. The probability of ZR is relatively high when the 850 T is higher than its 50% value (standardized value > 0.0) and when the BLPT is lower than its 50% value (standardized value < 0.0). This situation, with relatively warm air aloft and cold air near the surface, is conducive to the occurrence of ZR.

b. Regionalization

As in the OPER system, we developed EXP equation sets for each of several geographic regions. The regions were determined in the following manner. The REEP screening program was run on the developmental sample for the 12-, 24-, 36-, and 48-h projections from both the 0000 GMT and 1200 GMT cycle times with data combined from 229 conterminous United States stations--the so-called generalized-operator approach. For the purpose of establishing regions, we categorized precipitation type into two binary-type predictands, one being SNOW and the other being ZR and RAIN combined. The two statistics used to help determine the regions were the relative probability bias and the categorical bias. To obtain these statistics, we evaluated REEP probability of SNOW equations to obtain forecasts for each station on the developmental sample. The relative probability bias for each station was computed by,

$$\text{Rel. Prob. Bias} = \frac{\overline{\text{P(SNOW)}} - \text{RF(SNOW)}}{\text{RF(SNOW)}}, \tag{2}$$

where $\overline{\text{P(SNOW)}}$ is the average probability of SNOW forecast for each station and RF(SNOW) is the relative frequency of SNOW for each station from the developmental sample. To compute the categorical bias, the probability of SNOW forecast for each case of the developmental sample was transformed into a categorical forecast; that is, a categorical SNOW forecast resulted if the probability forecast exceeded 50%. The categorical bias was then computed for each station from,

$$\text{Cat. Bias} = \frac{\text{FCST (SNOW)}}{\text{OBS (SNOW)}} , \qquad (3)$$

where FCST (SNOW) is the number of forecast SNOW events and OBS (SNOW) is the number of observed SNOW events. Figs. 6 and 8 show the relative probability bias and categorical bias, respectively, averaged for the 12- and 24-h projections from 0000 GMT and 1200 GMT, while Figs. 7 and 9 show similar statistics averaged for the 36- and 48-h projections from the 0000 GMT and 1200 GMT cycles. For these maps, we didn`t analyze the data in areas below the dotted lines because the number of SNOW cases was not sufficient to give meaningful results. The analysis in Fig. 6 indicates that the bias for the 12- to 24-h period was generally small (within ± 10%). However, for the 36- to 48-h period (Fig. 7) the probability forecasts were generally too high over the Rocky Mountain area and portions of Northern Texas, Oklahoma, Kansas, and Nebraska, and generally too low over the Atlantic Coastal Plain. The analyses in Figs. 8 and 9 show that the SNOW event was generally overforecast over the Rocky Mountain area and portions of Northern Texas, Oklahoma, Kansas, and Nebraska, and generally underforecast over the Atlantic Coastal Plain, the Pacific Northwest, and portions of the south. In general, the areas of overforecasting and underforecasting were similar for both types of bias.

From these analyses, we determined the regions shown in Fig. 10 by grouping stations with similar characteristics. Other factors considered included the climatic frequency of the SNOW event and the density of stations. We tried to make the regions reasonably large so that the sample used to derive the EXP forecast equations for each region would be as large as possible; sample size is an important consideration for ZR, which is a relatively rare event.

c. Development of Regionalized EXP Equations

After specifying the appropriate regions, we developed two sets of REEP EXP equations and one set of logit equations for each region for the 12-, 24-, 36-, and 48-h projections by combining data from all stations within a region. The REEP screening procedure was used determine the predictors to use in the logit equations; Bocchieri and Glahn (1976) showed that logit equations should include about 10 continuous predictors. The REEP equation set which resulted from the screening runs to determine predictors for logit is called the REEP1 set and typically includes 10 to 15 predictors, depending on the region, in both continuous and binary form. The number of predictors included in the REEP1 set was determined by the fact that 10 predictors were to be included in the logit equations. Also, we allowed the screening procedure to continue until 20 predictors were selected for each region; the resulting equation set is called REEP2.

Table 2 lists the 10 most important predictor types as given by the REEP screening procedure for the 12-, 24-, 36-, and 48-h projections from 0000 GMT. This list was determined by both frequency and order of selection; for the purpose of this ranking, all predictor projections, smoothings, and binary limits were combined for each type of variable. The results indicate that:

(1) as in the OPER system, the joint predictors were dominant among the first several; (2) the 850 T + BLPT (SNOW) joint predictor was generally the most important and was chosen because of its ability to discriminate between SNOW and other precipitation types; (3) the 8.5-5 Th + 10-8.5 Th (SNOW), which was not included in the OPER system, was also relatively important especially for the longer range projections; (4) the FREQ ZR predictor, which was not included in the OPER system, also ranked relatively high, especially for the longer range projections; and, of course, (5) the OBS SFC T and OBS SFC Td became less important as the projection increased.

Table 3 shows the total reduction of variance for the REEP1 set for the 12- and 36-h projections from the 0000 GMT LFM cycle for each region. Note that, as in the OPER system, the reduction of variance was generally quite high for the RAIN and SNOW categories but rather low for ZR. The relatively infrequent occurrence of ZR makes its prediction especially difficult; the relatively high values for ZR in region 1, where ZR is very rare, is no doubt a case of over-fitting of the data. It`s also interesting that in regions 2 and 3 SNOW occurred more frequently than RAIN, and in region 5 the frequency of occurrence of ZR was close to that of SNOW.

3. VERIFICATION OF THE EXPERIMENTAL EQUATIONS

a. Comparison Between REEP and Logit

Bocchieri and Glahn (1976) showed the logit model was better than REEP for PoF forecasting. We decided to do a more extensive comparison between REEP and logit, since it would be more efficient to use REEP in developing new PoPT forecast equations. In particular, we compared the REEP1, REEP2, and logit EXP equation sets on independent data combined from 229 stations for the period September 1980 through March 1981. As noted previously, REEP1 typically included 10 to 15 predictors, depending on the region; logit included 10 predictors, which were determined by those selected for REEP1; and REEP2 included 20 predictors. The REEP2 system was included in this comparison because we hypothesized that the greater the number of binary predictors in the REEP equations the more likely that non-linear relationships between the predictors and the predictands would be accounted for. The logit model is inherently non-linear.

Table 4 shows the P-scores (Brier, 1950) for logit, REEP1, and REEP2 forecast equation sets for PoPT for the 12-, 24-, 36-, and 48-h projections from 0000 GMT. The percent improvement in P-score of logit over the best REEP is also shown. The results indicate that there was generally little difference between REEP1 and REEP2 and that logit was better than the best REEP by 4.0% at 12 hours and 1.8% at 24 hours; there was little difference between logit and REEP for 36 and 48 hours. These results generally agree with those found by Bocchieri and Glahn (1976) and indicate that we should continue to use the logit model for PoPT forecasting.

b. Comparison Between EXP and OPER

We also did a comparative verification between the logit EXP equation set and the OPER system using the same independent sample mentioned before. Both probabilistic and categorical precipitation type forecasts were verified.

Table 5 shows the P-scores for probability forecasts for the RAIN, ZR, and SNOW categories and the sum of the categorical P-scores (TOTAL P) for the EXP equation sets and the OPER system for the 12-, 24-, 36-, and 48-h projections from 0000 GMT. The percent improvement in P-score of EXP over OPER and the number of cases are also shown. The results indicate the following: (1) in terms of TOTAL P, the percent improvement of EXP over OPER ranged from 2.6% to 5.0%, except there was little difference between the two systems at 24 hours; (2) for the SNOW category, the improvement ranged from 2.0% to 5.2%; and (3) for the ZR category, there was little difference between the two systems for the 36- and 48-h projections, but there was substantial improvement of EXP over OPER at 12 hours (13.0%) and 24 hours (7.1%).

It should be noted that in the above verification a new definition of the ZR category was used; that is, freezing rain mixed with any other precipitation type was included in the ZR category. However, in the OPER system, this mixed precipitation event was included in the RAIN category. One may argue that the above verification gave a slight advantage to EXP. Therefore, we repeated the verification using the predictand categories as defined for the OPER system; the results (not shown) were similar to those in Table 5.

We also examined the reliability of the ZR and SNOW probability forecasts on the independent data sample. Reliable probability forecasts are such that for all of the SNOW forecasts of 20%, say, the relative frequency of SNOW is close to 20%. Figs. 11 and 12 show the reliability of probability forecasts for ZR and SNOW, respectively, for the 12-h projection from 0000 GMT. For ZR there was little difference between EXP and OPER for forecasts < 25%; however, for probabilities > 25%, EXP was generally more reliable. For the SNOW category, there was little difference between EXP and OPER, except that EXP was more reliable than OPER within the range 25% to 55%.

Figs. 13 and 14 show the reliability of probability forecasts for ZR and SNOW, respectively, for the 36-h projection from 0000 GMT. For ZR, there was little difference between EXP and OPER for probabilites < 35%. However, EXP was better than OPER for probabilities > 35%, although EXP tended to overestimate the actual relative frequency of ZR for this range of probabilities. It`s interesting to note that EXP made 66 forecasts > 25%, while OPER made only 13. In this sense, therefore, EXP was the better system, even though the P-scores in Table 5 showed little difference between EXP and OPER for ZR for the 36-h projection. For the SNOW category, EXP was generally more reliable than OPER; although the probability forecasts from both systems tended to be too high. Note also that, as compared to OPER, EXP had more cases for probabilities near 0% and 100% and less cases near 50%. This indicates that EXP had greater resolution than OPER; perfect resolution results when all forecasts are either at 0% or 100%.

In addition to comparing probability forecasts from EXP and OPER equations, we also verified categorical forecasts of precipitation type. The verification scores included threat score, bias, post-agreement, and prefigurance.[1] To

---

[1] The threat score = A/(B + C - A), the bias = B/C, the post-agreement = A/B, and the prefigurance = A/C, where A, B, and C are the number of correct forecasts, the total number of forecasts, and the number of observations of the event, respectively.

transform the probability forecasts into categorical forecasts, we determined threshold probability values for each region and projection from the developmental sample so as to obtain a relatively high threat score for the ZR and SNOW categories while restricting the bias to between 0.90 and 1.10 (see Bocchieri, 1979a for further details).

For the purpose of deriving threshold probabilities, we divided region 3 in Fig. 10 into three new regions; the resulting regions are shown in Fig. 15. This was done mainly for convenience in using computer programs to develop the thresholds. We felt that the differences in the categorical bias within the old region 3 (see Figs. 8 and 9) weren`t large enough to justify dividing this region for the purpose of deriving probability equations. However, we used these slight differences to divide the regions as shown in Fig. 15.

Table 6 shows the verification scores computed from categorical precipitation type forecasts from EXP and OPER equations for the 12- and 36-h projections from 0000 GMT. The independent data sample was the same as that used in the verification of the probability forecasts. The results indicate that: (1) for the ZR category, EXP was generally better than OPER for all scores and both projections, especially for the 12-h projection and especially for the bias; and (2) for the RAIN and SNOW categories, there was little difference between EXP and OPER for the 12-h projection, but EXP was better overall than OPER for the 36-h projection.

These results are generally consistent with those found in the comparative verification between EXP and OPER for probability forecasts. However, even though the P-scores in Table 5 showed little difference between EXP and OPER for the 36-h projection for the ZR category, the results for the categorical ZR forecasts indicate that EXP was better overall than OPER at 36 hours. A similar result was found when the reliability of the ZR probability forecasts was examined for the 36-h projection.

Note also in Table 6 that the scores for EXP were very good for RAIN and SNOW; the bias, for instance, was near perfect for both projections, and the post-agreement shows that there was a 90% to 95% chance that forecasts of these categories were correct. However, the scores for EXP for the ZR category were not nearly as good as those for RAIN and SNOW. The bias, for instance, was good at 12 hours, but EXP tended to underforecast ZR at 36 hours. Also, the post-agreement was only about 50% at 12 hours and decreased to about 30% at 36 hours.

In summary, the comparative verification between EXP and OPER for both probability and categorical precipitation type forecasts indicates that overall EXP was better than OPER and that EXP was substantially better for 12- to 24-h ZR forecasts. Therefore, we decided to develop PoPT forecast equations, which incorporate the new features of EXP, for operational implementation. Some aspects of the development of the new system are described in the next section.

4. DEVELOPMENT AND VERIFICATION OF NEW OPERATIONAL POPT FORECAST EQUATIONS

We developed new PoPT forecast equations for 6-, 12-, 18-, 24-, 30-, 36-, 42-, 48-, 54-, and 60-h projections from both 0000 GMT and 1200 GMT. For this development, we combined the dependent and independent data samples used in the experiments so that for the 6- through 24-h projections, nine winter seasons (1972-73 through 1980-81) were available, and for the 30- through 60-h projections, about five and one-half seasons (February 1976 through 1980-81) were available. To develop the new PoPT system, we used the same potential predictors (Table 1), the same regions (Fig. 10), and the same predictand categories as were used to develop EXP. The REEP screening procedure was used to determine the predictors to include in the logit equations for each projection. In addition to "primary" sets of PoPT equations, which contain surface observations valid at 0300 GMT or 1500 GMT as predictors, we also developed "backup" equations which don`t include surface observations. In order to transform the probability forecasts into categorical precipitation type forecasts, we derived threshold probabilities for ZR and SNOW for each of the seven regions shown in Fig. 15 in a manner similar to that used for EXP.

We verified the primary and backup PoPT equations for the 12-, 24-, 36-, and 48-h projections from both 0000 GMT and 1200 GMT for the new developmental data sample. Tables 7 and 8 show the threat scores and post-agreements for each region (see Fig. 15) for the ZR and SNOW categories for the 12- and 36-h projections from the 0000 GMT cycle time. The bias and prefigurance are not shown because the bias was forced to be between 0.90 and 1.10, and, therefore, the prefigurance was similar in value to the post-agreement. Probability threshold values are also shown in Tables 7 and 8. The verification results for the 1200 GMT cycle time (not shown) were generally similar to those for the 0000 GMT cycle time.

The results in Tables 7 and 8 are summarized as follows:

1) The conditional relative frequency of ZR ranged from < 1% for the Pacific Coast region to near 7% for the Central Plains region. Because of the small number of cases, the scores for ZR for the Pacific Coast are probably not reliable and will not be included in further discussion of the results. The conditional relative frequency of SNOW ranged from near 2% for the Pacific Coast to between 70 and 75% in the Northern Plains. The probability threshold values and verification scores generally varied in accordance with the frequency of the event; that is, the lower threshold values and scores were generally associated with the lower frequencies of occurrence. It should be noted that the conditional relative frequencies of ZR and SNOW vary according to time-of-day; the values shown in Tables 7 and 8 are valid at 1200 GMT. Relative frequencies valid at 1800 GMT or 0000 GMT were generally lower, especially for ZR.

2) The probability threshold values for ZR for the primary equations for the 12-h projection ranged from about 20% for the Rocky Mountain and Northern Plains regions to near 30% for the other regions. For the 36-h primary equations for ZR, the threshold values were slightly lower and ranged between 19% and 26%. For SNOW, the threshold values for the primary equations at 12 hours ranged from 26% for the Pacific Coast to 53% for the Rocky Mountain region; at 36 hours, the threshold values for SNOW were generally similar. These results show, overall, the lower threshold values were associated with lower conditional relative frequencies. The reason is that, as expected, the probability forecasts from the statistical equations were lower, in the mean, for the rarer events.

3) The threat scores and post-agreements for SNOW were quite high. For the primary equations at 12 hours (36 hours), the threat scores ranged from .63 (.49) for the Pacific Coast to .91 (.90) for the Northern Plains. The post-agreements at 12 hours (36 hours) ranged from about 75% (67%) for the Pacific Coast to about 95% (95%) for the Northern Plains. On the other hand, the scores for ZR were not nearly as good. For 12-h forecasts, the threat scores for the primary ZR equations ranged from .25 for the Rocky Mountain region to .39 for the South; at 36 hours, the threat scores ranged between .20 and .30. The post-agreements at 12 hours ranged from near 40% for the Rocky Mountains to about 56% for the South; at 36 hours, the post-agreements ranged between 35% and 40%. As was the case with the probability threshold values, the lower verification scores were generally associated with the lower frequencies of occurrence.

4) As expected, the verification scores for the backup equations were generally worse than the scores for the primary equations, more so for ZR than for SNOW and more so at 12 hours than at 36 hours. Also, the probability threshold values for the backup equations were generally about the same or slightly lower than those for the primary equations.

In addition to verifying categorical precipitation type forecasts for the new PoPT equations, we also examined the reliability of the probability forecasts. Figs. 16 and 17 show the reliability of the ZR probability forecasts from primary PoPT equations for the developmental sample for the 12- and 36-h projections, respectively, from the 0000 GMT cycle time. For this purpose, data were combined from 229 stations. The results for the 12-h projection (Fig. 16) indicate that the forecasts were quite reliable for probabilities < 65% but tended to slightly overestimate the relative frequency of ZR for forecasts > 65%. This tendency to overestimate was also apparent at the 36-h projection (Fig. 17) for probabilities > 35%. A similar result for ZR probability forecasts was found by Bocchieri (1979a) who noted that the logit model apparently is capable of forecasting high probabilities for ZR, but the system tends to be overconfident in view of the infrequent occurrence of the event. In contrast, the reliability values for the probability of SNOW (not shown) were generally quite good.

5. SUMMARY AND CONCLUSIONS

A new system for forecasting the conditional probability of precipitation type (PoPT) became operational within the National Weather Service in September 1978. That system, called OPER, was developed with the MOS technique and provided forecasts for three categories of precipitation type: snow or sleet (SNOW), freezing rain (ZR), and rain or mixed types (RAIN). In an effort to improve OPER, we developed a new, experimental set of PoPT forecast equations, called EXP.

EXP differs from OPER in several ways. First, the developmental sample used to develop EXP included about three winter seasons more data than OPER. Second, the "50% values" for several LFM predictors for each station were rederived using the larger data sample; a new constant, called "spread," also was developed. Most of the predictors used in the EXP equations were transformed from their original values through the application of these constants. Third, as compared to OPER, EXP included improved interactive predictors, additional LFM u- and v-wind component predictors, additional space-smoothing of predictors, and conditional climatic frequencies of ZR and SNOW. Finally, for EXP, freezing precipitation mixed with any other type was included in the ZR category; in OPER, this mixture of precipitation was defined as RAIN.

We developed both REEP and logit-based regionalized EXP equations for the 12-, 24-, 36-, and 48-h projections from 0000 GMT. The results of REEP screening indicate the joint predictors were dominant among the first several chosen and that the climatic frequency of ZR, which was not included in OPER, also ranked relatively high, especially for the longer range projections. We compared REEP and logit-based PoPT forecasts on independent data. The results indicate that we should continue to use the logit model for PoPT forecasting.

We also did a comparative verification on independent data between logit-based EXP and OPER probability and categorical forecasts. The results for the probability forecasts indicate that, for all three categories of precipitation type combined, EXP improved over OPER by about 2% to 5% in the P-score. For the ZR category in particular, there was little difference between the two systems at 36 and 48 hours, but there was substantial improvement of EXP over OPER at 12 hours (13%) and 24 hours (7%).

The comparative verification results between EXP and OPER for categorical forecasts on independent data were generally consistent with those found for probability forecasts. The scores for the EXP categorical forecasts were very good for RAIN and SNOW; the bias was near perfect, and 90% to 95% of the forecasts were correct. However, the scores for ZR were not nearly as good. The bias, for instance, was good at 12 hours, but EXP tended to underforecast ZR at 36 hours. Also, about 50% (30%) of the ZR forecasts were correct at 12 hours (36 hours).

Based on the comparative verifications between EXP and OPER, we concluded EXP was better overall than OPER and that EXP was substantially better for 12- to 24-h ZR forecasts. Therefore, we derived new PoPT forecast equations for operational use (see Section 4), incorporating the new features associated with the EXP equations. We plan to implement these new equations in the fall of 1982.

6. ACKNOWLEDGMENTS

We wish to acknowledge Paul Banas for his help with data processing and the many other members of the Techniques Development Laboratory who contribute to development and maintenance of the MOS system.

7. REFERENCES

Bocchieri, J. R., 1979a: A new operational system for forecasting precipitation type. Mon. Wea. Rev., 107, 637-649.

_____, 1979b: Use of the logit model to transform predictors for precipitation type forecasting. Preprints Sixth Conference on Probability and Statistics in Atmospheric Sciences, Banff, Amer. Meteor. Soc., 49-54.

_____, and H. R. Glahn, 1976: Verification and further development of an operational model for forecasting the probability of frozen precipitation. Mon. Wea. Rev., 104, 691-701.

Brelsford, W. M., and R. H. Jones, 1967: Estimating probabilities. Mon. Wea. Rev., 95, 570-576.

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

Gerrity, J. F., Jr., 1977: The LFM model-1976: A documentation. NOAA Technical Memorandum 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.

_____, and J. R. Bocchieri, 1975: Objective estimation of the conditional probability of frozen precipitation. Mon. Wea. Rev., 103, 3-15.

Jones, R. H., 1968: A nonlinear model for estimating probabilities of K events. Mon. Wea. Rev., 96, 383-384.

Miller, R. G., 1964: Regression estimation of event probabilites. Technical Report 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 Technical Procedures Bulletin No. 67, NOAA, U.S. Department of Commerce, 11 pp.

_____, 1977: High resolution LFM (LFM-II). NWS Technical Procedures Bulletin No. 206, NOAA, U.S. Department of Commerce, 6 pp.

Table 1. The potential predictors included in the development of EXP forecast equations.

Acronym	Definition
	a. Model Output Predictors
BLPT	Boundary-layer potential temperature
BL U	Boundary-layer east-west wind component
BL V	Boundary-layer north-south wind component
850 T	850-mb temperature
850 WBT	850-mb wet-bulb temperature
850 U	850-mb east-west wind component
850 V	850-mb north-south wind component
700 U	700-mb east-west wind component
700 V	700-mb north-south wind component
10-85 Th	1000-850 mb thickness
10-5 Th	1000-500 mb thickness
8.5-5 Th	850-500 mb thickness
	b. Model Output Joint Predictors
850 T + BLPT (ZR)	850-mb temperature and boundary-layer potential temperature for the ZR category
850 T + BLPT (SNOW)	850-mb temperature and boundary-layer potential temperature for the SNOW category
10-5 Th + BLPT (ZR)	1000-500 mb thickness and boundary-layer potential temperature for the ZR category
8.5-5 Th + 10-8.5 Th (ZR)	850-500 mb thickness and 1000-850 mb thickness for the ZR category
8.5-5 Th + 10-8.5 Th (SNOW)	850-500 mb thickness and 1000-850 mb thickness for the SNOW category
	c. Observed and Miscellaneous Predictors
OBS SFC T	Observed surface temperature
OBS SFC Td	Observed surface dew-point temperature
OBS SFC U	Observed surface east-west wind component
OBS SFC V	Observed surface north-south wind component
STA ELEV	Station elevation
SIN DOY	Sine of the day of year
COS DOY	Cosine of the day of year
FREQ ZR	Climatic frequency of ZR
FREQ SNOW	Climatic frequency of SNOW

Table 2. The 10 most important predictors as determined by the REEP screening procedure for the 12-, 24-, 36-, and 48-h projections from 0000 GMT. Ranking is based both on the order and frequency of selection. Predictor acronyms are defined in Table 1.

12-h Projection	24-h Projection	36-h Projection	48-h Projection
850 T + BLPT (SNOW)	850 T + BLPT (SNOW)	8.5-5 Th + 10-8.5 Th (SNOW)	850 T + BLPT (SNOW)
OBS SFC T	8.5-5 Th + 10-8.5 Th (SNOW)	FREQ ZR	8.5-5 Th + 10-8.5 Th (SNOW)
10-5 Th + BLPT (ZR)	OBS SFC T	OBS SFC T	FREQ ZR
8.5-5 Th + 10-8.5 Th (SNOW)	FREQ ZR	8.5-5 Th + 10-8.5 Th (ZR)	8.5-5 Th + 10-8.5 Th (ZR)
FREQ ZR	8.5-5 Th + 10-8.5 Th (ZR)	850 T + BLPT (SNOW)	FREQ SNOW
OBS SFC Td	10-5 Th + BLPT (ZR)	10-5 Th + BLPT (ZR)	850 T + BLPT (ZR)
850 T + BLPT (ZR)	10-5 Th	850 T + BLPT (ZR)	10-8.5 Th
10-8.5 Th	850 T + BLPT (ZR)	OBS SFC Td	10-5 Th + BLPT (ZR)
8.5-5 Th + 10-8.5 Th (ZR)	10-8.5 Th	10-5 Th	OBS SFC T
850 WBT	OBS SFC Td	850 V	OBS SFC Td

Table 3. The reduction of variance for the REEP1 equation set for the 12- and 36-h projections from 0000 GMT for each of the five regions shown in Fig. 10. The developmental data were from eight winter seasons (1972-73 through 1979-80) for the 12-h projection and about four winter seasons (1976-77 through 1979-80) for the 36-h projection. The predictors to include in the logit equations were chosen from this set. The total number of precipitation cases and the relative frequency of each precipitation type (%) (in parentheses) are also shown.

Region	Total Reduction of Variance (%)			Total Number of Cases
	SNOW	ZR	RAIN	
		12-h Projection		
1	56.4 (2.4)	30.0 (0.5)	61.7 (97.1)	4492
2	66.7 (69.8)	15.2 (2.8)	69.9 (27.4)	5654
3	79.0 (58.8)	22.7 (4.5)	82.1 (36.7)	25733
4	76.6 (18.7)	28.6 (3.8)	75.7 (77.4)	4101
5	69.8 (5.0)	36.4 (2.9)	70.7 (92.1)	6996
		36-h Projection		
1	41.7 (2.2)	23.1 (0.6)	42.9 (97.2)	2423
2	59.2 (70.5)	13.5 (3.1)	61.1 (26.4)	3117
3	74.0 (61.9)	16.2 (4.3)	76.3 (33.8)	14872
4	65.2 (21.2)	21.0 (3.8)	64.6 (75.0)	2329
5	60.4 (5.3)	26.4 (3.4)	60.3 (91.3)	4033

Table 4. The P-scores for the LOGIT, REEP1, and REEP2 EXP equation sets for PoPT forecasts for the 12-, 24-, 36-, and 48-h projections from 0000 GMT. The sample consisted of independent data combined from 229 stations for the period September 1980-March 1981. The percent improvement of LOGIT equation forecasts over the best REEP is also shown. The sample included about 4200 cases for the 24- and 48-h projection and 4600 cases for the 12- and 36-h projections.

System	Projection 12-h	 24-h	 36-h	 48-h
LOGIT	.096	.112	.131	.148
REEP1	.103	.115	.130	.147
REEP2	.100	.114	.131	.149
% Improvement LOGIT/REEP	+4.0	+1.8	-0.8	-0.7

Table 5. P-scores for PoPT forecasts from experimental (EXP) logit equations and the operational (OPER) system for the 12-, 24-, 36-, and 48-h projections from 0000 GMT. The sample consisted of independent data combined from 229 stations for the period September 1980-March 1981. The P-scores for each precipitation type category and the total P-score (TOTAL) are shown. The percent improvement in P-score of the EXP equation set over the OPER system is also shown.

System	Projection															
	12-h				24-h				36-h				48-h			
	RAIN	ZR	SNOW	TOTAL	RAIN	ZR	SNOW	TOTAL	RAIN	ZR	SNOW	TOTAL	RAIN	ZR	SNOW	TOTAL
EXP	.036	.020	.040	.096	.049	.013	.050	.112	.053	.024	.055	.132	.067	.014	.068	.149
OPER	.037	.023	.041	.101	.048	.014	.051	.113	.057	.024	.058	.139	.069	.014	.070	.153
Percent improvement	+2.7	+13.0	+2.4	+5.0	-2.1	+7.1	+2.0	+0.9	+7.0	+0.0	+5.2	+5.0	+2.9	+0.0	+2.8	+2.6
Number of Cases	2391	125	2006	4522	2310	59	1682	4051	2441	127	2068	4636	2295	61	1691	4047

Table 6. Verification scores computed from categorical precipitation type forecasts made by the EXP and OPER equations for the 12- and 36-h projections from 0000 GMT. The sample consisted of independent data combined from 229 stations for the period September 1980-March 1981 (about 4500 precipitation cases). The percent improvement in the scores of EXP over OPER is also shown.

System	Bias			Threat Score			Post-Agreement			Prefigurance		
	RAIN	ZR	SNOW	RAIN	ZR	SNOW	RAIN	ZR	SNOW	RAIN	ZR	SNOW
12-h Forecast												
EXP	1.00	.94	1.01	.91	.29	.88	.95	.47	.93	.95	.44	.94
OPER	.99	.79	1.03	.91	.24	.88	.96	.43	.92	.94	.34	.95
% Improvement EXP/OPER	-	-	-	0.0	+20.1	+0.0	-1.0	+9.3	+1.1	+1.1	+32.4	-1.1
36-h Forecast												
EXP	1.00	.61	1.03	.86	.12	.83	.93	.29	.90	.93	.17	.92
OPER	.95	.48	1.09	.84	.11	.82	.94	.31	.87	.89	.15	.94
% Improvement EXP/OPER	-	-	-	+2.4	+9.1	+1.2	-1.1	-6.5	+3.4	+4.5	+13.3	-2.1

Table 7. Verification scores for the new operational PoPT system computed by region (see Fig. 15) for the developmental sample. The scores are for the 12-h projection from 0000 GMT. Numbers in parenthesis are scores for the backup equations which do not contain observed predictors.

Region	Category	Probability Threshold (%)	Threat Score	Post-agreement (%)	Number of Cases	Relative Frequency (%)
1	ZR	30.0 (19.0)	0.48 (0.24)	64.0 (39.1)	24	0.5
(Pacific Coast)	SNOW	26.0 (30.0)	0.63 (0.55)	75.2 (71.6)	111	2.2
2	ZR	22.0 (19.0)	0.25 (0.21)	39.7 (33.8)	184	3.0
(Rocky Mountain)	SNOW	53.0 (50.0)	0.87 (0.85)	93.0 (91.8)	4273	69.3
3	ZR	19.0 (16.0)	0.30 (0.27)	44.8 (40.9)	498	4.3
(Northern Plains)	SNOW	45.0 (48.0)	0.91 (0.90)	94.9 (94.7)	8179	71.2
4	ZR	29.0 (29.0)	0.38 (0.34)	54.1 (48.4)	365	6.4
(Central Plains)	SNOW	42.0 (44.0)	0.82 (0.81)	91.8 (91.5)	2396	41.8
5	ZR	29.0 (27.0)	0.31 (0.30)	47.2 (45.8)	368	3.4
(Northeast)	SNOW	51.0 (49.0)	0.90 (0.90)	94.6 (94.3)	6116	56.7
6	ZR	29.0 (22.0)	0.37 (0.28)	54.6 (42.0)	170	3.8
(Mid-Atlantic)	SNOW	40.0 (49.0)	0.78 (0.76)	88.1 (89.5)	848	19.0
7	ZR	30.0 (24.0)	0.39 (0.30)	56.3 (44.8)	213	2.7
(South)	SNOW	35.0 (31.0)	0.69 (0.69)	85.7 (84.2)	421	5.3

Table 8. The same as Table 7 except scores are shown for 36-h projection.

Region	Category	Probability Threshold (%)	Threat Score	Post-agreement (%)	Number of Cases	Relative Frequency (%)
1	ZR	28.0 (19.0)	0.38 (0.20)	55.6 (33.3)	18	0.6
(Pacific Coast)	SNOW	30.0 (27.0)	0.49 (0.44)	66.7 (64.2)	59	2.0
2	ZR	21.0 (19.6)	0.27 (0.21)	42.5 (35.0)	122	3.4
(Rocky Mountain)	SNOW	58.0 (49.0)	0.83 (0.82)	91.7 (89.2)	2434	68.5
3	ZR	19.0 (19.0)	0.22 (0.22)	35.4 (35.7)	291	4.0
(Northern Plains)	SNOW	49.0 (48.0)	0.90 (0.90)	94.9 (94.9)	5455	74.0
4	ZR	26.0 (26.0)	0.29 (0.29)	45.2 (44.1)	242	7.1
(Central Plains)	SNOW	45.0 (31.0)	0.79 (0.78)	90.2 (87.7)	1534	44.7
5	ZR	23.0 (24.0)	0.21 (0.20)	33.6 (33.2)	214	3.0
(Northeast)	SNOW	50.0 (46.0)	0.88 (0.88)	94.1 (93.9)	4252	59.9
6	ZR	26.0 (22.5)	0.30 (0.28)	45.0 (41.8)	103	3.8
(Mid-Atlantic)	SNOW	36.0 (35.0)	0.70 (0.67)	82.9 (80.1)	585	21.6
7	ZR	22.0 (22.0)	0.24 (0.24)	38.0 (37.6)	145	2.9
(South)	SNOW	27.0 (25.0)	0.58 (0.58)	77.1 (77.4)	276	5.6

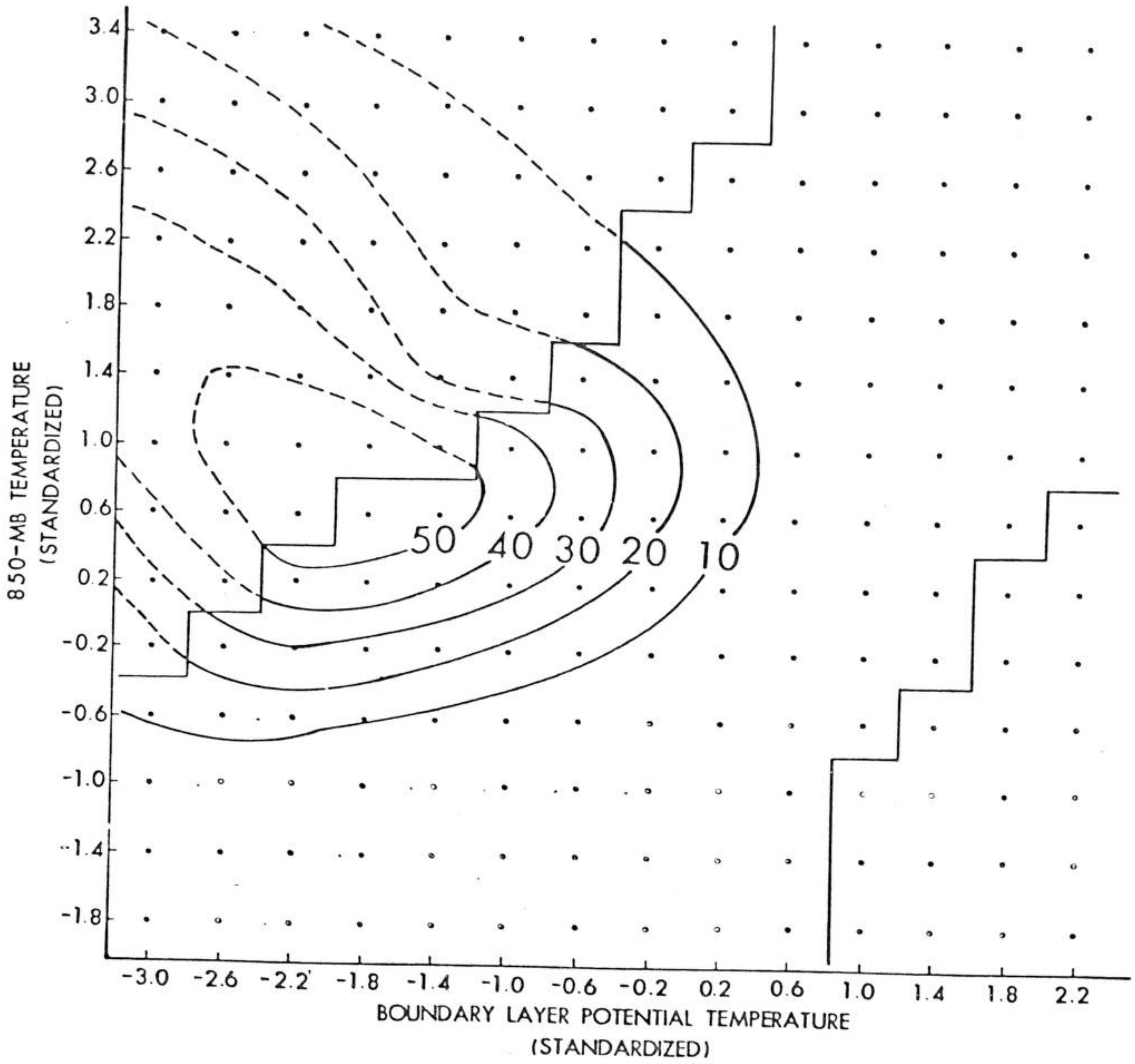

Figure 1. The empirical probability of ZR (%), given that precipitation occurs, as a function of the 850 T and BLPT. The predictors are forecast values from the LFM model and are standardized (see text for further explanation). The stepped lines separate areas with few or no data from areas with sufficient data.

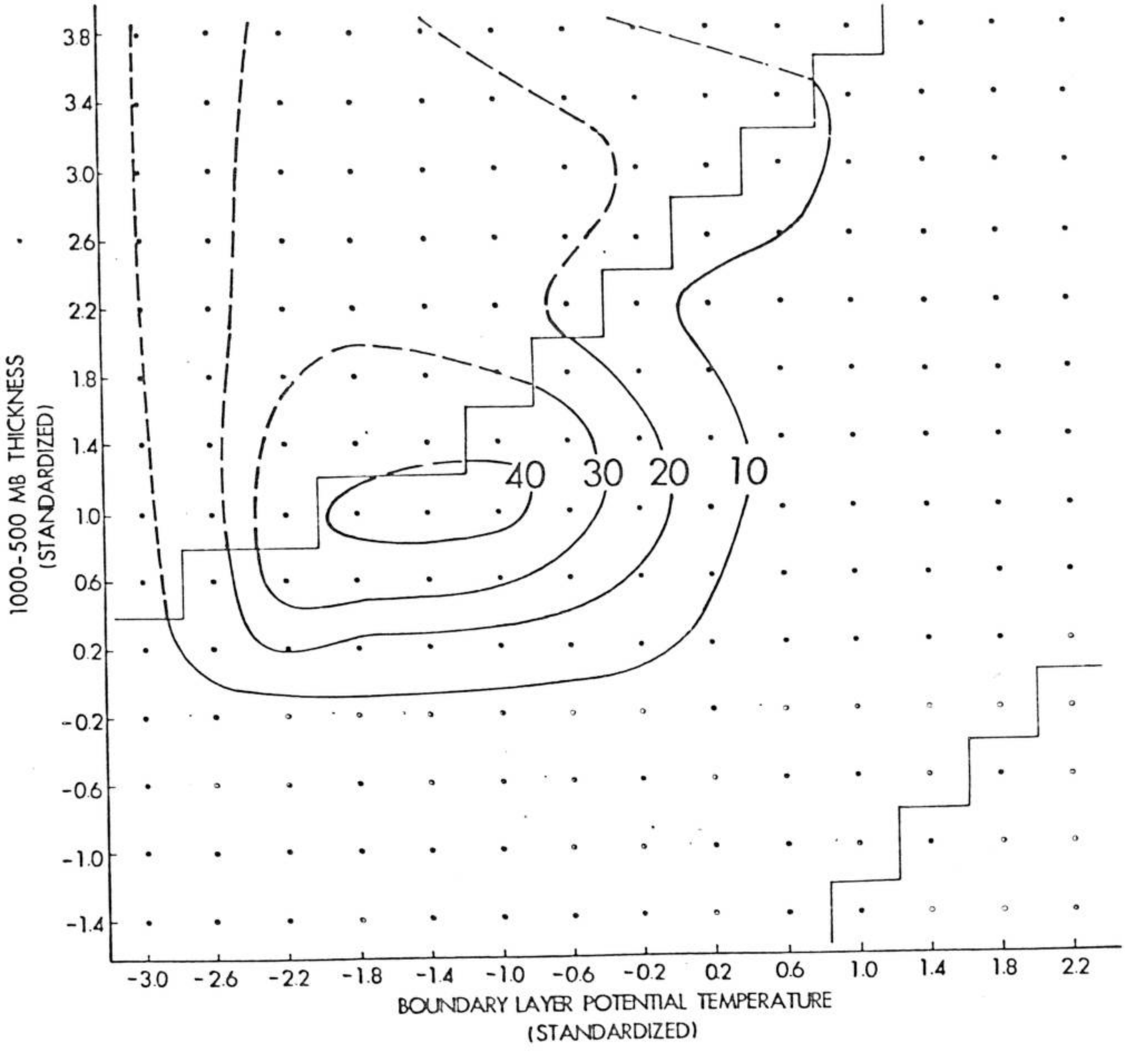

Figure 2. The same as Fig. 1 except that the 10-5 Th and BLPT are used as predictors.

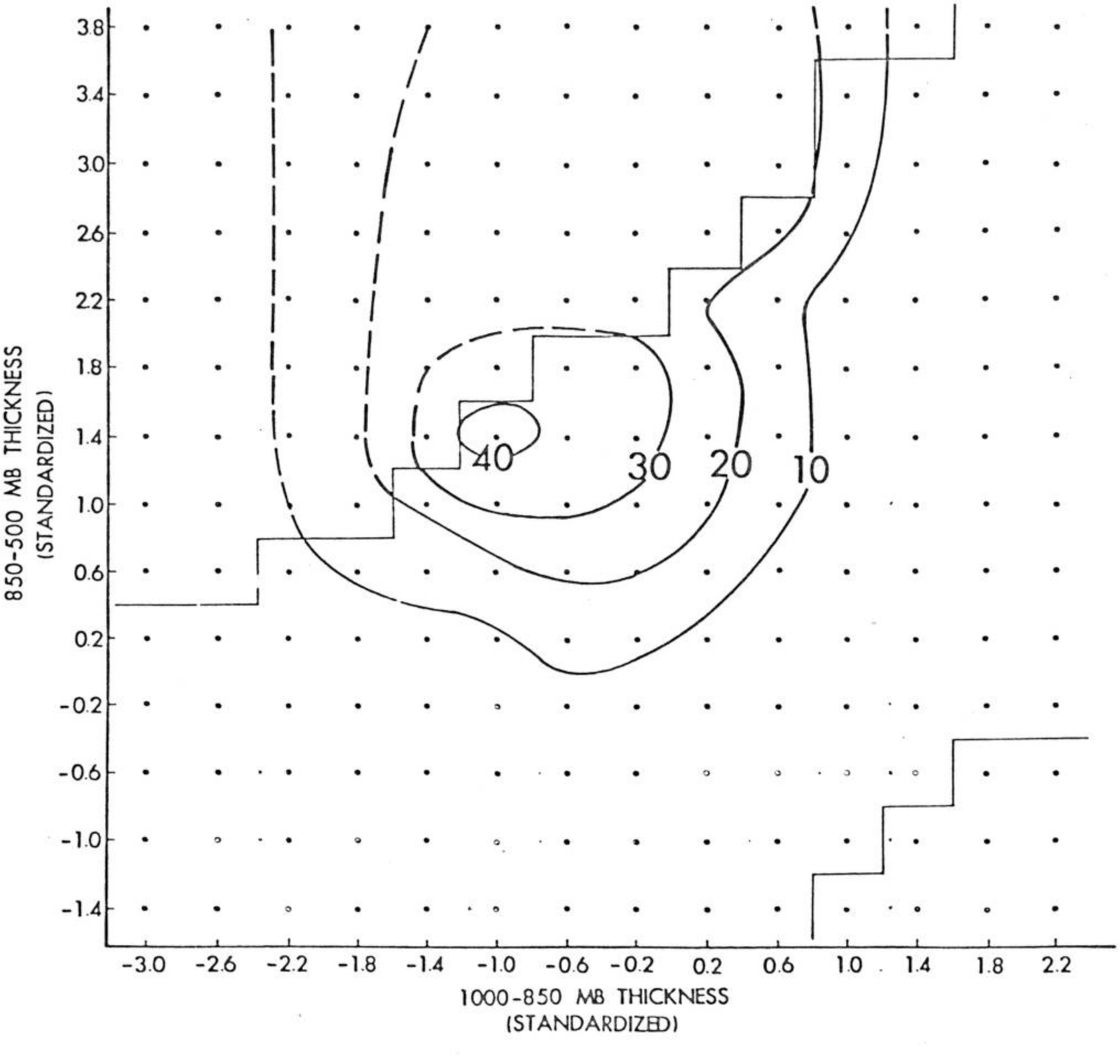

Figure 3. The same as Fig. 1 except that the 8.5-5 Th and 10-8.5 Th are used as predictors.

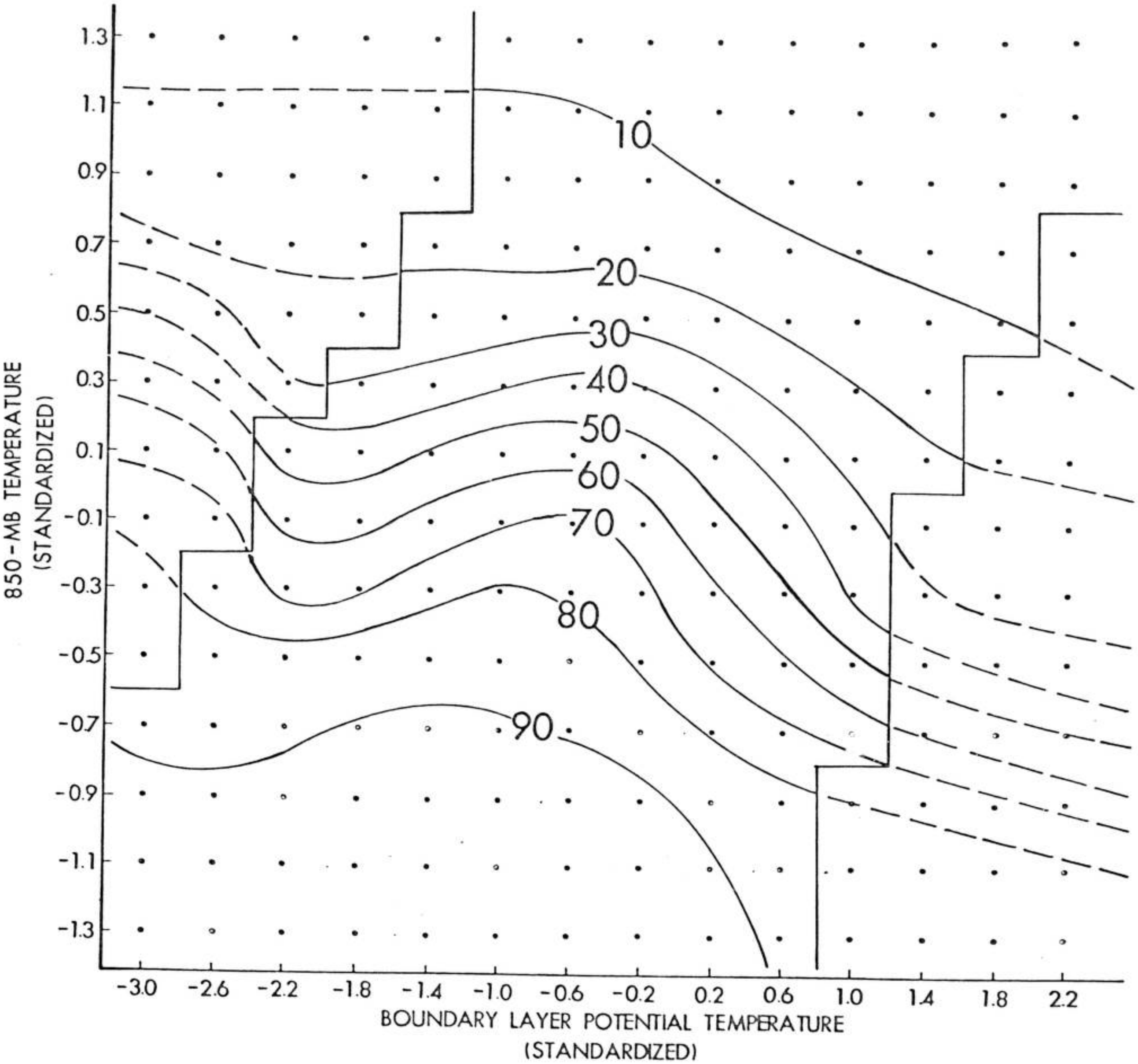

Figure 4. The same as Fig. 1 except that the empirical probability of the SNOW category is shown.

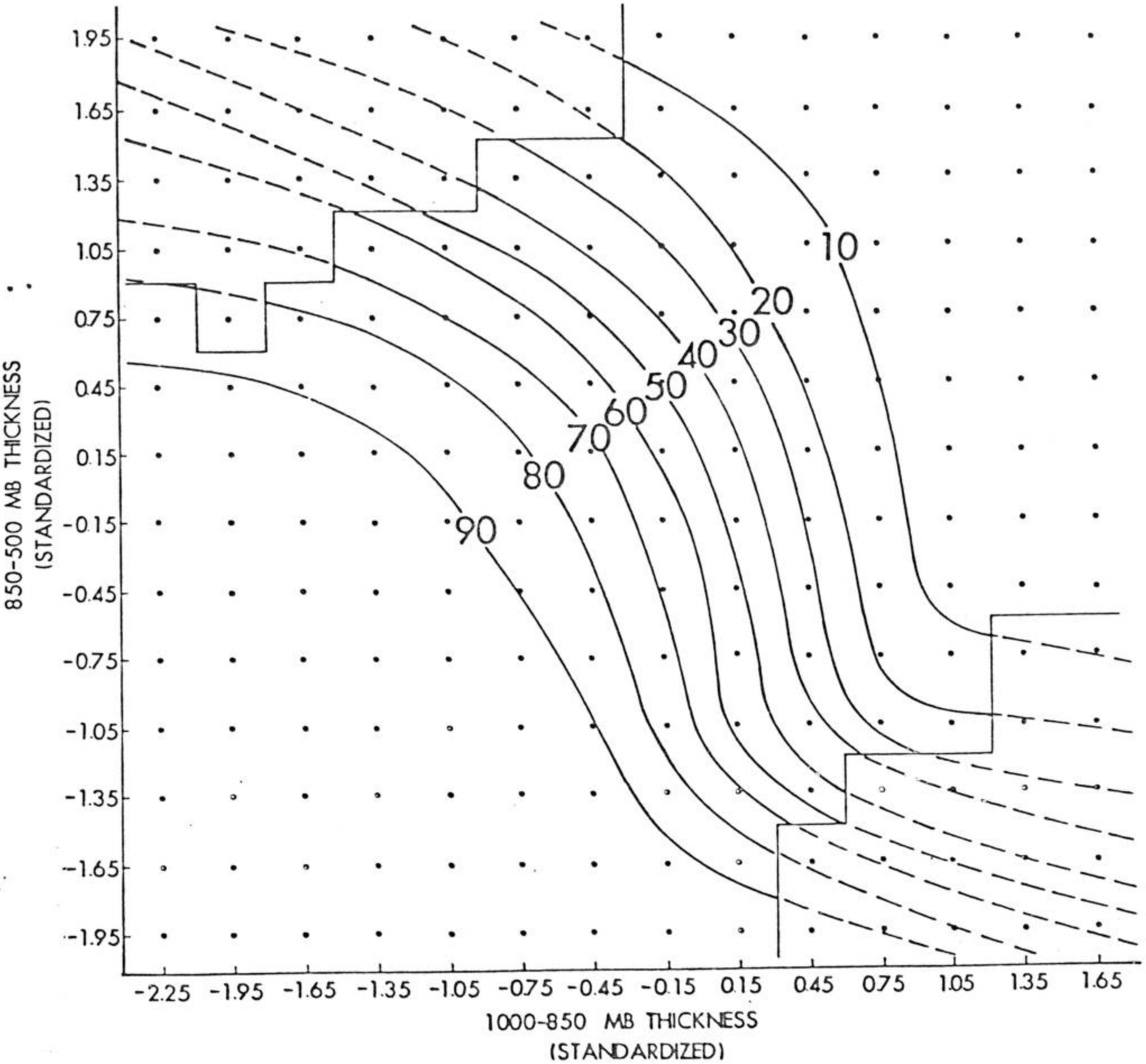

Figure 5. The same as Fig. 3 except that the empirical probability of the SNOW category is shown.

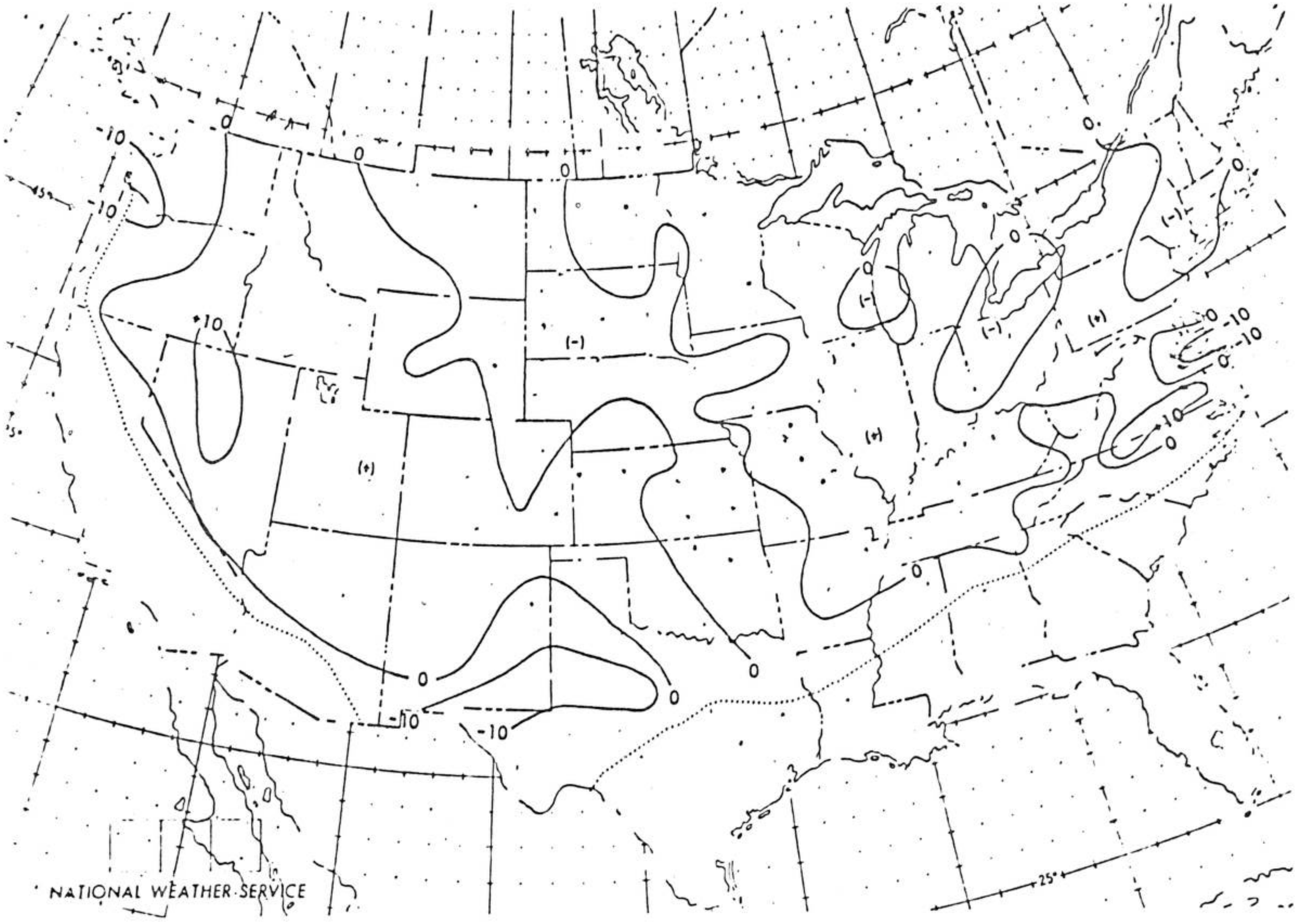

Figure 6. The relative probability bias (%) (see text for definition) for SNOW averaged for 12- and 24-h forecasts from both the 0000 GMT and 1200 GMT. The probability of SNOW forecasts were made from generalized-operator REEP equations for the developmental sample, the winter seasons of 1972-73 through 1979-80. Values >0% (<0%) indicate overforecasting (underforecasting) of the SNOW category. An analysis was not done for areas below the dotted line because the infrequent occurence of SNOW made analysis difficult.

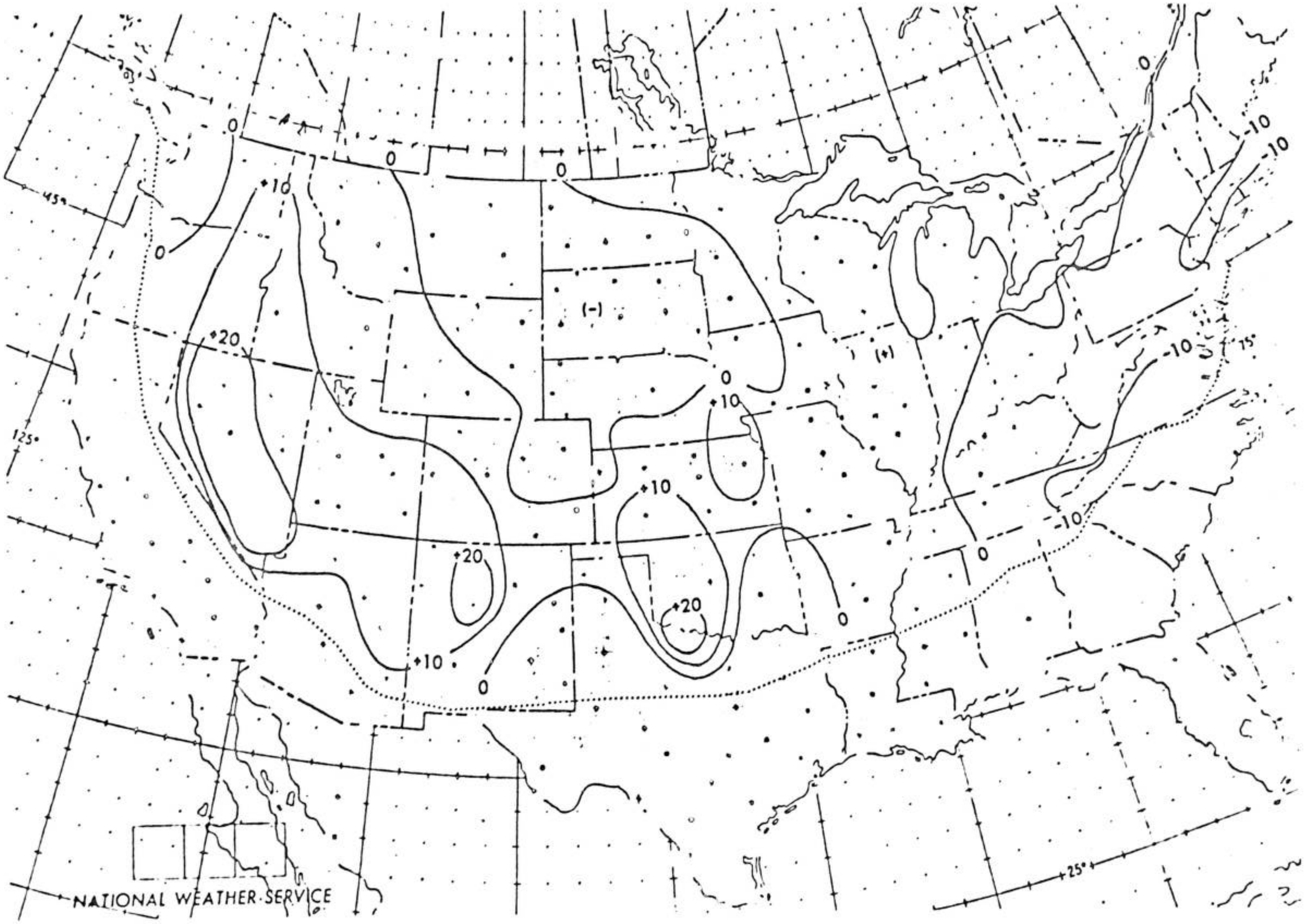

Figure 7. The same as Fig. 6 except that the relative probability bias was averaged from the 36- and 48-h projections from both the 0000 GMT and 1200 GMT cycles.

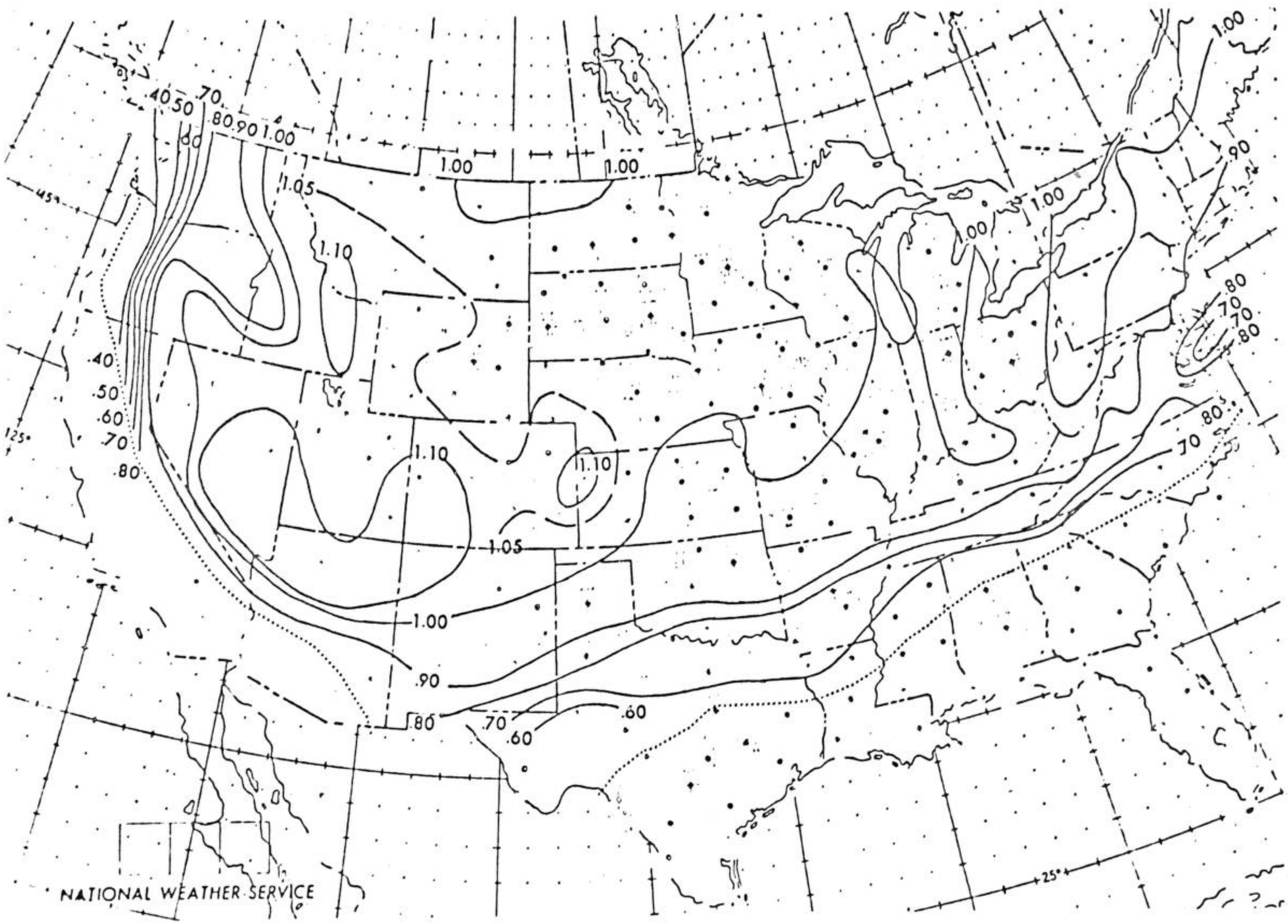

Figure 8. The same as Fig. 6 except that the categorical bias (see text for definitions) is shown. Values >1.00 (<1.00) indicate overforecasting (underforecasting) of the SNOW category.

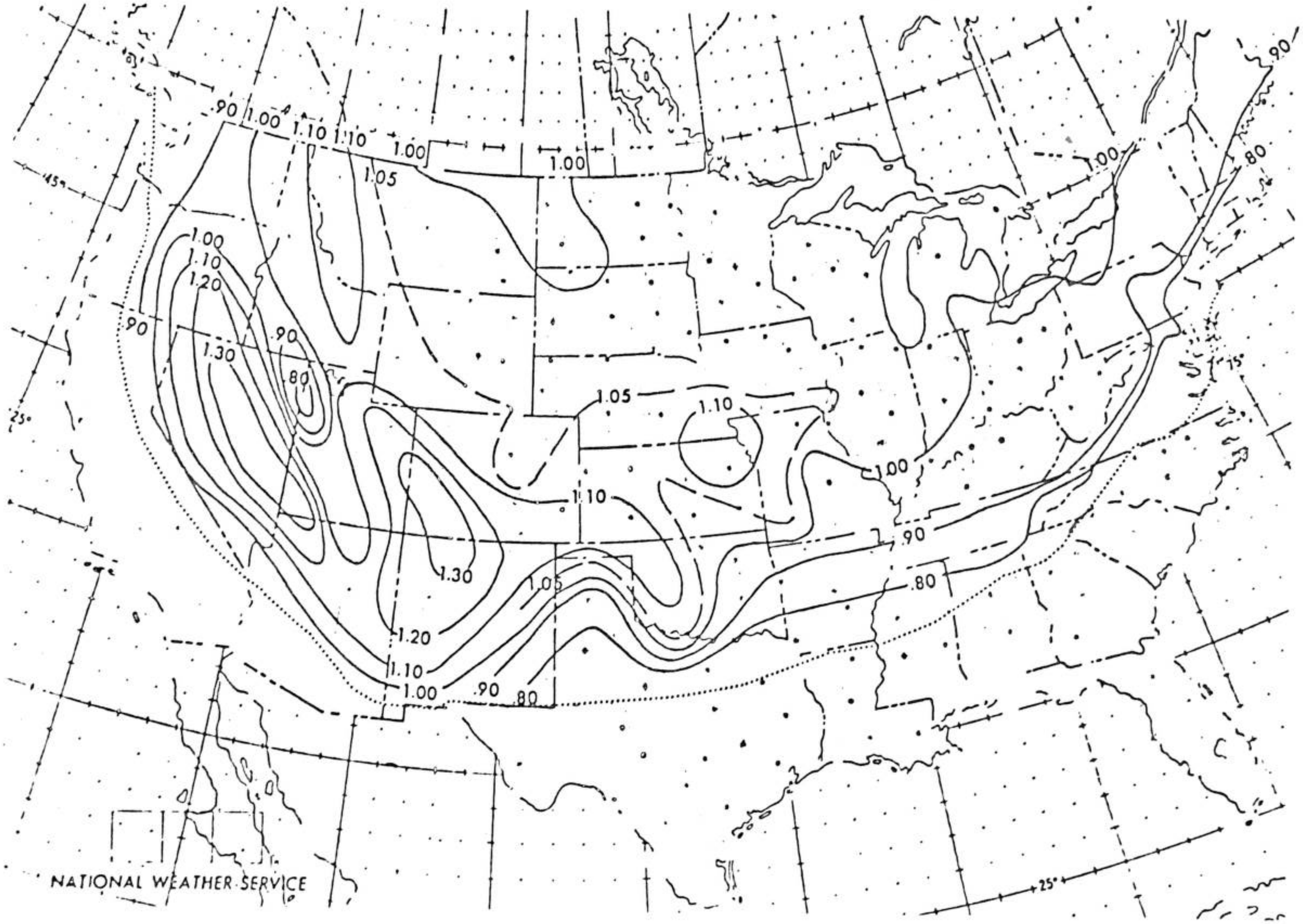

Figure 9. The same as Fig. 8 except that the categorical bias was averaged from the 36- and 48-h projections from both 0000 GMT and 1200 GMT.

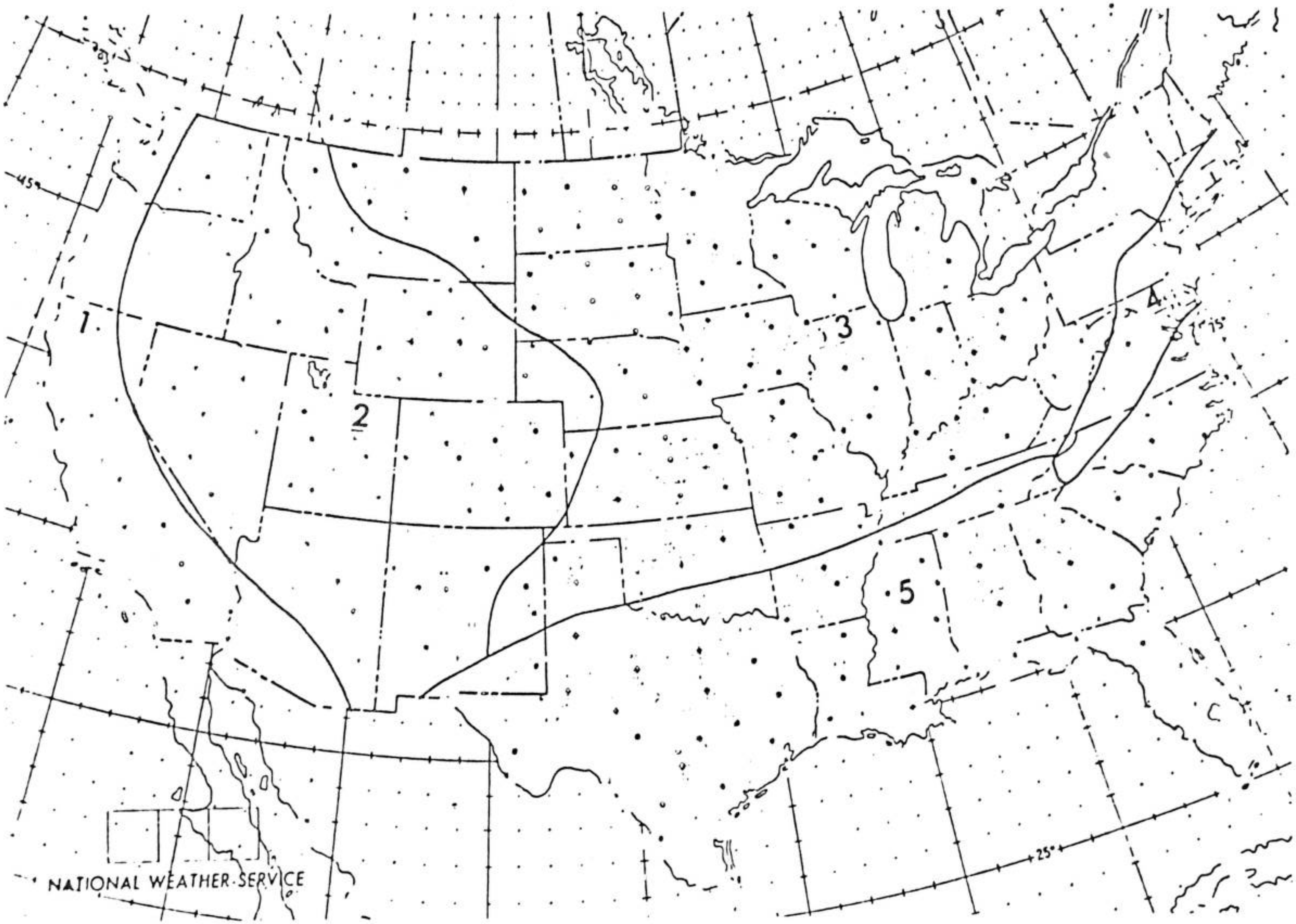

Figure 10. The regions used in the development of the EXP probability forecast equation sets for PoPT.

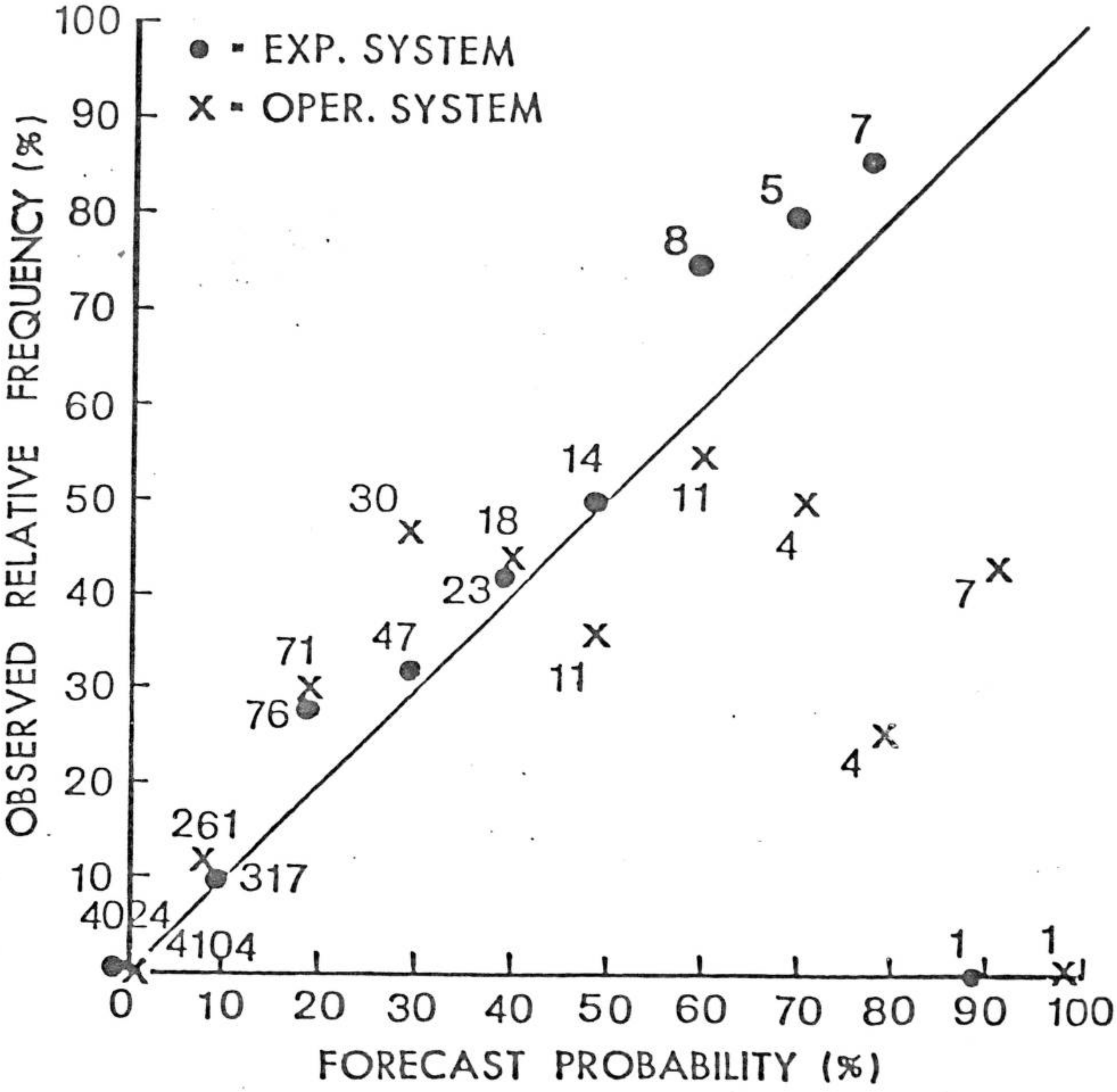

Figure 11. The reliability of the ZR probability forecasts for the 12-h projection from 0000 GMT. The dots (●) represent the EXP forecasts, while the crosses (X) represent the OPER forecasts. Independent data were combined from 229 stations for September 1980-March 1981. The number of cases for each dot and cross is also shown. The line denotes perfect reliability.

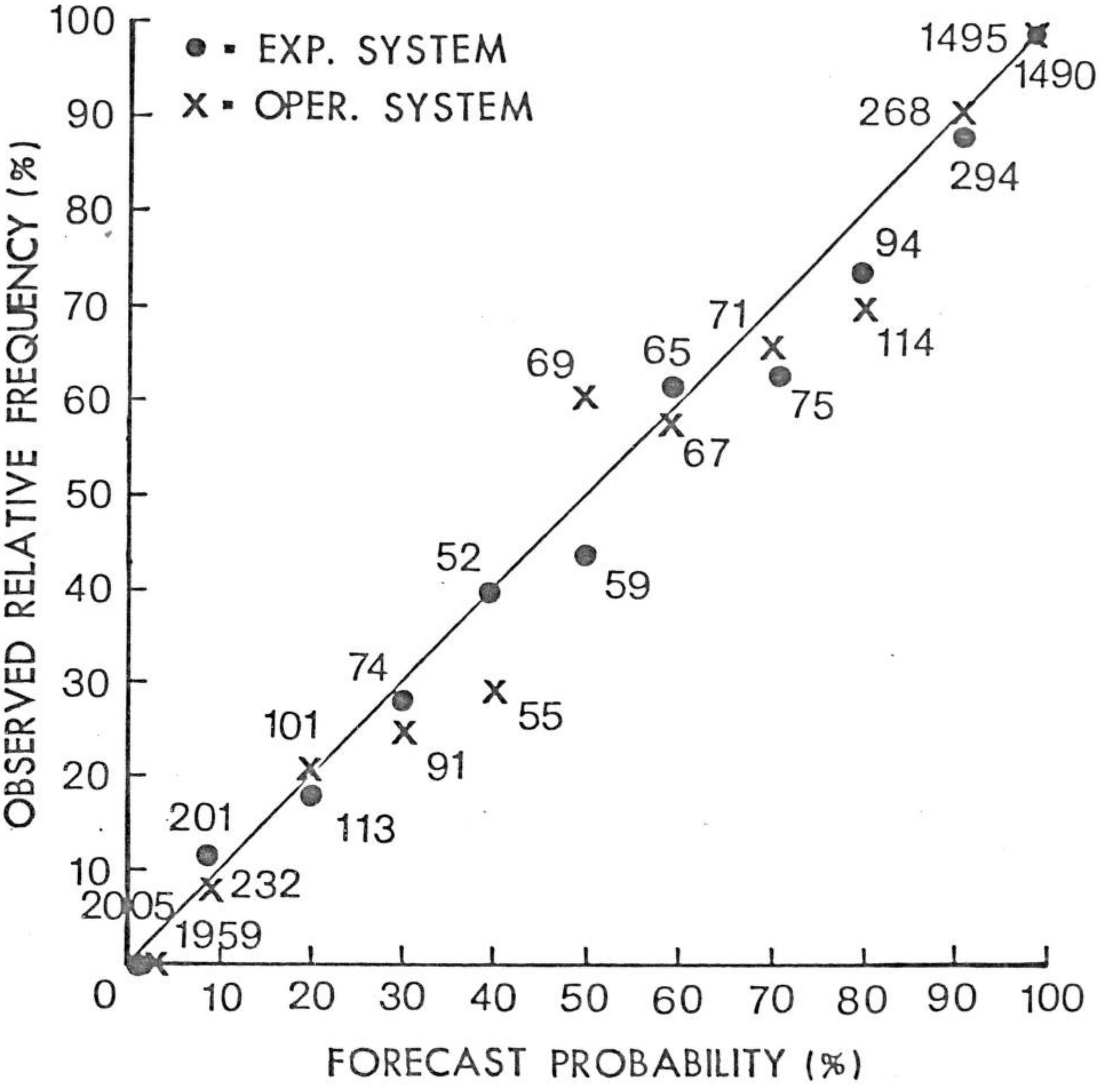

Figure 12. The same as Fig. 11 except that the reliability for the SNOW category is shown.

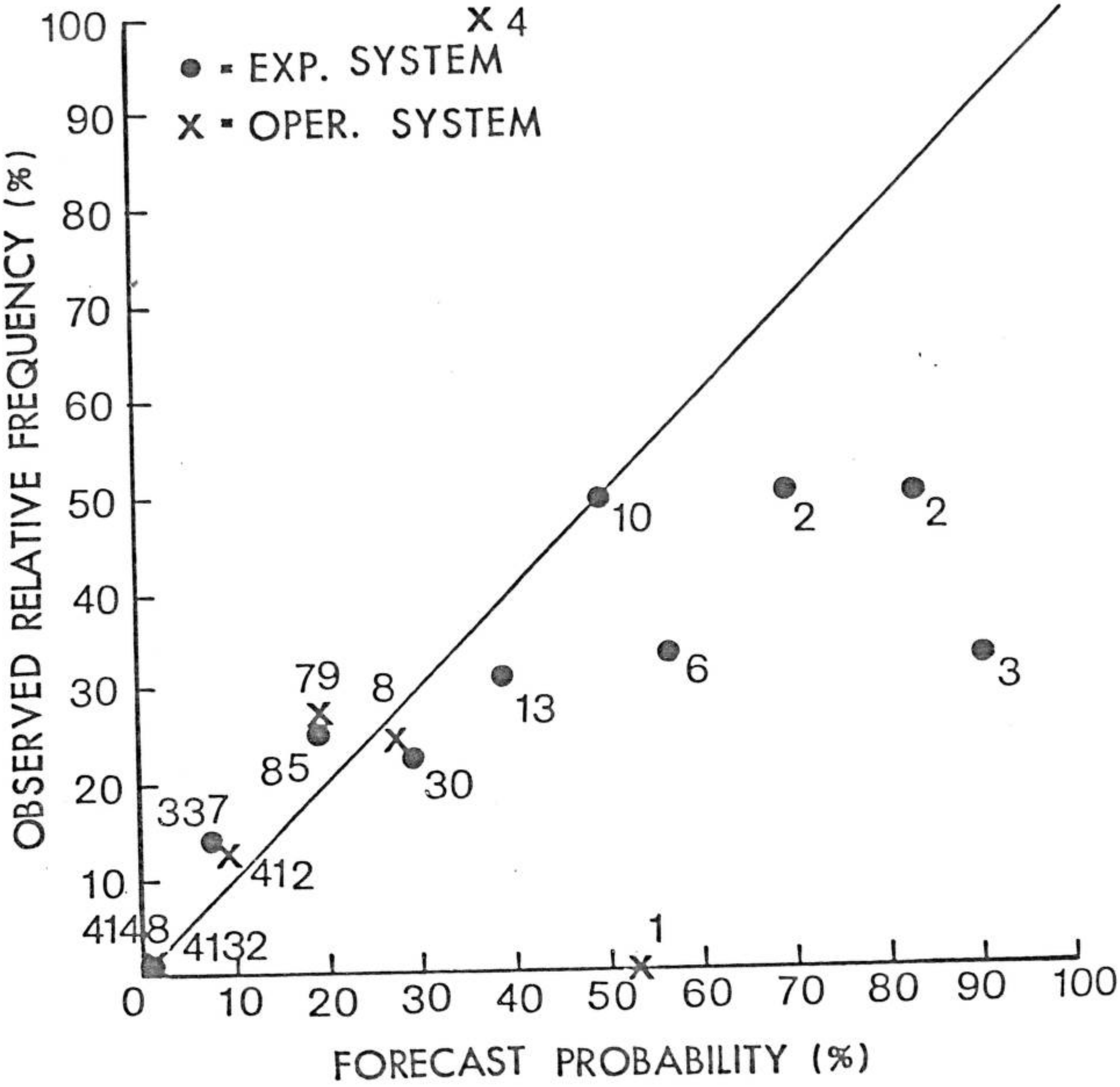

Figure 13. The same as Fig. 11 except that the reliability for the 36-h projection is shown.

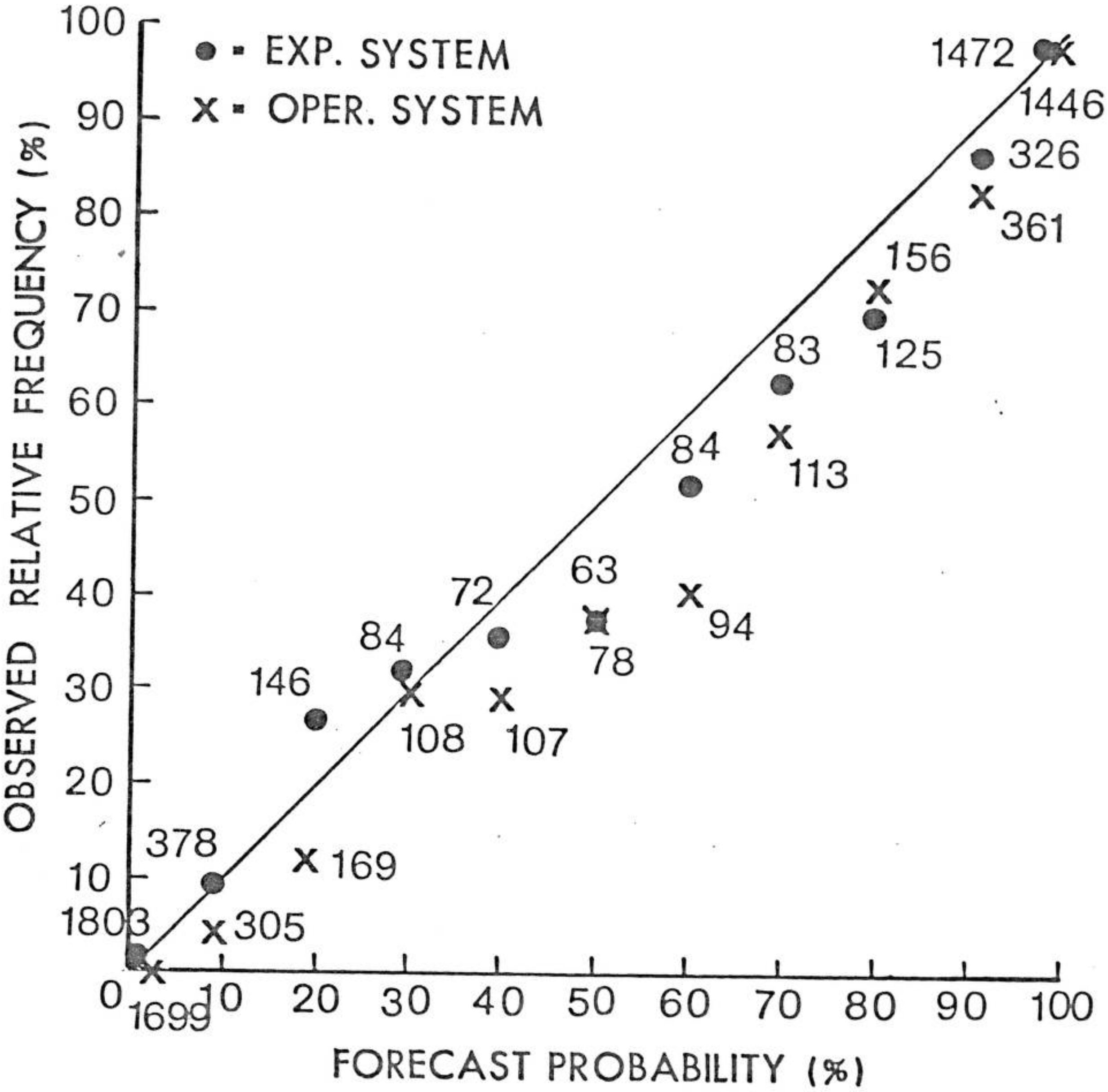

Figure 14. The same as Fig. 13 except that the reliability for the SNOW category is shown.

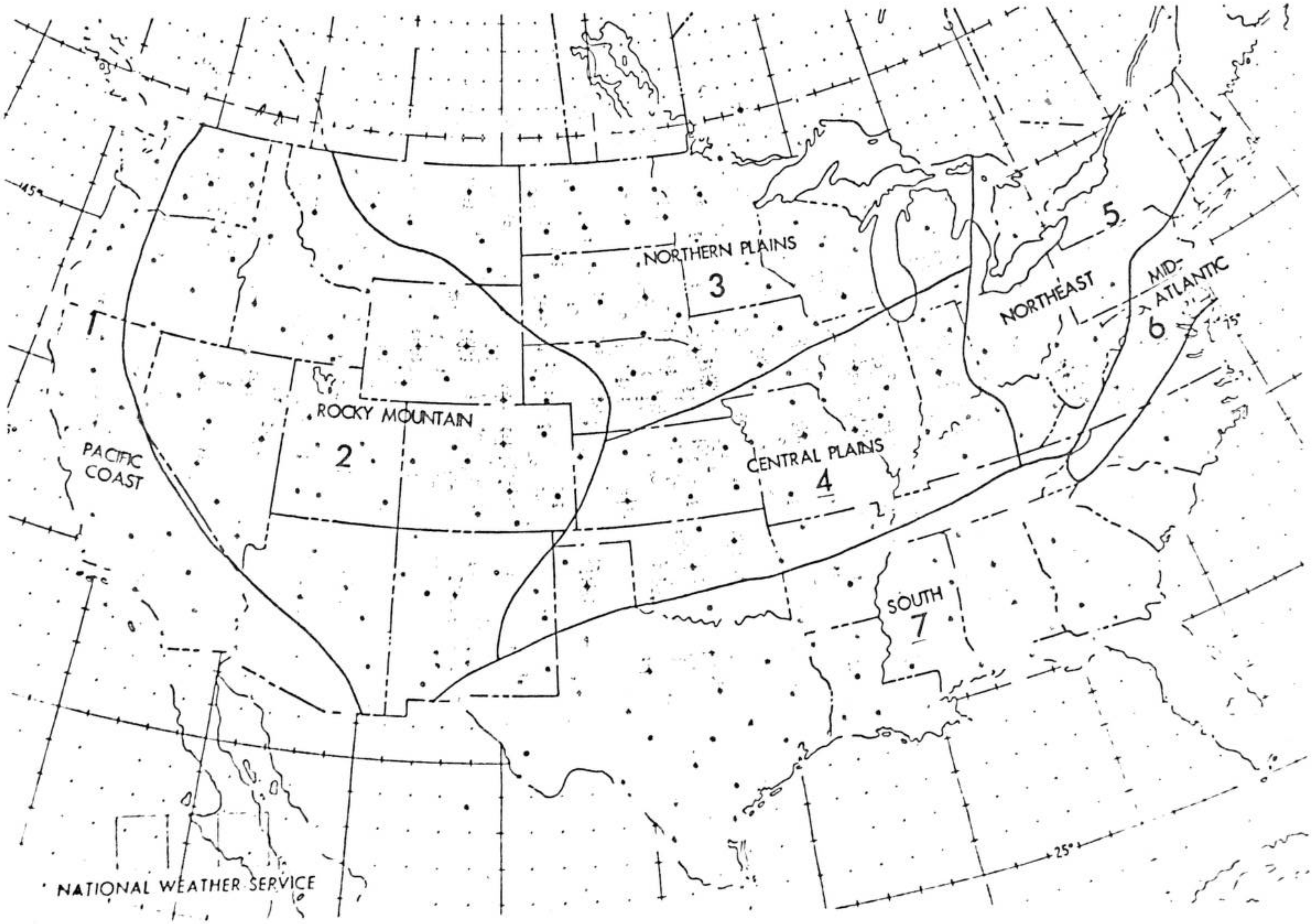

Figure 15. The regions used for the development of threshold probabilities for the ZR and SNOW categories for the EXP logit forecast equations. The threshold probabilities are used to transform probability forecasts into categorical forecasts.

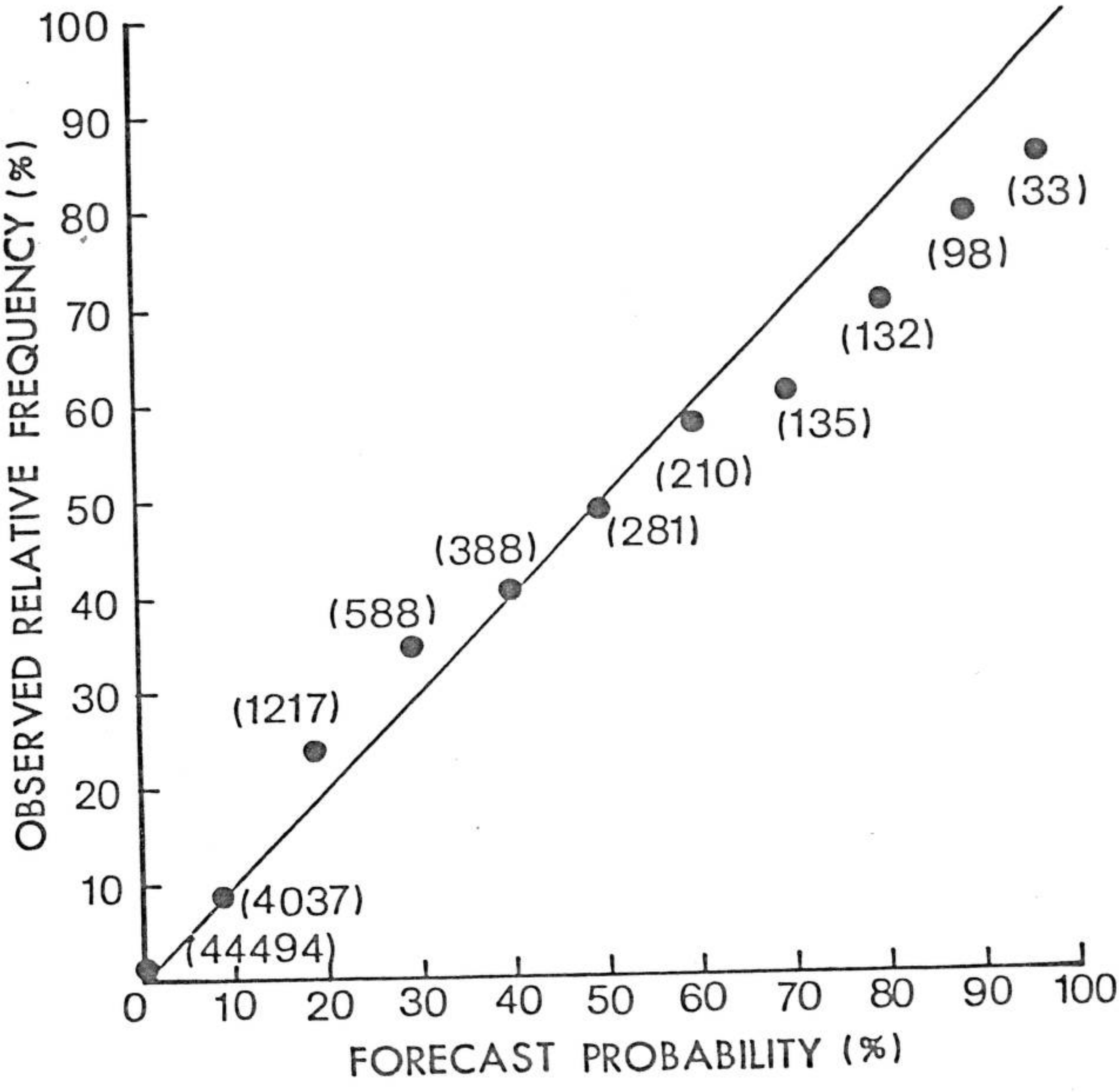

Figure 16. The reliability of the ZR probability forecasts for the developmental data sample for the 12-h projection from 0000 GMT for 229 stations. The number of cases represented by each dot is shown in parentheses. The line denotes perfect reliability.

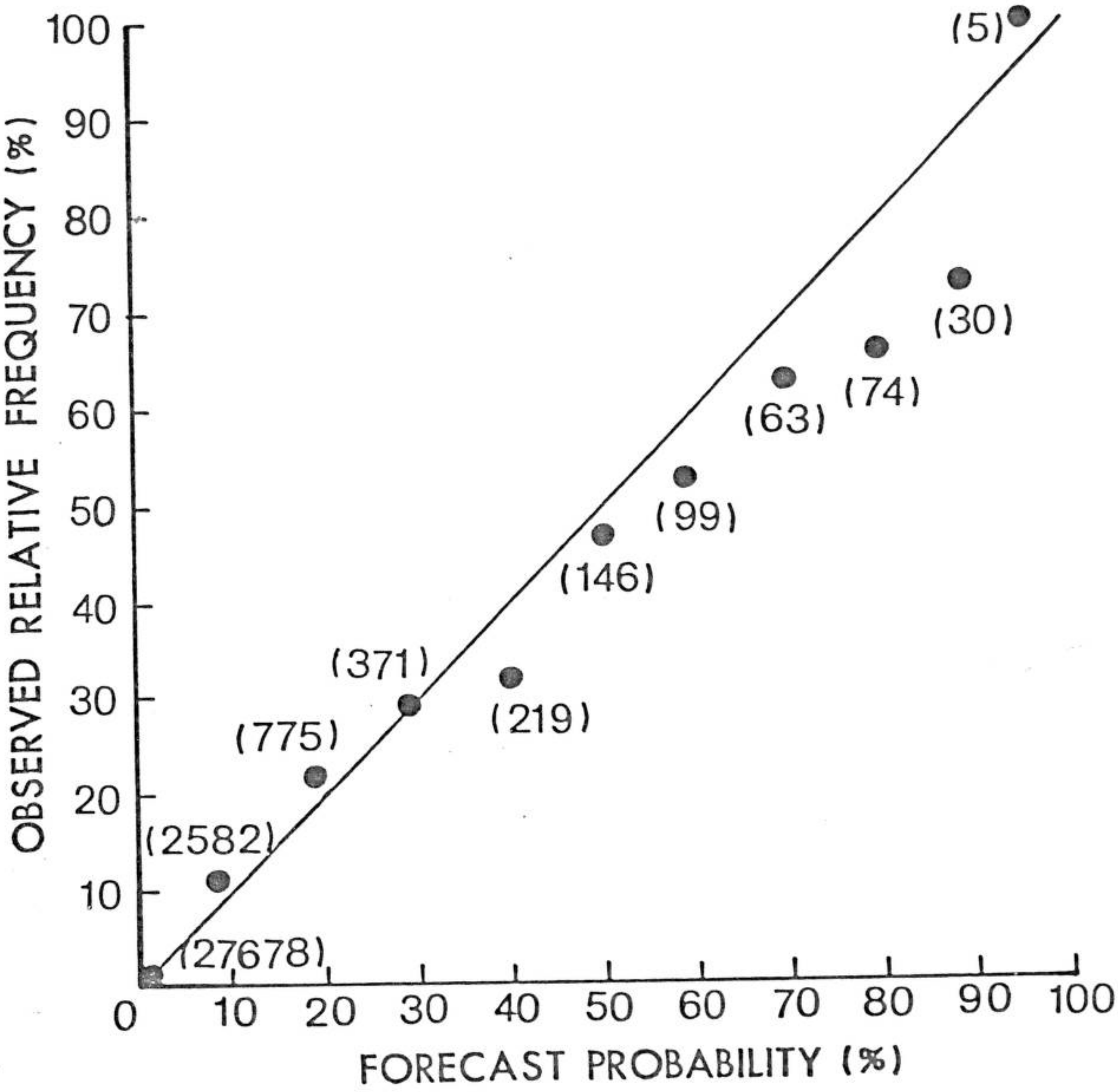

Figure 17. The same as Fig. 16 except the reliability is for the 36-h projection.