

An Improved Operational System for Forecasting Precipitation Type

JOSEPH R. BOCCHIERI¹ AND GEORGE J. MAGLARAS

Techniques Development Laboratory, National Weather Service, NOAA, Silver Spring, MD 20910

(Manuscript received 1 May 1982, in final form 7 December 1982)

ABSTRACT

A Model Output Statistics system for forecasting the conditional probability of precipitation type (PoPT) became operational within the National Weather Service in September 1978. Forecasts are provided for three precipitation type categories: snow or ice pellets, freezing rain, and rain. To develop the forecast equations, data are combined from different stations because of the limited amount of developmental data. To justify combining the data, the Limited-area Fine Mesh (LFM) model predictors are transformed from their original values through the use of the logit model. In one experiment, it is shown that probability of snow forecasts are made more accurate through an improved use of the logit model for predictor transformation.

The new transformation procedure is then used in the development of a set of experimental PoPT forecast equations. The experimental equations differ from the operational equations in other ways also. The developmental sample for the experimental equations included approximately three winter seasons more data than the sample used for the operational system. Also, improvements are made to the potential predictors used to develop the experimental equations. Finally, freezing rain mixed with any other precipitation type is defined as freezing rain in the experimental system; in the operational system, this mixture of precipitation is defined as rain.

A comparative verification between the experimental and operational systems on independent data indicates that, overall, the experimental PoPT forecasts are better than the operational forecasts, especially for 12–24 h freezing rain forecasts. Based on these results, new operational PoPT forecast equations are developed incorporating the features associated with the experimental equations. The new system was implemented in the fall of 1982.

1. Introduction

A system for forecasting the conditional probability of precipitation type (PoPT) (Bocchieri, 1979a) became operational within the National Weather Service in September 1978. PoPT gives forecasts for three precipitation type categories: snow or ice pellets (SNOW), freezing rain or drizzle (ZR), and rain or mixed 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.

The Model Output Statistics (MOS) technique (Glahn and Lowry, 1972) was used to develop the PoPT system. In MOS, the predictand is statistically related to variables that have been predicted by a numerical model or models. For PoPT, output from the Limited-area Fine Mesh (LFM) model (Gerrity, 1977) and a finer mesh version of the LFM called LFM-II (Newell and Deaven, 1981) was used. The PoPT system evolved from the conditional probability of frozen precipitation system (Glahn and Bocchieri, 1975; Bocchieri and Glahn, 1976) which be-

came operational within the National Weather Service in November 1972; in that system, explicit probability forecasts for ZR weren't available.

We have recently experimented with a number of ways to improve the operational system. For instance, to derive the forecast equations, we needed to combine data from different stations because of the limited amount of developmental data. To make combining the data more palatable, several of the more important LFM predictor variables were transformed from their original values through the use of the logit model (Brelford and Jones, 1967; Jones, 1968). The logit model provides a means of fitting a sigmoid or S-shaped curve when the dependent variable is binary and the independent variable is continuous. The pilot study described in Section 2 shows how we improved the use of the logit model for the purpose of transforming predictors.

As described in Section 3, we then used the improved transformation procedure to develop a set of experimental PoPT forecast equations. The experimental equations also differ from the operational system in other ways. For instance, the developmental sample for the experimental equations included approximately three winter seasons (September through April), more data than the sample used for the op-

¹ Present affiliation: National Weather Service Forecast Office, World Weather Building, Camp Springs, MD 20233.

erational equations. Also, we improved the potential predictors used to develop the experimental equations. Finally, freezing rain or drizzle mixed with any other precipitation type was included in the ZR category for the experimental equations; in the operational system, this mixture of precipitation was included with the RAIN category.

We did a comparative verification between the experimental and operational systems on independent data (Section 4). The results indicate that, overall, the experimental PoPT forecasts were better than those from the operational system, especially for 12–24 h ZR forecasts. Based on these experimental results, we derived new operational PoPT forecast equations. The new system was implemented in the fall of 1982.

2. Experiments in the transformation of predictors

Prior to developing the operational PoPT forecast equations, we transformed the LFM thermodynamic predictor variables into deviations from 50% values. 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% values for each predictor and each station by using the logit model to fit the data.

As discussed in Glahn and Bocchieri (1975), the 50% value of a variable can vary quite a bit from station to station depending on local factors, especially station elevation. Our initial assumption was that a given *deviation* of a predictor from its 50% value would 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 mb temperature between the 50% and the 95% points of the curve might be 4 K; however, for a shallow logit curve, this difference might be 8 K.

We hypothesized that we could improve the accuracy of PoPT forecasts by transforming predictors to account not only for the difference in 50% values between stations but also for the difference in steepness or spread of the logit curves. We tested this hypothesis in the experiment described later. First, some properties of the logit model are presented.

a. Some properties of the logit model

For the case of one independent variable, the logit model provides a means of fitting an S-shaped, symmetric curve where the dependent variable Y is binary and the independent variable X is continuous. The probability of the binary variable having the value of 1 can be expressed as

$$P(Y = 1|X) = \frac{1}{1 + \exp - (\alpha + \beta X)}. \quad (1)$$

For our purpose, Y takes the value 1 for an observation of SNOW and 0 otherwise; the probability is conditional on the event that precipitation occurs. Hereafter, the expression $P(\text{SNOW})$ will be used instead of $P(Y = 1|X)$.

The computer program we use (Jones, 1968) determines the maximum likelihood estimates for the model parameters. As discussed in Cox (1970) and illustrated by Fig. 1, the logit curve has the following properties:

- 1) The value of X at which $P(\text{SNOW})$ is 50% is given by $-\alpha/\beta$.
- 2) The parameter β measures the steepness or slope of the logit curve; the larger the $|\beta|$, the steeper the curve.
- 3) The parameter β is such that the distance between the 95% point and the 50% point is approximately $3/\beta$. Bocchieri (1979b) showed that this latter distance is actually equal to $2.944/\beta$.

For our purposes, we define “slope” and “spread” parameters by

$$\text{Slope} = |\beta|, \quad (2)$$

$$\text{Spread} = |2.944/\beta|. \quad (3)$$

The logit curve, therefore, is completely determined by two parameters—the 50% value locates the midpoint of the curve (with respect to the X axis) and the spread gives the shape of the curve.

The problem in combining data from different stations is illustrated in Fig. 2 which shows the logit curves for Fort Smith, Arkansas and Sheridan, Wyoming. For these stations, $P(\text{SNOW})$ is shown as a function of the LFM 850 mb temperature forecast. The developmental sample for the curves consisted of five winter seasons, 1972–73 through 1976–77. For each station, we matched 850 mb temperature forecasts from the LFM model and corresponding surface observations of precipitation type for seven projections—6, 9, 12, 15, 18, 21 and 24 h. The data from all projections and from both the 0000 and 1200 GMT forecast cycles were combined into one sample so that as many SNOW cases as possible would be included. We then fit the logit model to the data and computed the 50%, slope and spread parameters. In Fig. 2, note that not only does the 50% value of the curves differ but also do the slope and spread. The logit curve for Fort Smith is steeper (slope = 1.19) than the curve for Sheridan (slope = 0.34), and therefore, the spread of the curve for Fort Smith (2.47) is less than that for Sheridan (8.66). Obviously, if the 850 mb temperature is transformed into deviations from 50% values, then a given deviation would *not* give the same $P(\text{SNOW})$ at Fort Smith and Sheridan. In fact, a -2 K deviation gives about a 90% chance of SNOW at Fort Smith but only a 65% chance at Sheridan. This illustrates the need to account not only for the difference in the 50% value between the curves but also for the difference in the slope or spread.

The variation between stations of the 50% value of a given predictor has been well documented and illustrated by Glahn and Bocchieri (1975) and Bocchieri and Glahn (1976). Fig. 3 shows the variation of the slope parameter for the 850 mb temperature logit curves for the conterminous United States; the logit curve for each station was derived from the same sample used to obtain the curves in Fig. 2. The analysis shows that the slopes for stations in the Rocky Mountain region are generally less than the slopes for stations to the east and west of that region. In the Rocky Mountain region, the slopes are generally <0.5; in the eastern half of the United States and near the Pacific Northwest coast, the slopes are generally between 0.6 and 0.9.

b. Testing of predictor transformation methods

We compared two predictor transformation procedures, called the centered and standardized procedures. In the centered procedure,

$$X_T = X - (50\% \text{ value}), \tag{4}$$

where X_T is the transformed variable, and X the original variable. In this manner, the difference in the 50% value of the logit curve from station to station is accounted for. The centered procedure was used in the development of the operational PoPT system (Bocchieri, 1979a). In the standardized procedure,

$$X_T = \frac{X - (50\% \text{ value})}{\text{spread}}. \tag{5}$$

That is, the original variable is transformed such that the differences in 50% values and spread parameters from station to station are accounted for. The standardized procedure is analogous to transforming a variable having a normal distribution, for instance,

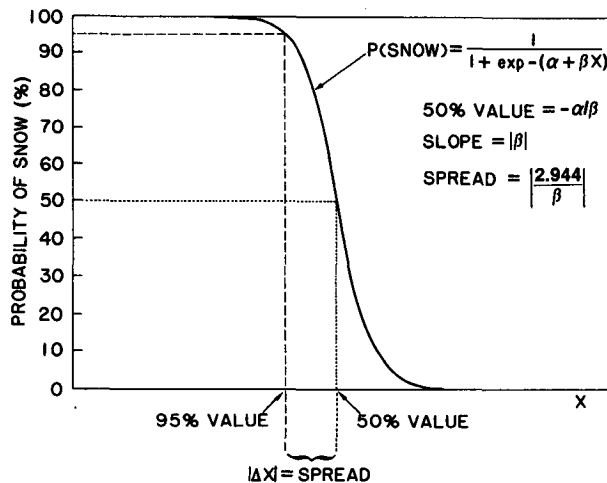


FIG. 1. A hypothetical logit curve and its associated parameters (see text for further explanation).

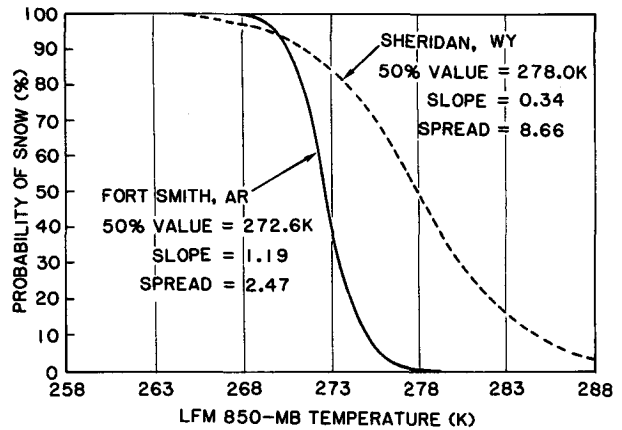


FIG. 2. The logit curves for Fort Smith, Arkansas and Sheridan, Wyoming. The curves give the conditional probability of SNOW as a function of the LFM 850 mb temperature forecast and were derived with data from five winter seasons, 1972-73 through 1976-77. The 50%, slope and spread parameters are discussed in the text.

into a standard normal deviate by subtracting the mean and dividing by the standard deviation (Brownlee, 1967).

In order to compare the centered and standardized procedures, we developed two systems to forecast the conditional probability of SNOW for the 18 h projection from 0000 GMT. For each system, we used three predictors in transformed form: 850 mb temperature, 1000-500 mb thickness, and boundary layer wet-bulb temperature. These predictors account for much of the useful information from the LFM model with respect to probability of SNOW forecasting. The two systems are called the centered and standardized systems, respectively, to correspond to the transformation procedure used in each. In the centered system, Eq. (4) was used to transform each predictor at each station for each precipitation case in the sample; similarly, Eq. (5) was used for the standardized system. After transforming the predictors and combining data from 174 conterminous U.S. stations, we used the logit model to develop the forecast equations. The developmental data period was the same as that used for Figs. 2 and 3.

We verified the centered and standardized systems on both developmental and independent samples; the independent sample consisted of data combined from all 174 stations from September 1977 through February 1978. The Brier scores (Brier, 1950) for both the developmental and independent samples and the improvement in Brier score of the standardized system over the centered system are shown in Table 1. The results indicate that the standardized system was better than the centered system by 1.9% for the developmental sample and by 2.4% on the independent sample. We concluded that the accuracy of probability of SNOW forecasts could be improved through

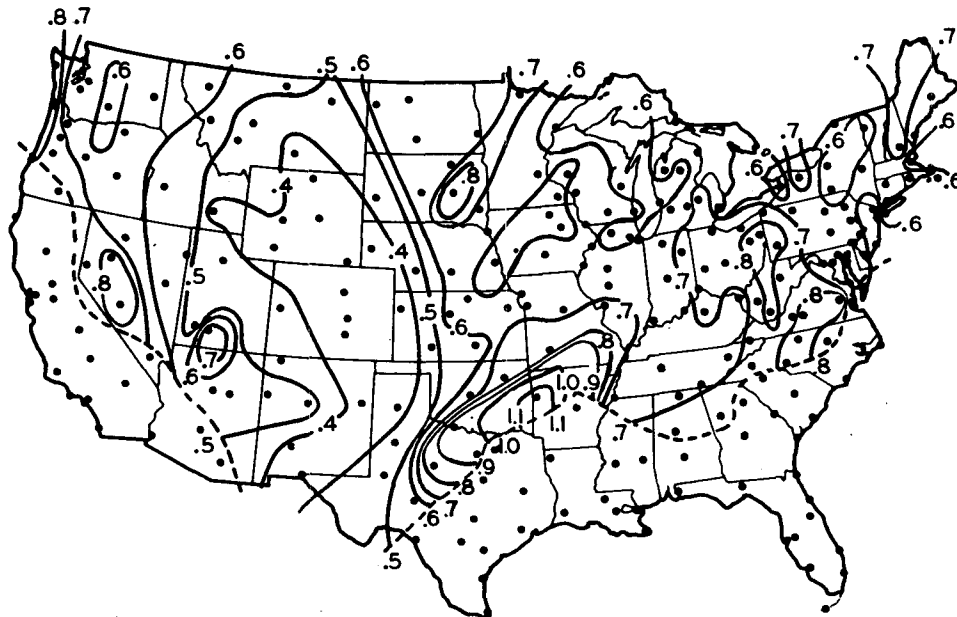


FIG. 3. An analysis of the slope of the 850 mb temperature logit curves. The logit curve for each station (shown by dots) was developed from the same sample used for the curves in Fig. 2. The logit curves for the stations below the dashed line couldn't be determined objectively due to insufficient data.

the use of the standardized transformation procedure. A similar improvement should be realized for ZR forecasts since these forecasts are a function of predictors which are transformed with respect to the SNOW category. As described in the next section, we used the standardized transformation procedure in the development of an experimental set of PoPT forecast equations.

3. Development of experimental PoPT equations

We developed a set of experimental PoPT forecast equations, called EXP, for the 12, 24, 36 and 48 h projections from 0000 GMT. 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, developmental data were available for 4½ winter seasons, February 1976 through April 1980. The developmental sample for the operational PoPT system, hereafter called OPER, consisted of about five winter seasons for the ≤24 h projections and 2–3 winter seasons for the longer-range projections. Data from September 1980 through March 1981 were set aside for use as an independent sample to make comparisons between the EXP and OPER systems.

As in the OPER system, the logit model was used to develop the statistical forecast equations for EXP. Our present logit computer program doesn't have a screening option; therefore, as for OPER, we used forward screening in a statistical technique known as Regression Estimation of Event Probabilities

(REEP) (Miller, 1964) to select predictors for the logit equations. A good description of the REEP screening procedure can be found in Glahn and Lowry (1972).

a. The potential predictors

Table 2 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 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

TABLE 1. The Brier scores for conditional probability of SNOW forecasts made from the centered and standardized forecast systems for the 18 h projection from 0000 GMT. The percent improvement in Brier score of the standardized system over the centered system is also shown. The number of precipitation cases is shown in parentheses. Data were combined from 174 stations. The developmental sample consisted of data from the winter seasons of 1972–73 through 1976–77. Independent data were from September 1977 through February 1978.

Forecast system	Brier score	
	Developmental data (23292)	Independent data (4737)
Centered	0.106	0.083
Standardized	0.104	0.081
Percent improvement	1.9	2.4

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 h. Also, for EXP we used LFM *u*- and *v*-wind component forecasts at various levels as predictors 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 h. Another difference is that LFM boundary layer wet-bulb temperature was not included as a potential predictor for EXP, the reason being that a change in October 1979 of the earth's terrain used in the LFM model affected the computation of wet-bulb temperature. We also included climatic frequencies of the ZR and SNOW categories as predictors in EXP. These frequencies weren't used to develop OPER. We derived the frequencies for each of ~ 230 stations for each of the months September through April by combining the 3 h surface observations of weather for the period 1972-73 through 1979-80. This sample was stratified by month but not by time of day. Only those observations which contained reports of precipitation were used; the frequencies are therefore conditional on the event that precipitation occurs.

Before the REEP screening procedure was used, the BLPT, 850 T, 850 WBT, 10-8.5 Th, 10-5 Th and 8.5-5 Th predictors were transformed by the standardized procedure [(5)]. For this purpose, we determined the 50% values and spread constants for each station and each of the forementioned predictors using the developmental sample in a manner similar to that described in Section 2a. Throughout the remainder of this paper, predictors transformed according to (5) will be referred to as "standardized" predictors.

The joint predictors in Table 2 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 derived in a manner similar to that used in the OPER system, except that more data were available for their development. Graphs were constructed using the developmental sample 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 was 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 for reasons explained previously. Before constructing the graphs for EXP, all predictors were standardized.

TABLE 2. The potential predictors used in the development of EXP forecast equations.

Acronym	Definition
<i>Model output predictors</i>	
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
10-8.5 Th	1000-850 mb thickness
10-5 Th	1000-500 mb thickness
8.5-5 Th	850-500 mb thickness
<i>Model output joint predictors</i>	
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
<i>Observed and miscellaneous predictors</i>	
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

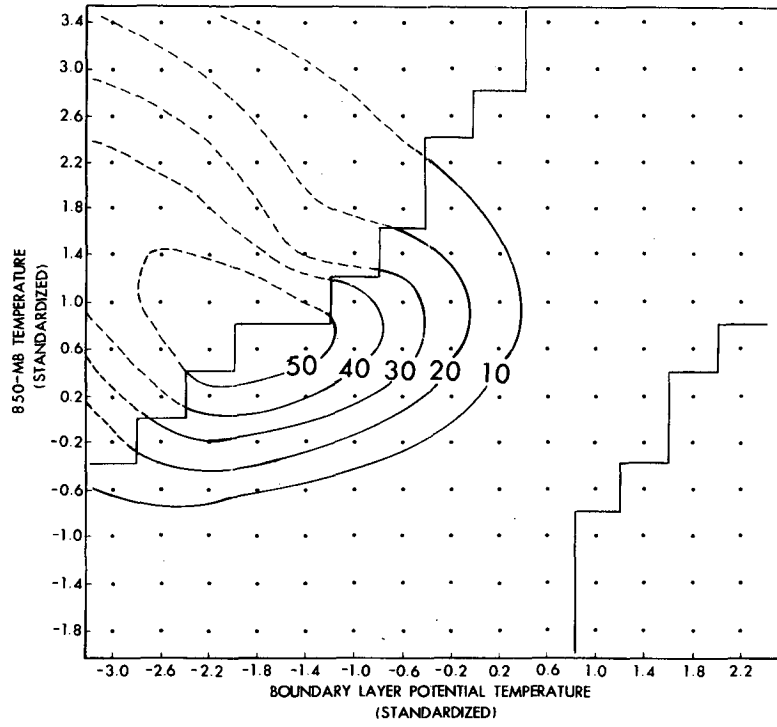


FIG. 4. The empirical probability of ZR (%), given that precipitation occurs, as a function of 850 T and BLPT. The predictors are forecast values from the LFM model and are standardized (see text for further explanation). The stepped lines separate areas with few or no data from areas with sufficient data.

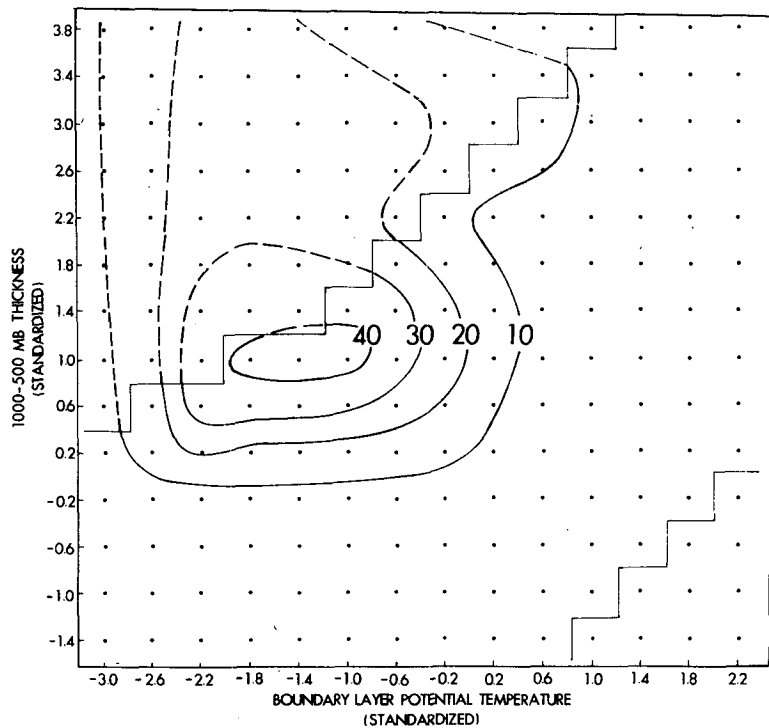


FIG. 5. As in Fig. 4, except that the 10-5 Th and BLPT are used as predictors.

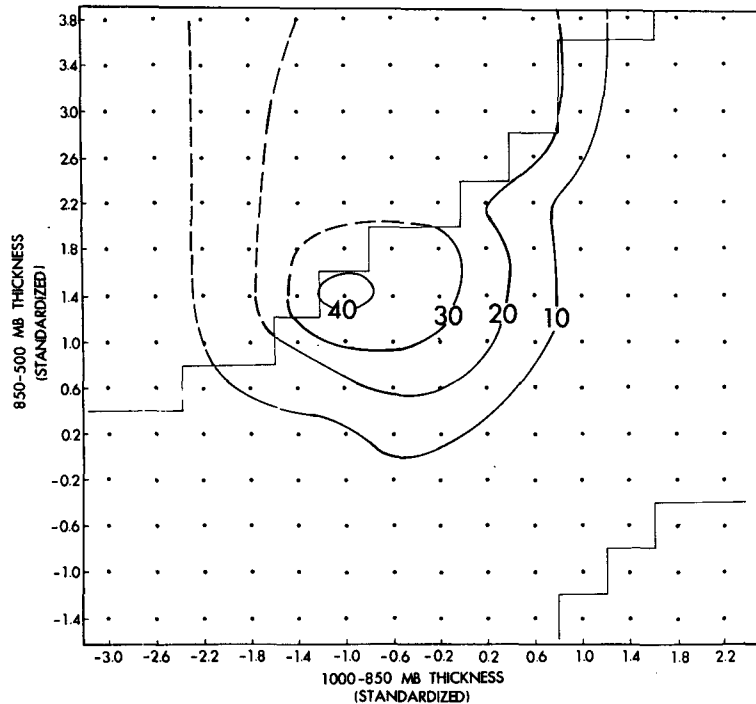


FIG. 6. As in Fig. 4, except that the 8.5–5 Th and 10–8.5 Th are used as predictors.

Figs. 4, 5 and 6 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. 7 and 8 indicate 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. 1–4 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. 4 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 freezing rain.

b. Development of regionalized equations

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 projec-

tions from both the 0000 and 1200 GMT cycle times with data combined from 229 conterminous U.S. 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

$$\text{Rel. Prob. Bias} = \frac{\overline{P(\text{SNOW})} - RF(\text{SNOW})}{RF(\text{SNOW})}, \quad (6)$$

where $\overline{P(\text{SNOW})}$ is the average probability of SNOW forecast for each station and $RF(\text{SNOW})$ 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

$$\text{Cat. Bias} = \frac{\text{FCST}(\text{SNOW})}{\text{OBS}(\text{SNOW})}, \quad (7)$$

where FCST (SNOW) is the number of forecast

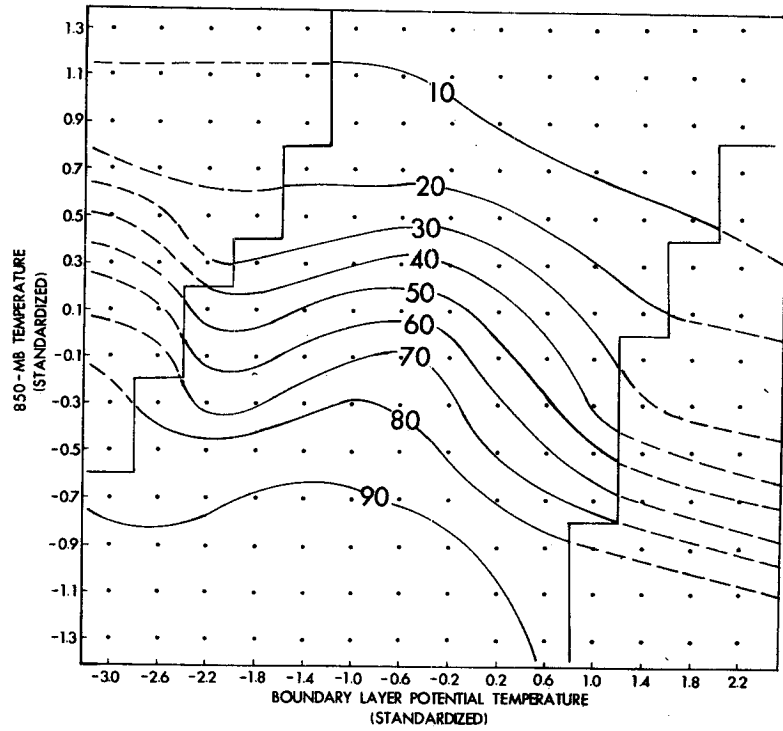


FIG. 7. As in Fig. 4, except that the empirical probability of the SNOW category is shown.

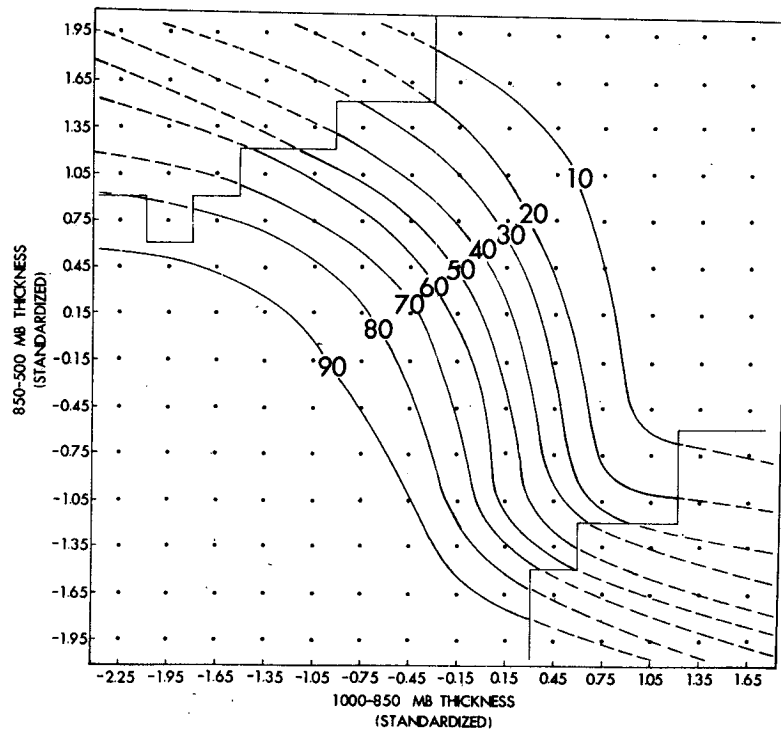


FIG. 8. As in Fig. 4, except that the empirical probability of the SNOW category is shown as a function of the 8.5-5 Th and 10-8.5 Th.

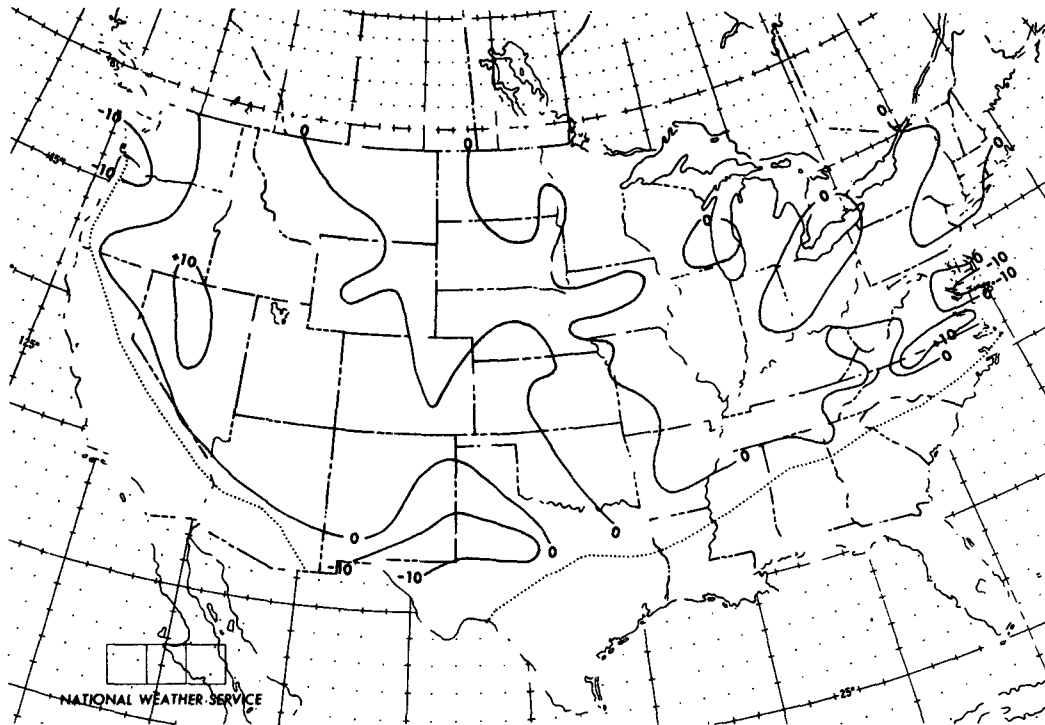


FIG. 9. The relative probability bias (%) (see text for definition) for SNOW averaged for 12 and 24 h forecasts from both the 0000 and 1200 GMT for the developmental sample, the winter seasons of 1972–73 through 1979–80. Values $> 0\%$ ($< 0\%$) indicate overforecasting (underforecasting) of the SNOW category. An analysis was not done for areas south and west of the dotted line because the infrequent occurrence of SNOW made analysis difficult.

SNOW events and OBS (SNOW) is the number of observed SNOW events. Figs. 9 and 11 show the relative probability bias and categorical bias, respectively, averaged for the 12 and 24 h projections from 0000 and 1200 GMT, while Figs. 10 and 12 show similar statistics averaged for the 36 and 48 h projections from the 0000 and 1200 GMT cycles. The analysis in Fig. 9 indicates that the relative probability bias for the 12–24 h period was generally small (within $\pm 10\%$). However, for the 36–48 h period (Fig. 10) the probability forecasts were generally too high over the Rocky Mountain area and portions of northern Texas, Oklahoma, Kansas and Nebraska, and the probabilities were generally too low over the Atlantic Coastal Plain. The analyses of the categorical bias in Figs. 11 and 12 show the SNOW event was generally overforecast over the Rocky Mountain area and portions of northern Texas, Oklahoma, Kansas and Nebraska, and was generally underforecast over the Atlantic Coastal Plain, the Pacific Northwest, and portions of the south. Overall, the areas of overforecasting and underforecasting were similar for both types of bias.

From these analyses, we determined the regions shown in Fig. 13 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.

After specifying the appropriate regions, we developed logit PoPT equations for each region by combining data from all stations within a region. The REEP screening procedure was used to determine the predictors to use in the logit equations. Table 3 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 selected; 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 important, especially for the longer range projections;

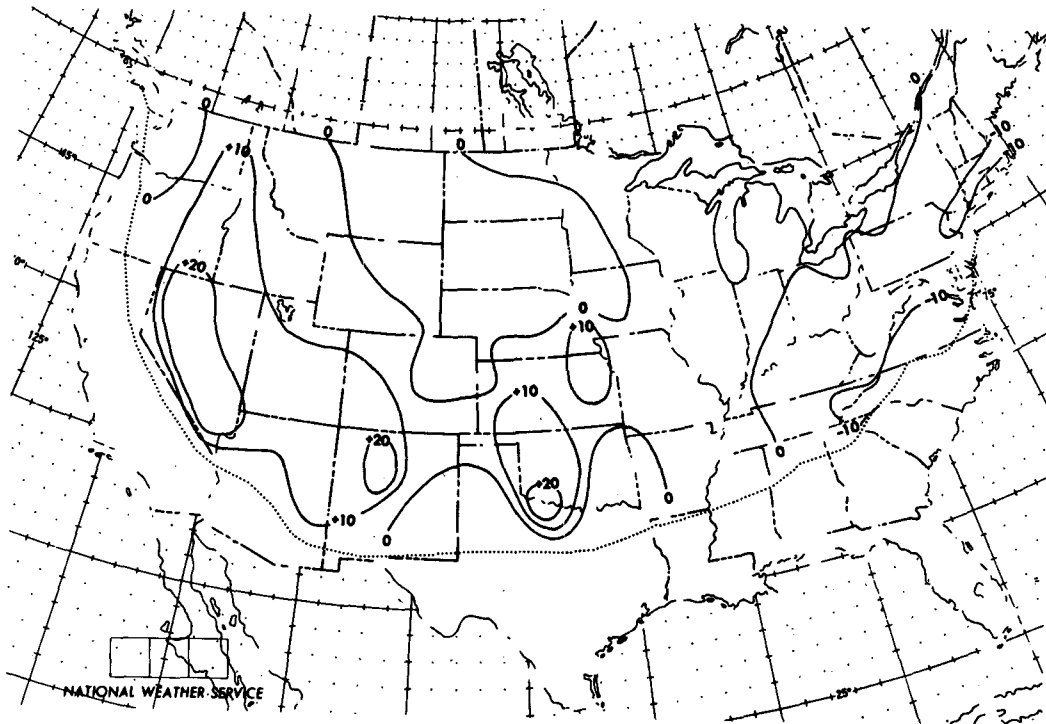


FIG. 10. As in Fig. 9, except that the relative probability bias was averaged from the 36 and 48 h projections from both the 0000 and 1200 GMT cycles.

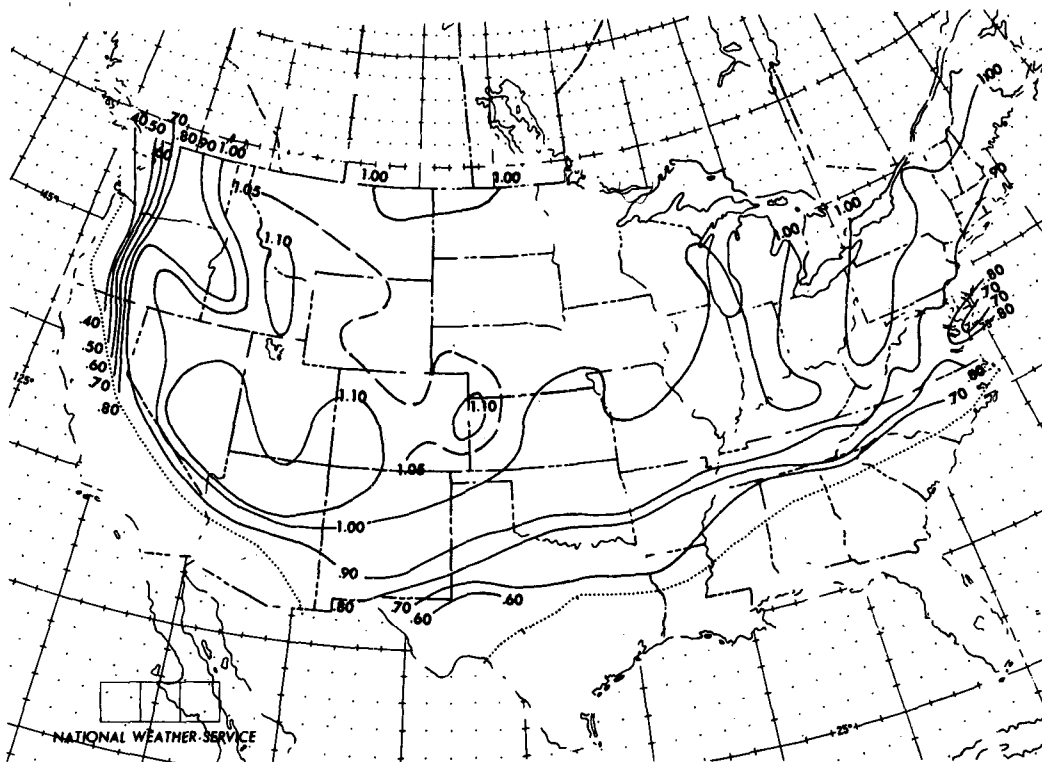


FIG. 11. As in Fig. 9, except that the categorical bias (see text for definition) is shown. Values > 1.00 (<1.00) indicate overforecasting (underforecasting) of the SNOW category.

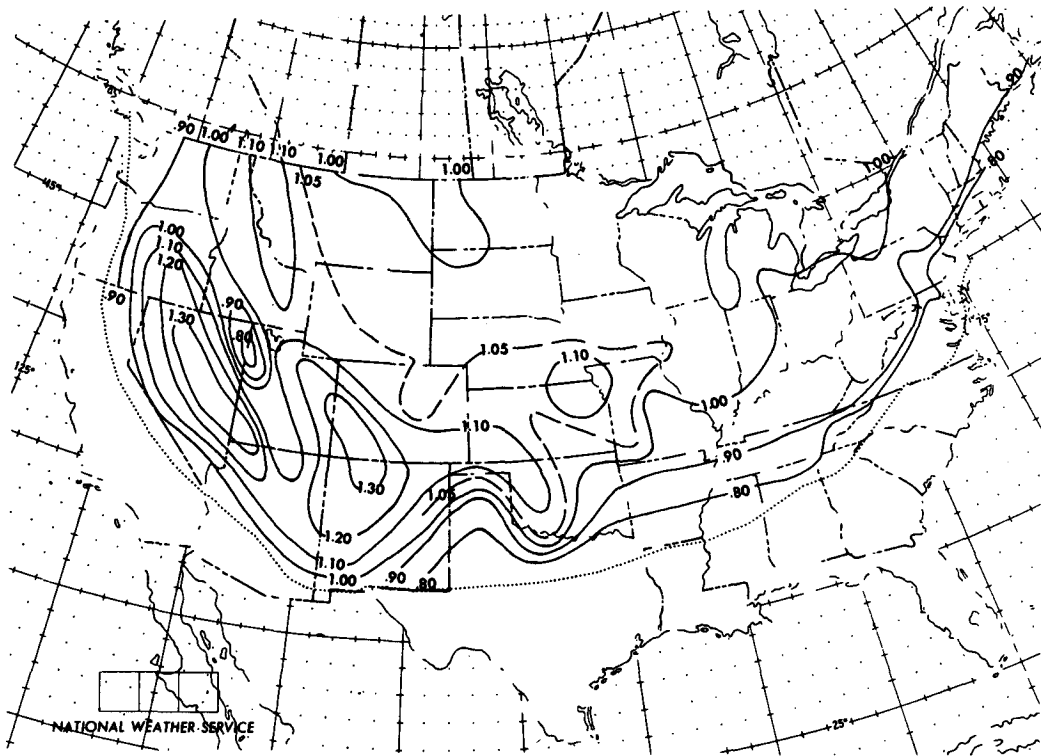


FIG. 12. As in Fig. 11, except that the categorical bias was averaged from the 36 and 48 h projections from both 0000 and 1200 GMT.

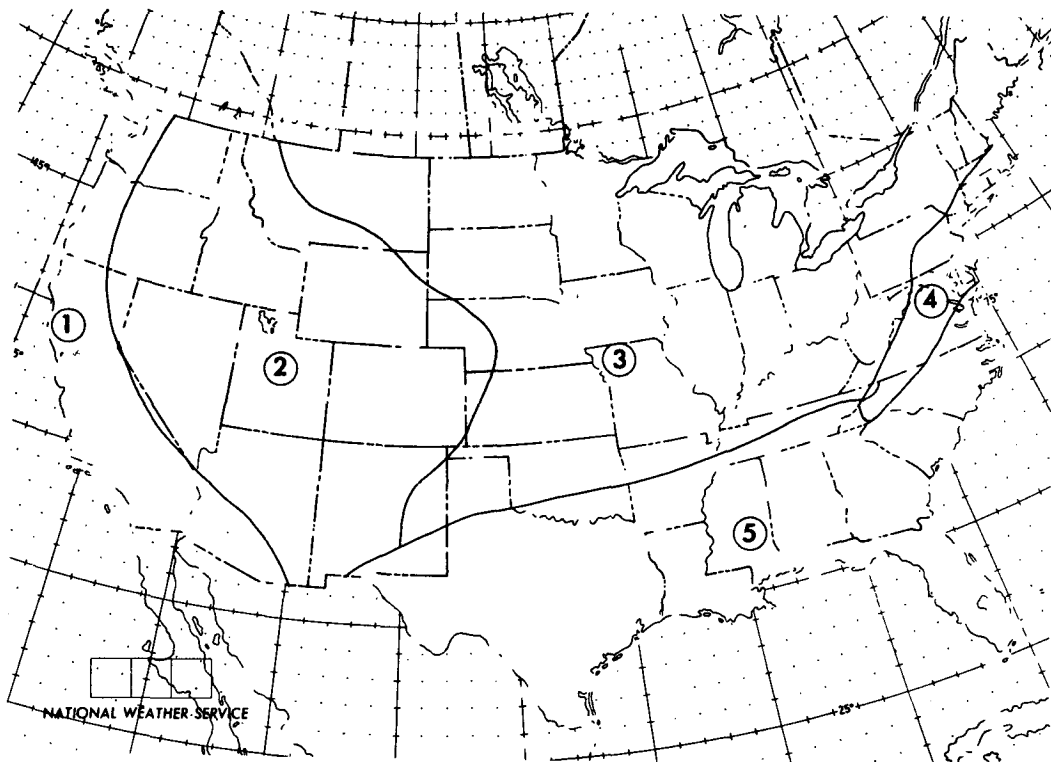


FIG. 13. The regions used in the development of the EXP probability forecast equation sets for PoPT.

TABLE 3. 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 2.

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

4) the FREQ ZR predictor, which was not included in the OPER system, also ranked relatively high, especially for the longer range projections; and 5) the OBS SFC T and OBS SFC Td became less important, of course, as the projection increased.

4. Verification of forecasts from the experimental equations

We did a comparative verification between the EXP equation set and the OPER system on independent data combined from 229 stations for the period September 1980 through March 1981. Both probabilistic and categorical precipitation type forecasts were verified.

Table 4 shows the Brier scores for probability forecasts for the RAIN, ZR, and SNOW categories and

the sum of those Brier scores (TOTAL) 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 Brier score of EXP over OPER and the number of cases are also shown. The results indicate the following: 1) in terms of the TOTAL, 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 h; 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 h (13.0%) and 24 h (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 4.

In addition to comparing probability forecasts from EXP and OPER equations, we also verified categorical forecasts of precipitation type. The verification scores included the categorical bias [(7)], threat score, post-agreement and prefigurance.¹ To 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)

¹ The threat score $[A/(B + C - A)]$, 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.

TABLE 4. Brier 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 Brier scores for each precipitation type category and the total Brier score (TOTAL) are shown. The percent improvement in Brier 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	0.036	0.020	0.040	0.096	0.049	0.013	0.050	0.112	0.053	0.024	0.055	0.132	0.067	0.014	0.068	0.149
OPER	0.037	0.023	0.041	0.101	0.048	0.014	0.051	0.113	0.057	0.024	0.058	0.139	0.069	0.014	0.070	0.153
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

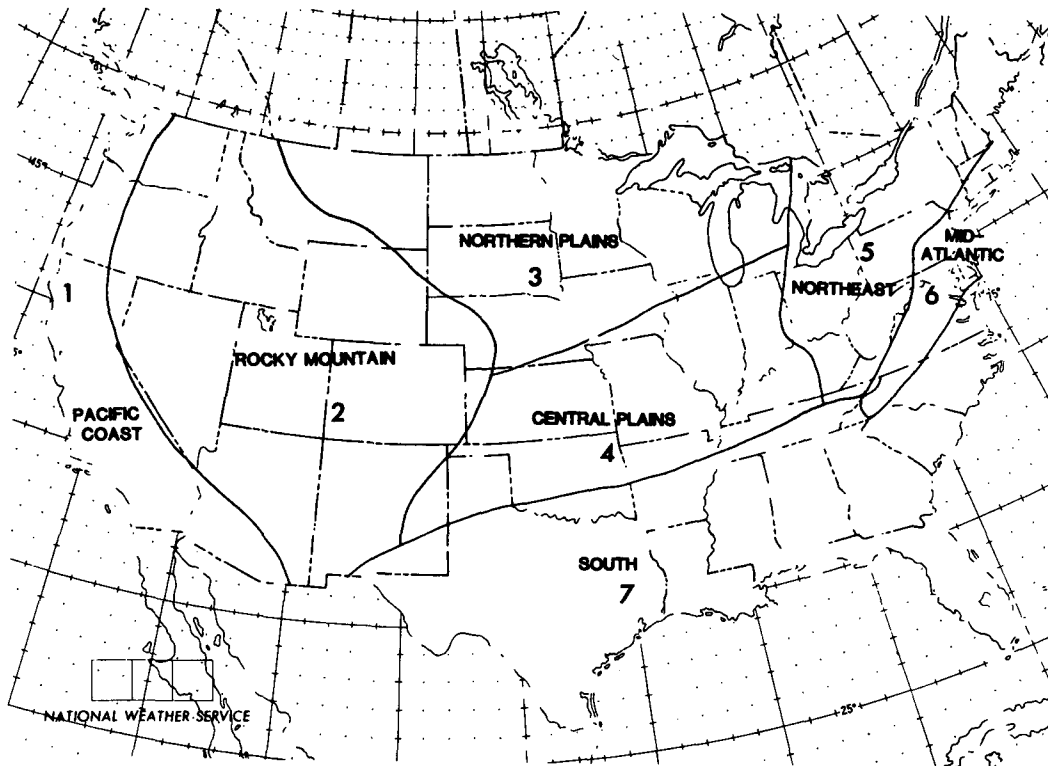


FIG. 14. The regions used for the development of threshold probabilities for the ZR and SNOW categories for the EXP logit forecast equations. The threshold probabilities are used to transform probability forecasts into categorical forecasts.

for further details]. For the purpose of deriving threshold probabilities, we divided region 3 in Fig. 13 into three new regions; the resulting regions are shown in Fig. 14. This was done mainly for convenience in using computer programs to develop the thresholds. To divide the old region 3, we used the small differences in the categorical bias (see Figs. 11 and 12).

Table 5 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;

TABLE 5. 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 (~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
<i>12 h Forecast</i>												
EXP	1.00	0.94	1.01	0.91	0.29	0.88	0.95	0.47	0.93	0.95	0.44	0.94
OPER	0.99	0.79	1.03	0.91	0.24	0.88	0.96	0.43	0.92	0.94	0.34	0.95
Percent improvement	—	—	—	0.0	+20.1	+0.0	-1.0	+9.3	+1.1	+1.1	+32.4	-1.1
<i>36 h Forecast</i>												
EXP	1.00	0.61	1.03	0.86	0.12	0.83	0.93	0.29	0.90	0.93	0.17	0.92
OPER	0.95	0.48	1.09	0.84	0.11	0.82	0.94	0.31	0.87	0.89	0.15	0.94
Percent improvement	—	—	—	+2.4	+9.1	+1.2	-1.1	-6.5	+3.4	+4.5	+13.3	-2.1

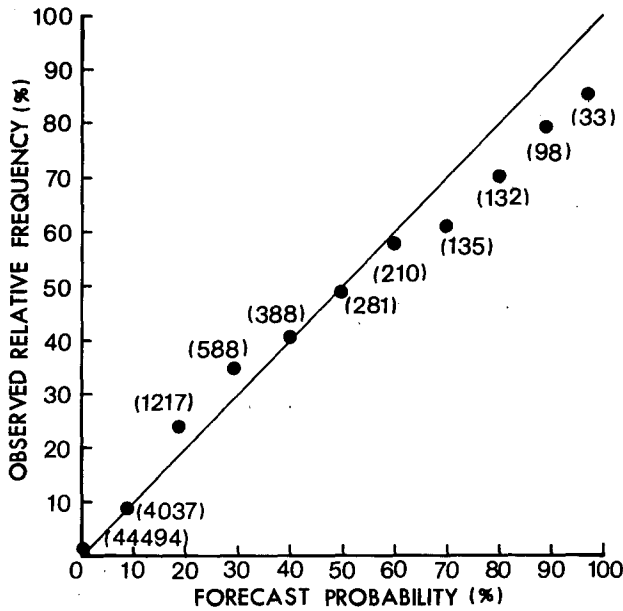


FIG. 15. The reliability of ZR probability forecasts from the new operational system for the 12 h projection from 0000 GMT. The sample consists of developmental data combined from 229 stations. The number of cases represented by each dot is shown in parentheses. The line denotes perfect reliability.

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 Brier scores in Table 4 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 h.

Note also in Table 5 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 90–95% of the 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 h, but EXP tended to underforecast ZR at 36 h. Also, for EXP, the post-agreement was only ~50% at 12 h and decreased to ~30% at 36 h.

5. Development of new operational PoPT forecast equations

Based on the verification results shown in Section 4, we decided to derive new operational PoPT forecast equations incorporating the features associated with the EXP equations. For this system, we developed equations for 6, 12, 18, 24, 30, 36, 42, 48, 54 and 60 h projections from both 0000 and 1200 GMT. The dependent and independent data samples used

in the experiments were combined so that, for the 6–24 h projections, nine winter seasons (1972–73 through 1980–81) were available and, for the 30–60 h projections, about 5½ winter seasons (February 1976 through 1980–81) were available. The new system will be implemented in the fall of 1982, and, twice daily, the PoPT forecasts will be sent to ~230 conterminous U.S. stations over the National Weather Service's Automation of Field Operations and Services (AFOS) network and the National Digital Facsimile (DIFAX) network.

To develop the new operational equations, we used the same potential predictors (Table 2), the same regions (Fig. 13), and the same predictand categories as were used to develop the EXP equations. The REEP screening procedure was used to determine the predictors to include in the logit equations for each projection. Also, in order to transform the probability forecasts into categorical precipitation type forecasts, we derived threshold probabilities for ZR and SNOW for each projection and for each of the seven regions shown in Fig. 14 in a manner similar to that used for the EXP equations. In addition to "primary" sets of PoPT equations, which contain surface observations valid at 0300 or 1500 GMT as predictors, we also developed "backup" equations which don't include surface observations.

We verified the primary and backup PoPT equations for several projections for the new developmental data sample. One desirable characteristic of probability forecasts is reliability; for all of the ZR forecasts of 20%, for instance, the relative frequency of the ZR event should be as close to 20% as possible. Figs. 15 and 16 show the reliability of the ZR prob-

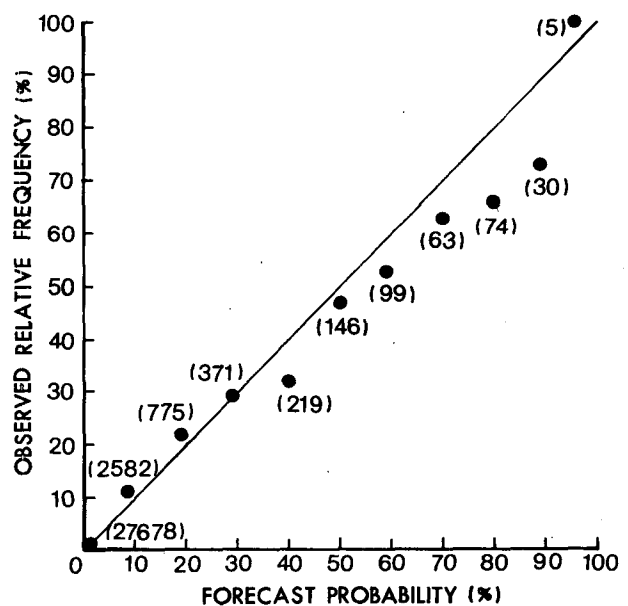


FIG. 16. As in Fig. 15, except that the reliability is for the 36 h projection.

ability forecasts from the operational PoPT primary 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. 15) 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. 16) for probabilities > 35%. A similar result for the old operational 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.

6. Summary

In an effort to improve the operational MOS PoPT system, we developed an experimental set of PoPT forecast equations. As compared to the operational system, the experimental equations were developed with a larger data sample and included improved predictors. Also, we devised a better method to transform the LFM model output predictors so that data could be combined from different stations when developing the forecast equations.

We did a comparative verification on independent data between the experimental and operational PoPT equations for both probabilistic and categorical precipitation type forecasts. The verification results indicate that the experimental system was better overall than the operational system especially for 12–24 h freezing rain forecasts. Therefore, we derived new PoPT forecast equations for operational use incorporating the features associated with the experimental

system. The new system was implemented within the National Weather Service in the fall of 1982.

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.

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 Conf. 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.
- Brownlee, K. A., 1967: *Statistical Theory and Methodology in Science and Engineering*. Wiley, 42–48.
- Cox, W. R., 1970: *Analysis of Binary Data*. Methuen, 25–26.
- 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. [NTIS PB-279-419].
- 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 probabilities. Tech. Rep. No. 1, Contract CWB-10704, The Travelers Research Center, Inc., 153 pp. [NTIS AD 602037].
- Newell, J. E., and D. G. Deaven, 1981: The LFM-II model—1980. NOAA Tech. Memo. NWS NMC-66, NOAA, U.S. Department of Commerce, 20 pp. [NOAA S/T 18-121].