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

TDL Office Note 77-8

AUTOMATED PREDICTION OF PROBABILITY OF PRECIPITATION (PoP) IN ALASKA--SUMMER SEASON

David B. Gilhousen

May 1977

AUTOMATED PREDICTION OF PROBABILITY OF PRECIPITATION (POP) IN ALASKA--SUMMER SEASON

David B. Gilhousen

1. INTRODUCTION

Since January 1972, the Techniques Development Laboratory has been providing automated PoP forecasts for 267 stations in the conterminous United States. These forecasts, made twice daily, are produced from equations made with the model output statistics (MOS) technique (Glahn and Lowry, 1972). One Region of the National Weather Service (NWS) has documented recent improvements in the PoP forecasts going to the public through judicious use of this product (MacDonald, 1977). In this paper, we describe the development and testing of PoP equations for the 14 Alaskan stations listed in Table 1. These 14 stations have a wide variety of local topographic influences and exhibit different seasonal climates. To account for this, we developed single station equations based on a 3-month summer season.

In particular, our development and testing involved several steps. First, we developed single station equations using multiple linear regression for several forecast projections. Secondly, we compared these forecasts to NWS subjective "local" forecasts for three Alaskan stations. Thirdly, we developed equations using a multivariate logit model (Brelsford and Jones, 1967). Then we compared forecasts produced by the logit model with those produced by linear regression. Finally, we developed our operational equations for all forecast projections using the logit model since this model produced better forecasts.

2. DEVELOPMENT OF REGRESSION EQUATIONS

Using the MOS approach, we generated six test sets of prediction equations for the 0000 GMT run of the primitive equation (PE) model (Shuman and Hovermale, 1968). Each test set consists of fourteen separate equations for each of the Alaskan stations. The predictand is the occurrence of measurable precipitation (.01 in or greater) in either a 12- or 6-hr forecast period. Two test sets produced forecasts for two 12-hr periods (called first period ending at 24 hours, and second period ending at 36 hours). Four test sets produced forecasts for four 6-hr periods ending at 18, 24, 30, and 36 hours after 0000 GMT. We derived the PoP equations for a 12-hr period simultaneously with equations for the two 6-hr periods within the 12-hr period. With this procedure, all three equations at any given station have the same predictors and more consistent[1] forecasts are likely.

---

[1] The PoP forecasts are inconsistent when either one of the 6-hr PoPs is greater than the 12-hr PoP, or the sum of the 6-hr PoPs is less than the 12-hr PoP.

Table 2 shows the potential predictors we screened. The developmental data sample consisted of the periods June through August of the years 1971 through 1975. This 3-month "summer" season was defined by the Alaska Region and is now being used in all our MOS development for Alaska. In addition to the PE model output, we screened surface observations (available 6 hours after the PE model run time) at the forecast site for all projections of 24 hours or less. Backup equations which don't use surface observations were also derived for these projections.

Table 3 is a summary of the predictors selected most frequently by the screening process for the first and second period equations. In addition to the predominant humidity predictors, wind components are particularly important in Alaska. Also, most of the predictors selected have valid times that are at or beyond the end of the forecast interval. This shows the slowness of the PE model.

3. TESTING

We carried out a verification experiment in order to determine how our automated forecasts compare with those prepared at Weather Service Forecast Offices (WSFO's) in Alaska. In particular, we verified objective forecasts based on MOS and subjective NWS local forecasts for Juneau, Fairbanks, and Anchorage made during June, July, and August of 1976. The objective forecasts were produced from regression equations developed on the five summer seasons of 1971 through 1975. These forecasts were produced solely for verification purposes, so they were not available as guidance to the field forecasters.

Table 4 shows the verification results in terms of the Brier score[1] for three stations. The objective forecasts scored better than the subjective ones for all stations for both forecast intervals. This is surprising, since forecasters can use surface data which are three hours later than those input to our MOS forecasts and only three hours earlier than the start of the first period. A possible explanation may be that local topographic differences between the city and the airport observing sites significantly affect the chances of rainfall. For the public forecast, the forecaster would obviously transmit his assessment of the probability for the urban area. He may be sending us urban forecasts which we are verifying with airport rainfall.

4. TESTING THE LOGIT MODEL

We decided to test forecasts from a multivariate logit model against those produced by our multiple regression model. A similar comparison for forecasts of conditional probability of frozen precipitation (PoF) showed the logit forecasts to be slightly better (Bocchieri and Glahn, 1976). A one-preductor logit model will attempt to fit a symmetric S-shaped curve to the data. In our PoF work, we found this curve to fit

---

[1]The Brier score is defined to be one-half the score proposed by Brier (1950).

the data well in a plot of 1000-500 mb thickness versus relative frequency of snow (Glahn and Bocchieri, 1975). Figure 1 shows a logit curve that fits the occurrence of measurable precipitation to PE mean relative humidity forecasts for Juneau, Alaska. In examining these curves for all Alaskan stations, we found nothing to discourage us from using the logit model.

A multivariate logit model fits a surface in multidimensional space. The mathematical form of the model can be expressed as:

$$P(Y) = \frac{1}{1 + \exp(a+b_1X_1+b_2X_2+\ldots b_NX_N)}$$

where Y is our binary predictand, P(Y) is the probability of precipitation, the X's are continuous predictors, and the a's and b's are constants to be determined. The computer program we use determines the maximum likelihood estimates of the a and b coefficients at the fourteen individual stations. Our program does have several limitations. It doesn't have a screening option; up to 10 independent variables can be included in the program, but they must be selected subjectively. Therefore, we gave our logit program the first 10 predictors selected by the screening regression program. We recognize that these 10 predictors may not be the best 10 for the logit model.

Table 5 shows our comparative verification between the forecasts made from our logit equations and forecasts made from multiple regression equations. Logit forecasts produced 3 to 4% lower Brier scores than regression forecasts when all data were pooled. Based on this result, we concluded that we should use the logit model to produce operational PoP equations, the predictors being the 10 best selected by regression.

5. OPERATIONAL SYSTEM

The operational system provides 12-hr PoP forecasts for periods ending at 24, 36, 48, and 60 hours along with 6-hr PoP forecasts for periods ending at 18, 24, 30, and 36 hours after model run time. There are separate equations for each station, forecast period, and model run time. In developing most of these equations, we screened the predictors whose forecast times were 48 hours or less in Table 2. The development data were for the 3-month summers of 1971 to 1976. However, for the 0000 GMT model run, we developed 12-hr PoP forecasts for the 36-48 hr third period and the 48-60 hr fourth period using only summer 1973 through 1976 data. This enabled us to screen PE fields beyond 48 hours, plus those fields marked by an asterisk in Table 2. Due to model slowness, 48-60 hr precipitation occurrence is better correlated with 60- and 72-hr forecast fields than with 48-hr fields. The 1200 GMT PE model produced forecasts only out to 48 hours. Therefore, third and fourth period forecasts from the 0000 GMT model run should be better than those from the 1200 GMT run.

Figure 2 shows the locations of the 14 Alaskan stations. We plotted their relative frequency of measurable precipitation and the range of the forecasts on the dependent data for 0000 GMT fourth period forecasts. The range shows the highest and lowest PoP forecasts that the MOS equations gave in the developmental sample. The wider the range of the forecasts, the better the quality and utility of the product. Notice that at Annette and Nome the full range of forecast values from 0 to 100% were produced. At King Salmon and Cold Bay more limited ranges were produced, yet the frequency of precipitation at those stations was higher than that reported at Nome. We can perhaps explain this by observing that the mean summer storm track for Arctic cyclones lies in the northern part of the Bering Sea close to Nome (Read and Kunkel, 1960). The high frequency of precipitation in the Aleutians is due to local effects such as the advection of warm air over colder water and the subsequent interaction with terrain. Obviously, rainfall from moving baroclinic disturbances can be better modeled by the PE than rainfall due to orographic or advective causes.

Our MOS approach attempted to account for these local effects. The screening process produced equations for Cold Bay and King Salmon that included many of the wind component and temperature advection terms and the coefficients give heavy weight to these terms. Yet despite this, a rather limited range of forecast values were produced. This effect was more apparent in the fourth period equations, but was observed in most of the equations for Cold Bay and King Salmon.

6. FUTURE WORK

We will continue to use this same approach to develop Alaskan PoP equations for the fall, winter, and spring seasons. We will also start work on forecasting the conditional probability of frozen precipitation for Alaska; our target date for implementing these PoF equations is September 1, 1977.

ACKNOWLEDGMENTS

Special thanks are extended to Evelyn Boston of TDL for assistance in verifying the objective and subjective PoP forecasts. We are also grateful to the Technical Procedures Branch of the Office of Meteorology and Oceanography for providing us with the local forecasts.

REFERENCES

Bocchieri, J. R., 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.

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

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.

MacDonald, A. E., 1977: The MAN/MOS program. NOAA Technical Memorandum NWS WR-112, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, 21 pp.

Reed, R. J., and B. A. Kunkel, 1960: The Arctic circulation in summer. J. Meteor., 17, 489-506.

Shuman, F. G., and J. B. Hovermale, 1968: An operational six-layer primitive equation model. J. Appl. Meteor., 7, 525-547.

Table 1. Fourteen stations used to develop an automated PoP forecasting system for Alaska along with their location identifiers.

ANC	Anchorage	JNU	Juneau
ANN	Annett	AKN	King Salmon
BRW	Barrow	OTZ	Kotzebue
BTI	Barter Island	MCG	McGrath
BET	Bethel	OME	Nome
CDB	Cold Bay	SNP	St. Paul Island
FAI	Fairbanks	YAK	Yakutat

Table 2. Potential predictors available to the screening regression program for the summer season. Certain predictors have been space smoothed by 5, 9, or 25 points to eliminate small scale noise. Predictors at 60 and 72 hours are available for the 0000 GMT run only. An asterisk indicates that the field is only available starting June 1, 1973 (see text for details).

Field	Smoothing[1] (Points)	Time (Hours from model run time)	Form[2]
(a) PE Model Output			
Mean Rel. Humidity	1,5,9,25	12,18,24,30,36,42,48,60,72	B,C
Boundary Layer Humidity	1,5,9,25	12,24,36,48,60,72	B,C
Precipitation Amount	1,5,9	12,24,36,48,60,72	B
850-mb Height	5,9	12,24,36,48,60,72	B
850-mb Temperature	1	24,36,48,60,72	B,C
850-mb Temperature Advection	5	24,36,48,60,72	B,C
500-mb Minus 850-mb Temp. Difference	1,5	24,36*,48*,60*,72*	B
500-mb Geostrophic Vorticity Advection	5	12,24,36,48,60,72	B,C
850-mb Vertical Velocity	5	12,24,48*,60*,72*	B,C
650-mb Vertical Velocity	5	12,24	B,C
Boundary Layer Moist. Convergence	1,5	12,24,36,48,60,72	B,C
Boundary Layer U,V Wind	5	12,24,36,48,60,72	B,C
850-mb U,V Wind	1,5	24,36,48,60,72	B,C
500-mb U,V Wind	1,5	24,36,48,60,72	B,C
Mean Rel. Humidity[3], $\Delta\tau 12$	5	48	B,C
(b) Other Predictors			
Sine and Cosine Day of Year	--	--	C
Observed Ceiling	--	06	B
Observed Sky Cover	--	06	B
Observed Weather	--	06	B
Observed Dew Point	--	06	B

1 A one in this column means no smoothing.
2 B = binary form, C = continuous form.
3 12-hr trend ending at time shown.

Table 3. PE forecast and 0600 GMT observed predictors listed according to the total number of times they are used in the Alaskan summer season PoP equations for the 0000 GMT forecast cycle. The number following the predictor indicates its projection. (Note: geo.=geostrophic, conv.=convergence, adv.=advection, S5=five-point smoothing to eliminate small scale noise.)

Rank	Forecast Period	
	First (12-24 hr)	Second (24-36 hr)
1	Mean Rel. Humidity S5 30	Mean Rel. Humidity S5 36
2	Bound. Layer U S5 24	Precip. Amount S5 36
3	Mean Rel. Humidity S5 24	500-mb. Geo. V 36
4	Observed Sky Cover	Bound. Layer U S5 36
5	500-mb Geo. Vorticity Adv. S5 12	Mean Rel. Humidity S5 42
6	Precip. Amount S5 24	Bound. Layer V S5 36
7	Observed Weather	850-mb. Temp. Adv. S5 36
8	Bound. Layer Moist. Conv. S5 12	Mean Rel. Humidity S5 30
9	Bound. Layer Rel. Humidity S5 24	Bound. Layer Rel. Humidity S9 36
10	Bound. Layer Moist. Conv. S5 24	500-mb. Geo. U 36
11	500-mb. Geo. U S5 24	850-mb. Geo. V 36
12	Bound. Layer V S5 24	850-mb. Temperature 36

Table 4. Brier scores for 0000 GMT TDL objective and NWS subjective local PoP forecasts for three stations in Alaska during June through August of 1976.

Station	First Period			Second Period		
	No. of Cases	Brier Scores Objective	Local	No. of Cases	Brier Scores Objective	Local
Anchorage	78	.089	.093	78	.079	.081
Juneau	78	.083	.132	78	.092	.114
Fairbanks	70	.077	.091	70	.070	.081
Overall	226	.083	.106	226	.081	.093

Table 5. Comparative verification between forecasts from logit equations and forecasts from multiple regression equations for various station groupings. Independent data are for 0000 GMT forecasts for June through August of 1976. Panhandle stations include Annette, Juneau, and Yakutat. SW coast stations include Cold Bay, King Salmon, and St. Paul Island. Interior stations include Fairbanks, McGrath, and Bethel.

Stations	First Period			Second Period		
	No. of Cases	Brier Scores Regression	Logit	No. of Cases	Brier Scores Regression	Logit
Panhandle	261	.087	.082	261	.089	.090
SW Coast	191	.090	.092	232	.086	.082
Interior	261	.070	.068	261	.083	.075
All 14 Stations	1145	.074	.071	1189	.077	.074

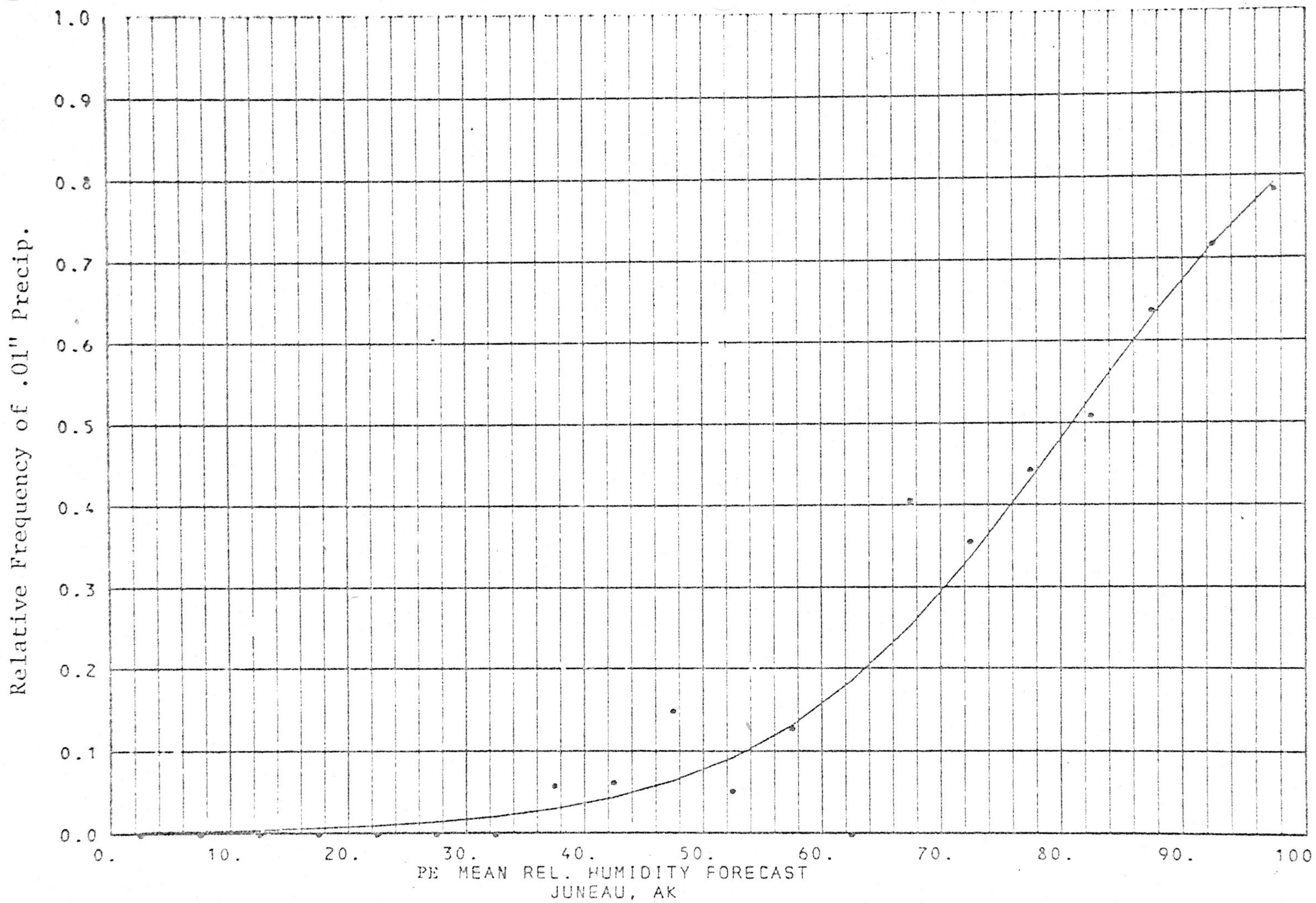

Figure 1. The logit curve describes the 12-24 hr occurrence of measurable precipitation as a function of 24-hr PE mean relative humidity forecasts at Juneau, Alaska. The dots are the relative frequency of measurable precipitation in 5% intervals. The data sample consists of the 3-month summer seasons of 1971 through 1975.

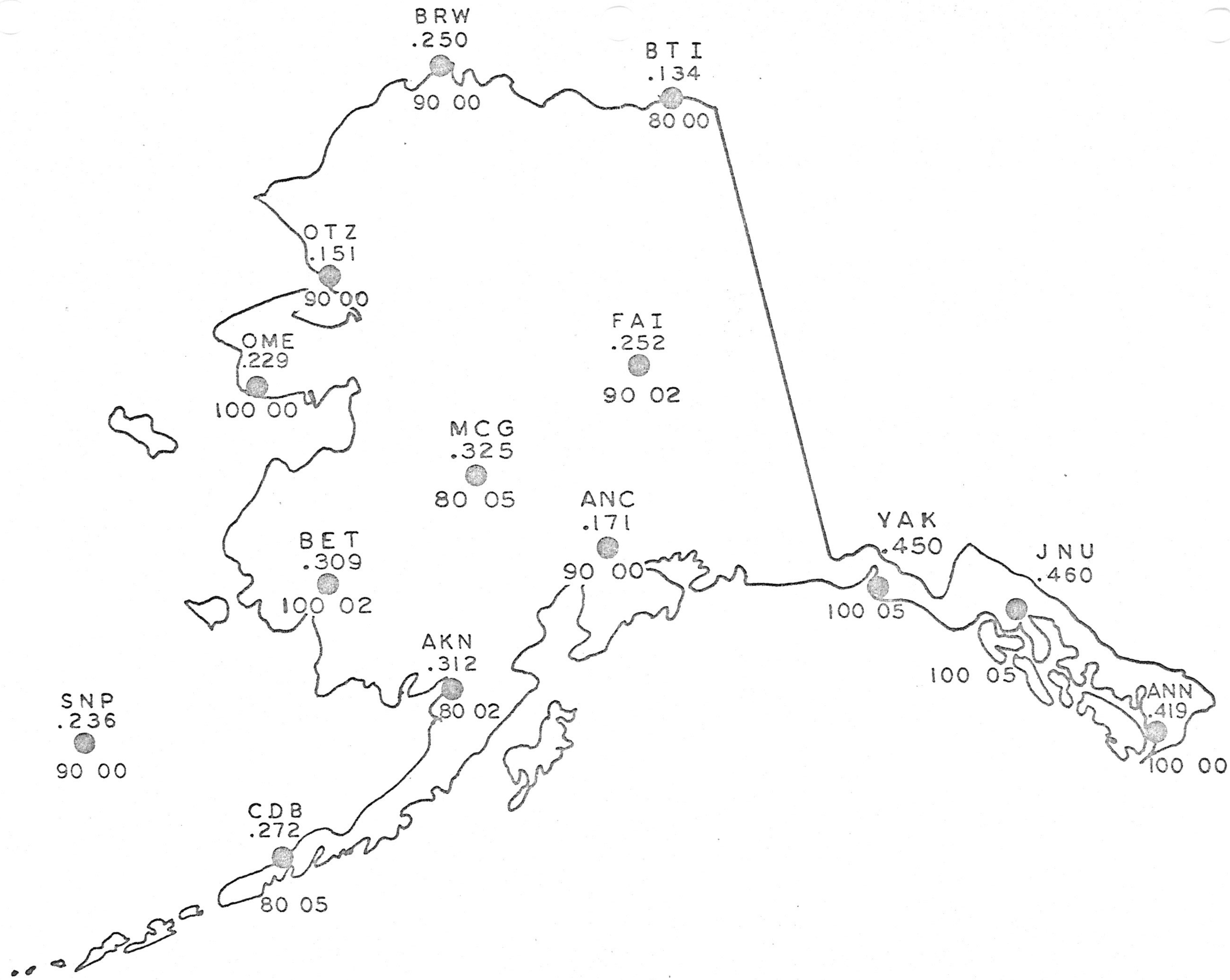

Figure 2. Relative frequency of precipitation and range of 0000 GMT PoP forecasts for the fourth period. Frequencies are plotted above the station location; the maximum and minimum PoPs that the MOS equations gave in the developmental sample are plotted below.