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

TDL OFFICE NOTE 85-3

DEVELOPMENT OF AN IMPROVED AUTOMATED SYSTEM FOR FORECASTING THE PROBABILITY OF LIQUID PRECIPITATION TYPE

George J. Maglaras and Edward L. Chiang

January 1985

DEVELOPMENT OF AN IMPROVED AUTOMATED SYSTEM FOR FORECASTING THE PROBABILITY OF LIQUID PRECIPITATION TYPE

George J. Maglaras and Edward L. Chiang

1. INTRODUCTION

The Techniques Development Laboratory (TDL) has been producing computer worded forecasts (CWF's) (Glahn, 1978; Bermowitz et al., 1980; Bermowitz and Miller, 1984; National Weather Service, 1983a, 1983b) for several years. Part of the input used by the CWF is the conditional probability of liquid precipitation type (PoLPT). When liquid precipitation is likely, the CWF uses these forecasts to decide whether the liquid precipitation event should be described as "drizzle," "rain," or "showers." A new system for forecasting PoLPT for 267 conterminous United States stations became operational in September 1978 (Gilhousen, 1979). The probability forecasts are conditional because the system assumes liquid precipitation will occur; i.e., only liquid precipitation cases were included in the developmental sample. To develop the forecast equations for this system, the Model Output Statistics (MOS) technique (Glahn and Lowry, 1972) was used with output from the Limited-area Fine Mesh (LFM) model (Gerrity, 1977; National Weather Service, 1977). The Regression Estimation of Event Probability (REEP) statistical model (Miller, 1964) was used to relate the LFM model output to observations of liquid precipitation type.

In an effort to improve the PoLPT forecast equations, we developed an experimental set of equations for both the cool (October-March) and warm (April-September) seasons, called EXP. We compared the forecasts from these experimental equations to those from the operational system (OPER) and found the forecasts from EXP to be superior. As a result, we decided to develop a new set of PoLPT forecast equations for operational use.

The new set of PoLPT forecast equations, called NEW, differs from the operational system in three ways. First, NEW was developed with more than six (seven) cool (warm) seasons of LFM data for all projections; OPER had been developed with less LFM data and the sample varied from projection to projection as shown in Table 1. Second, NEW was developed with data from nearly 500 stations; OPER used data from only 233 stations. Finally, NEW does not use predictors based on LFM sigma-layer output, thus making NEW less sensitive to model changes; OPER uses sigma-layer predictors.

2. DEVELOPMENT AND TESTING OF THE EXP EQUATIONS

A. Predictor and Predictand Data

The developmental sample for the cool season included more than five seasons (1976-77 through 1981-82); the warm season sample included six seasons (1977 through 1982) of LFM model output. Surface observations were from the Techniques Development Laboratory's (TDL's) archive of hourly surface reports for approximately 500 stations in the conterminous United States. Table 2 shows the potential predictor variables we used to develop the cool and warm season equations. These included model output variables valid for 12-, 18-, 24-, 30-, 36-, 42-, and 48-h projections. The model output variables for the

12-h projection were unsmoothed; for the 18-, 24-, and 30-h projections the variables were five-point space-smoothed; and for the 36-, 42-, and 48-h projections, the variables were nine-point space-smoothed. For the experiments we developed equations with data from only the 0000 GMT cycle so all observed predictor variables were valid at 0300 GMT.

B. Regions

Grouping stations into regions is desirable because only liquid precipitation cases are included in the developmental sample and in some locations during certain times of the year precipitation is considered a rare event. Grouping stations increases the sample size used to develop equations. In order to determine regional boundaries, we calculated the observed relative frequencies of showers and of drizzle for each station for the 18-h projection from 0000 and 1200 GMT. Stations were grouped into regions if they had similar relative frequencies of showers and of drizzle. The 20 (15) regions we selected for the cool (warm) season are shown in Fig. 1 (Fig. 2).

C. Cool Season EXP Equation Development

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 give estimates of the probabilities of occurrence for a given set of binary predictands. In PoLPT, liquid precipitation is divided into three binary predictands: drizzle, rain, and showers. 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 which liquid precipitation type was reported. The potential predictors were either in binary or continuous form. The use of binary predictors helps to account for non-linear relationships between the predictand and predictor. A good description of the screening procedure can be found in Glahn and Lowry (1972).

In order to determine if LFM sigma-layer predictors were needed, we developed two sets of EXP equations for the cool season. For each equation set, we combined data from all stations within a region (see Fig. 1) and developed separate sets of equations for the 18-, 30-, and 42-h projections from 0000 GMT. The first EXP equation set included forecasts of LFM sigma-layer fields as predictors; the second set did not. However, the second set did contain several new derived predictors which we considered good substitutions for the sigma-layer predictors. The first set did not include these new predictors (see Table 2 for an explanation of which predictors were screened in each equation set). Also, in order to determine if observed predictors were important, we developed two more equations sets, both with sigma-layer predictors, for the 18-h projection from 0000 GMT. One set included observed weather elements as predictors; the other did not. For these two equation sets, the REEP screening procedure was continued until 12 terms had been selected. For the other EXP equation sets, the REEP screening process was continued as long as the addition of a new term added at least 0.75% to the reduction of variance for at least one of the three predictands or until a maximum of 12 terms had been selected.

The most important predictors for the equations with sigma-layer predictors were the three sigma-layer relative humidity fields, the mean relative humidity, differences in relative humidity between sigma layers, differences in temperature between various levels, and temperature fields including the boundary-layer potential temperature. For the equations without sigma-layer predictors, the most important predictors were the same except that sigma-layer relative humidity fields and differences in relative humidity between sigma-layers were replaced by dew-point depression fields and differences in dew-point depression between various levels. Also, we removed the boundary-layer potential temperature as a potential predictor because it is a sigma-layer predictor.

D. Verification of the Cool Season EXP Equations

For the cool season EXP equation sets, we performed two comparative verifications on independent data combined from 218 stations for the period October 1982 through March 1983. In each verification, we calculated the P-score (Brier, 1950) over all three categories for the 18-, 30-, and 42-h projections from 0000 GMT. We transformed the probability forecasts to categorical forecasts by choosing the category with the highest probability. From these categorical forecasts, we computed the percent correct and Heidke skill score (Panofsky and Brier, 1968).

For the first verification, we compared EXP forecasts from equations which included sigma-layer variables as predictors (EXPWSL) to those from the OPER system. We also compared EXPWSL equations with and without observed weather elements as predictors to OPER in this verification, but for the 18-h projection only.

Table 3 shows the scores for EXPWSL and OPER. The results indicate that EXPWSL was better than OPER in terms of all three scores (especially for the Heidke skill score), and for all three projections (especially for the 42-h projection). Table 4 shows the scores for EXPWSL equations with and without observed predictors, and OPER. The results indicate that, in terms of percent improvement over OPER, EXPWSL equations with observed predictors were only slightly better than EXPWSL equations without observed predictors.

For the second verification, we compared the EXPWSL forecasts to those from EXP equations with no sigma-layer predictors (EXPNSL). Our concern was that future models may not include sigma-layers similar to those used in the present LFM model.

Table 5 shows the scores for EXPWSL and EXPNSL. The results indicate that EXPNSL is better than, just as good as, and worse than EXPWSL in terms of P-score, percent correct, and Heidke skill score, respectively, for the three projections combined. Overall, the quality of the forecasts produced from the two equation sets was not greatly different.

E. Warm Season EXP Equation Development and Testing

For the warm season EXP equations, we combined data from all stations within a region (see Fig. 2) and developed separate sets of equations for the 18-, 30-, and 42-h projections from 0000 GMT. Based on the cool season experiments, the warm season EXP equations were developed without any sigma-layer or observed predictors. The REEP screening process for the EXP equations was continued as long as the addition of a new term added at least 0.5% to the reduction of variance for at least one of the three predictands or until a maximum of 12 terms had been selected.

The most important predictors for the warm season EXP equations, similar to the cool season, were temperature and dew-point depression fields at various levels, differences in temperature and dew-point depression between various levels, and the mean relative humidity.

The warm season verification involved a comparison of EXP and OPER forecasts on independent data combined from 218 stations for the period April 1983 through September 1983. We verified forecasts for the same projections using the same scores that we had used for the cool season.

Table 6 shows the scores for EXP and OPER and the percent improvement of EXP over OPER for each of the scores. The results indicate that EXP was better than OPER in terms of all three scores and for all three projections. The results are similar to those obtained for the cool season, except that the magnitude of the improvement is less for the warm season equations.

3. DEVELOPMENT OF NEW POLPT FORECAST EQUATIONS

The results of the verification tests presented in Section 2 showed the forecasts from cool and warm season EXP equation sets were more accurate than those from OPER and that sigma-layer and observed predictors did not add appreciably to the accuracy of the EXP equations. Hence, we decided to develop new operational equations which incorporate the features of EXP without any sigma-layer or observed predictors.

We derived sets of new forecast equations for the cool and warm seasons for the 18-, 30-, 42-, and 54-h projections after 0000 and 1200 GMT. To develop the new PoLPT system we combined the dependent and independent data samples used in the experiments discussed previously, thus creating developmental samples of more than six seasons for the cool season and seven seasons for the warm season. For NEW, we used the same regions (see Fig. 1 and Fig. 2) and the same potential predictor list (see Table 2) as for EXP. The REEP screening process was continued as long as the addition of a new term added at least .75% (.50%) to the reduction of variance for the cool (warm) season for at least one of the three predictands or until a maximum of 12 terms had been selected. The most important predictors for NEW were the same as for the EXP equation sets.

4. SUMMARY

A system for forecasting PoLPT for the conterminous United States became operational in September 1978. That system, here called OPER, was developed with the MOS technique and output from the LFM model. In an effort to improve OPER, we developed an experimental set of PoLPT forecast equations called EXP, also based on LFM model output.

Based on the results of several experiments, we determined that the cool and warm season EXP equation sets which did not include LFM sigma-layer fields or observed weather elements as predictors were better than OPER for both seasons. Therefore, we derived new operational PoLPT forecast equations, incorporating the features associated with the EXP equations. Separate sets of equations were derived for both forecast cycles (0000 and 1200 GMT) for the cool (October-March) and warm (April-September) seasons. These new equations were implemented in December 1984. The PoLPT forecasts from the new operational system are used as input to the CWF program. The probability forecasts are converted to categorical forecasts by the CWF by choosing the predictand category (drizzle, rain, or showers) with the highest probability. An added constraint used for the drizzle category is that the probability must also be greater than 40%. This is done in order to reduce the number of drizzle forecasts in the CWF. Although the PoLPT system does not overforecast the occurrence of drizzle, forecasts of drizzle are not often made by local forecasters, and we are attempting to make the appearance of drizzle in the CWF forecast realistic in terms of local forecaster use. As such, these forecasts are not being disseminated to NWS forecasters in the United States directly, but indirectly through the CWF.

5. ACKNOWLEDGMENTS

We are grateful to Belinda Howard for typing the manuscript, and to the many other members of the Techniques Development Laboratory who contribute to the development and maintenance of the MOS system.

REFERENCES

Bermowitz, R. J., M. M. Heffernan, H. R. Glahn, 1980: Computer worded forecasts: An update. Preprints Eighth Conference on Weather Forecasting and Analysis, Denver, Amer. Meteor. Soc., 453-456.

_____, and R. L. Miller, 1984: A field evaluation and some recent changes in the computer worded forecast. Preprints Tenth Conference on Weather Forecasting and Analysis, Clearwater Beach, Amer. Meteor. Soc., 92-97.

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, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, 68 pp.

Gilhousen, D. B., 1979: Improved prediction of liquid precipitation type. TDL Office Note 79-8, National Weather Service, NOAA, U.S. Department of Commerce, 10 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.

_____, 1978: Computer worded public weather forecasts. NOAA Technical Memorandum NWS TDL-67, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, 24 pp.

Miller, R. G., 1964: Regression estimation of event probabilities. Technical Report No. 1, Contract CWB-10704, The Travelers Research Center, Inc., Hartford, Conn., 153 pp.

National Weather Service, 1977: High resolution LFM (LFM II). NWS Technical Procedures Bulletin No. 206, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, 6 pp.

_____, 1983a: Computer worded forecasts. NWS Technical Procedures Bulletin No. 336, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, 17 pp.

_____, 1983b: Computer worded forecasts (APPENDIX). NWS Technical Procedures Bulletin No. 337, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, 8 pp.

Panofsky, H. A. and G. W. Brier, 1968: Some applications of statistics to meteorology. Pennsylvania State University, University Park, Pa., 224 pp.

Table 1. Number of seasons of data used to develop cool and warm season OPER equations.

Projection	No. of Seasons of Data Cool Season	Warm Season
18	4	5
30	3	3
42	2	2
54	2	2

Table 2. The potential predictors included in the development of the EXP PoLPT forecast equations. One (two) asterisks indicate predictors included in equations sets with (without) sigma-layer predictors. No asterisk indicates the predictor was included in all types of equation sets. (BL = boundary layer).

Definition		Levels
a. Model Output Predictors		
East-west wind component		850 mb
North-south wind component		850 mb
Mean relative humidity		SFC-500 mb
BL relative humidity	*	--
Layer-one relative humidity	*	Top BL-700 mb
Layer-two relative humidity	*	700-490 mb
BL east-west wind component	*	--
BL north-south wind component	*	--
Temperature		1000 mb, 850 mb
Dew-point		1000 mb, 850 mb
Vertical Velocity		850 mb
Precipitable water		SFC-500 mb
Precipitation amount		--
b. Model Output Derived Predictors		
Geostrophic east-west wind component	**	1000 mb
Geostrophic north-south wind component	**	1000 mb
Temperature difference		850-1000 mb, 700-1000 mb, 700-850 mb, 500-850 mb
Dew-point depression	**	1000 mb, 850 mb, 700 mb, 500 mb
Dew-point depression differences	**	1000-850 mb, 1000-700 mb, 1000-500 mb, 850-700 mb, 850-500 mb
BL potential temperature	*	--
BL wind speed	*	--
K Index		--
Total-totals Index		--
c. Observed and Geoclimatic Predictors		
Observed weather		--
Observed east-west wind component		--
Observed north-south wind component		--
Observed temperature		--
Observed dew-point		--
Observed wind speed		--
Sine of the day of the year		--
Cosine of the day of the year		--
Station latitude		--
Station longitude		--

Table 3. P-scores for EXPWSL and OPER cool season equation sets for PoLPT forecasts for the 18-, 30-, and 42-h projections after 0000 GMT. The sample consisted of independent data combined from 218 stations for the period October 1982 through March 1983. The percent improvement of EXPWSL over OPER is also shown. The sample included an average of 3150 cases for each projection.

Projection (h)	Forecast System	Verification Scores		
		P-Score	Percent Correct	Skill Score
18	EXPWSL	.4465	67.7	.3411
	OPER	.4633	66.6	.2841
	% Improvement EXPWSL/OPER	3.6	1.7	20.1
30	EXPWSL	.4850	62.6	.2664
	OPER	.5012	62.1	.2252
	% Improvement EXPWSL/OPER	3.2	0.8	18.2
42	EXPWSL	.4857	64.0	.2296
	OPER	.5302	59.7	.1245
	% Improvement EXPWSL/OPER	8.4	7.2	84.4

Table 4. Same as Table 3 except EXPWSL equations with (OBS) and without (NOBS) observed predictors are compared to OPER for the 18-h projection only. The sample included an average of 3100 cases.

Projection (h)	Forecast System	Verification Scores		
		P-Score	Percent Correct	Skill Score
18	OBS	.4429	68.1	.3530
	OPER	.4618	66.8	.2871
	% Improvement OBS/OPER	4.1	1.9	23.0
18	NOBS	.4446	67.7	.3404
	OPER	.4633	66.6	.2841
	% Improvement NOBS/OPER	4.0	1.7	19.8

Table 5. Same as Table 3 except EXPWSL and EXPNSL are compared. The sample included an average of 3200 cases for each projection.

Projection (h)	Forecast System	Verification Scores		
		P-Score	Percent Correct	Skill Score
18	EXPWSL	.4464	67.6	.3389
	EXPNSL	.4417	68.0	.3256
	% Improvement EXPWSL/EXPNSL	-1.1	-0.6	3.9
30	EXPWSL	.4845	62.6	.2675
	EXPNSL	.4785	63.1	.2689
	% Improvement EXPWSL/EXPNSL	-1.2	-0.8	-1.3
42	EXPWSL	.4849	64.0	.2178
	EXPNSL	.4851	63.4	.1930
	% Improvement EXPWSL/EXPNSL	0	0.9	15.4

Table 6. Same as Table 3 except warm season EXP and OPER equations are compared. The sample consisted of independent data combined for the period April 1983 through September 1983. The sample included an average of 2100 cases for each projection.

Projection (h)	Forecast System	Verification Scores		
		P-Score	Percent Correct	Skill Score
18	EXP	.4616	64.8	.3620
	OPER	.4689	64.5	.3397
	% Improvement EXP/OPER	1.6	0.5	6.6
30	EXP	.4625	65.9	.3827
	OPER	.4794	63.4	.3343
	% Improvement EXP/OPER	3.5	3.9	14.5
42	EXP	.4907	61.3	.2913
	OPER	.5254	58.7	.2326
	% Improvement EXP/OPER	6.6	4.4	25.2

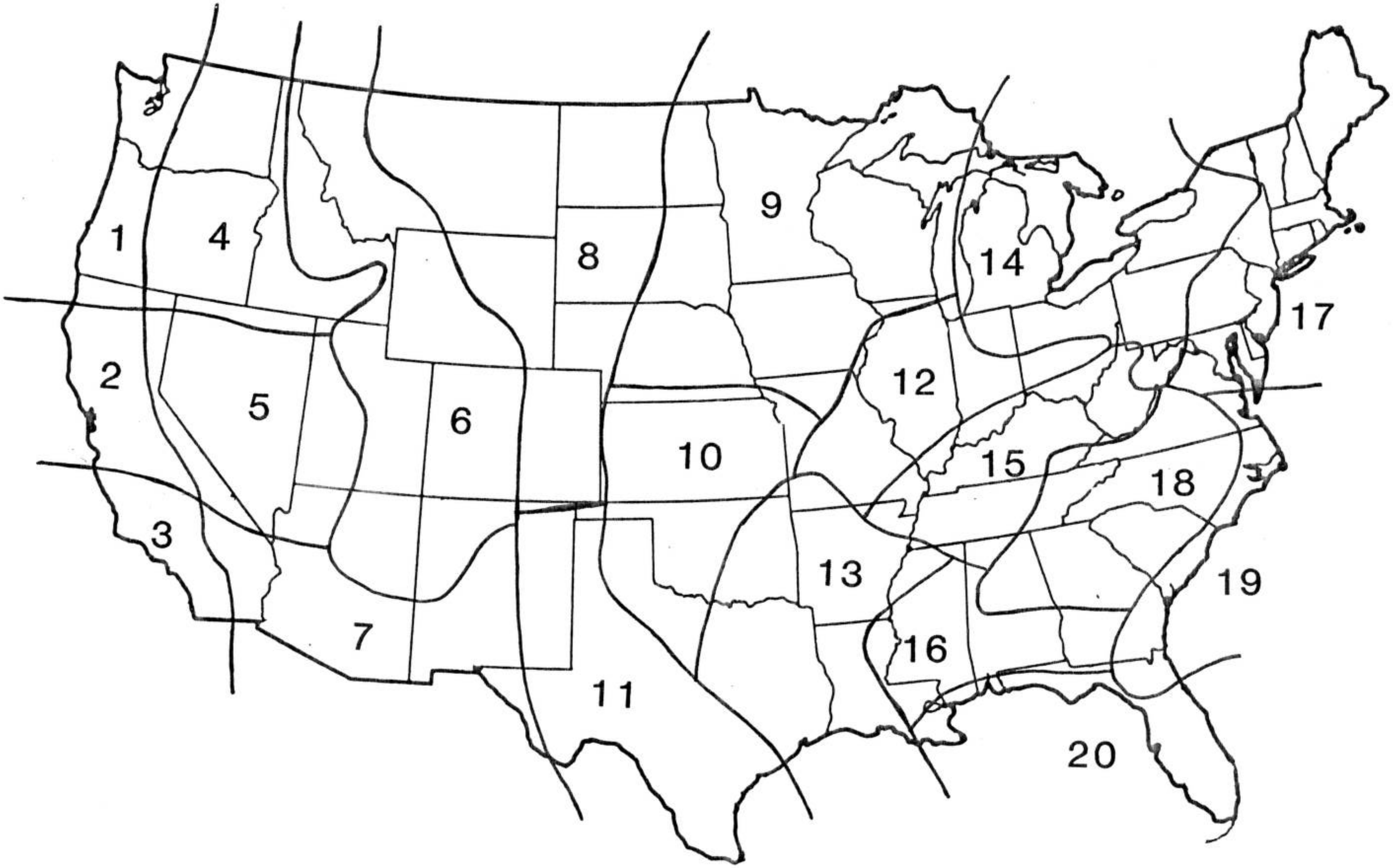

Figure 1. Regions used to develop PoLPT equations for the cool season.

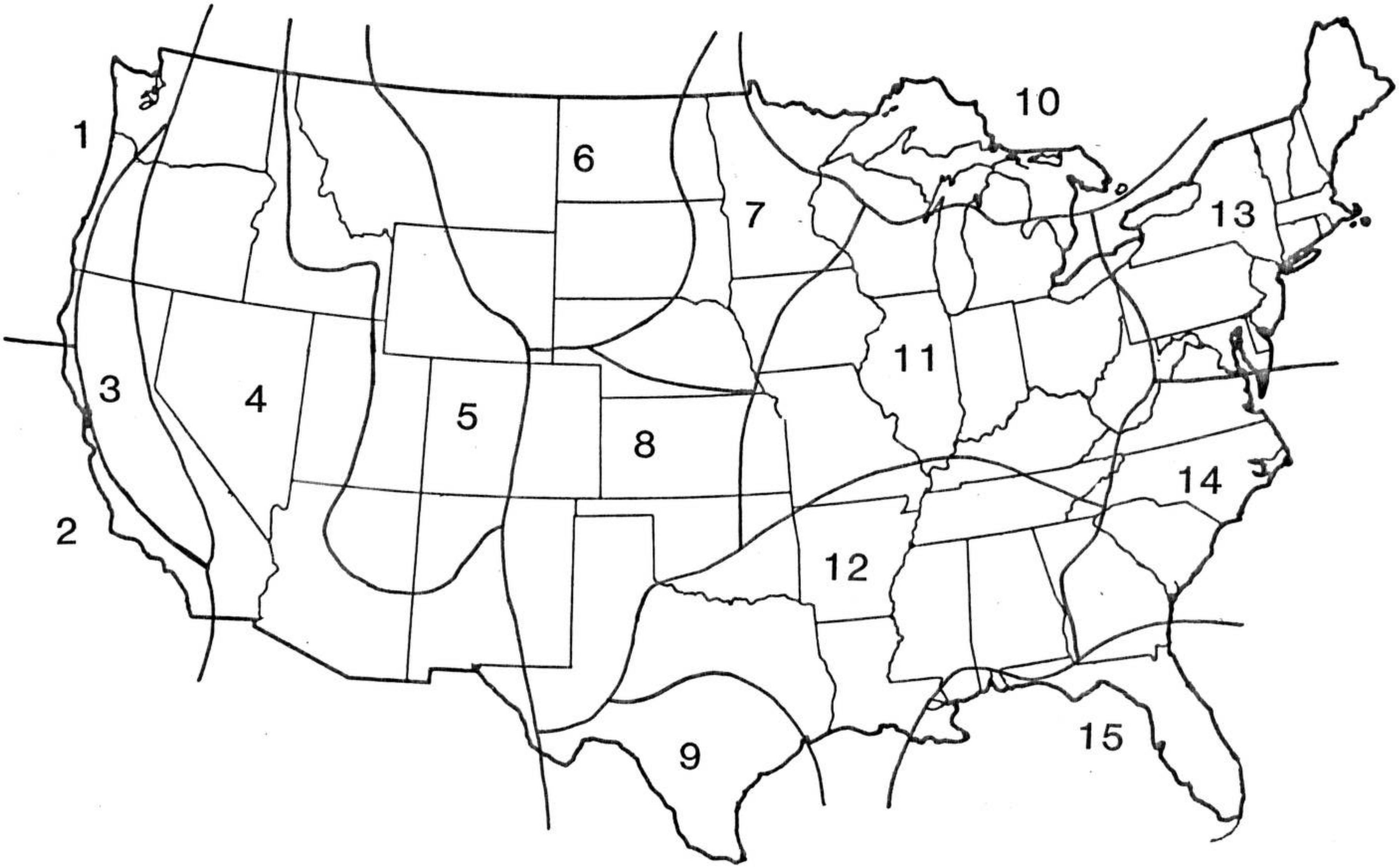

Figure 2. Same as Fig. 1 except for the warm season.