OBJECTIVE FORECASTS OF LOCAL WEATHER BY MEANS OF MODEL OUTPUT STATISTICS

BY

W. KLEIN

NATIONAL WEATHER SERVICE, NOAA

SILVER SPRING, MD.

U.S.A.

ABSTRACT

The method called Model Output Statistics (MOS) is a very effective technique for combining statistical and numerical weather prediction. MOS has been successfully applied by the Techniques Development Laboratory of the National Weather Service to prepare automated guidance forecasts of numerous weather elements on the synoptic scale in all parts of the United States. This talk will describe the use of MOS in forecasting weather for the public. To illustrate the method and its performance under operational conditions, sample forecast equations and teletype output will be presented. The utility of MOS will then be evaluated with the aid of comparative verification figures.

1. Introduction

For many years statistical weather prediction had a rather poor reputation because of its questionable physical basis. However, the past decade has witnessed a marked revival of interest in the subject, and it now enjoys a high level of respectability among weather forecasters in the United States.

I see two main reasons for the rejuvenation of statistical forecasting. First the use of digital electronic computers has made possible rapid processing of large quantities of data by sophisticated statistical techniques such as screening regression, discriminant analysis, and factor analysis. Second, the success of numerical weather prediction has furnished a reliable and skillful base upon which statistics can build.

Figure 1 illustrates two methods of combining statistical and numerical techniques. The first, called the perfect prog method, utilizes observed historical data to specify local weather elements from concurrent (or nearly concurrent) weighted combinations of meteorological parameters. To use the derived equations for making a forecast, we apply them to the output of numerical prognostic models which simulate the observed circulation, as shown by the dashed arrow. Although errors in the numerical prediction will inevitably produce corresponding errors in the statistical forecast, the latter will improve each time the former is improved. An advantage of this method is that stable forecasting relations can be derived for individual locations and seasons from a long period of record. A disadvantage is that it takes no account of errors and uncertainties in the numerical model. The perfect prog method has been applied many times since its initial use by the author (Klein et al., 1959), but in the

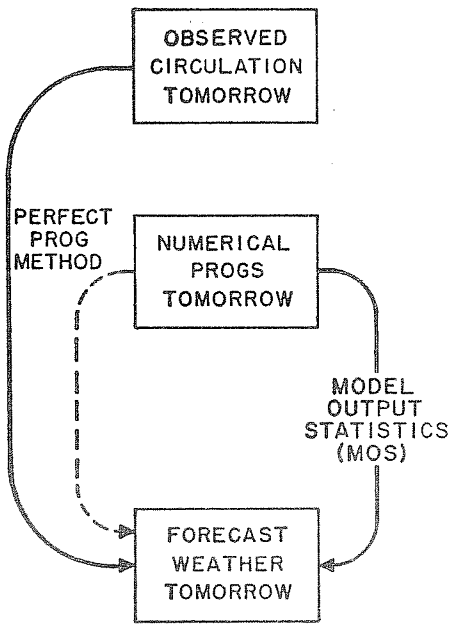

Figure 1. Two methods of combining numerical and statistical weather forecasting in schematic form.

past few years it has been gradually replaced by the second method. However, it is still useful for forecasting relatively rare events where local factors are very important. For example, it is being applied on an operational basis in the National Weather Service (NWS) to produce automated forecasts of the water level on Lake Erie (Richardson and Pore, 1969) and extratropical storm surges along the Atlantic coast (Pore et al., 1974).

The second statistical technique has been named Model Output Statistics (MOS) by Glahn and Lowry (1972) and developed in the Techniques Development Laboratory (TDL) of the Systems Development Office. Instead of a long period of observed data, the predictor sample in MOS usually consists of a relatively short period of prognostic data produced by numerical models. Thus, the MOS method archives the output from numerical models and matches it with observations of local weather. Forecast equations are then derived by using a variety of statistical techniques. In this way the bias and inaccuracy of the numerical model, as well as the local climatology, can be automatically built into the forecast system. Another characteristic of MOS is that it can include many predictors not readily available to the perfect prog method, such as vertical velocity, boundary layer wind and temperature, 3-dimensional air trajectories, etc. Because of these advantages, the MOS technique has, for most uses, proven to be more successful than the perfect prog method in recent years. TDL now applies MOS routinely to make automated forecasts of nearly every weather element in all parts of the United States except Hawaii (Klein and Glahn, 1974).

In this lecture I shall limit myself to public weather and shall explain how TDL applies MOS to forecast temperature, precipitation, thunderstorms, winds, and clouds. I shall also present comparative verification figures to demonstrate that MOS is about as skillful as experienced forecasters. MOS has also been successfully applied to marine, aviation, and agricultural weather, but these special services will not be discussed here. Definition of statistical terms is given in the Appendix.

2. The MOS system

In the MOS technique, observations of local weather are matched with prognostic data produced by numerical models. These data are used as potential predictors, together with station observations and climatological terms. Forecast equations are then derived by using a variety of statistical techniques. In this way the predictors are selected and weighted in accordance with the accuracy (or inaccuracy) of the numerical model instead of the true relations in the real atmosphere.

Most of the TDL MOS products have been based upon the output of the six-level baroclinic Primitive Equation (PE) model of Shuman and Hovermale (1968) and the 3-dimensional trajectory (TJ) model of Reap (1972). Systematic archiving of output from these models began in July 1969 and has continued to date. Later TDL added a third model; namely, the Limited Area Fine Mesh (LFM) model (Gerrity, 1977; Howcroft and Desmaris, 1971). Archiving of this model began on October 1, 1972.

Local surface weather reports are acquired monthly from the National Climatic Center in Asheville, N. C. for each of the 254 basic observing stations plotted in Fig. 2. Numerical model output at each of these stations is obtained by biquadratic interpolation from PE, TJ, and LFM model grids. The observations and numerical predictors are then matched on a station by station basis.

Although the numerical predictors are always located at the same point as the predictand weather element, they are not necessarily valid at the same time. Because the numerical models can be systematically slow or fast, predictors within ± 24 hr of the predictand time are also useful in certain cases.

Another procedure which has increased the utility of the numerical predictors is space smoothing. Averaging over 5, 9, or 25 grid points frequently removes spurious perturbations from "noisy"

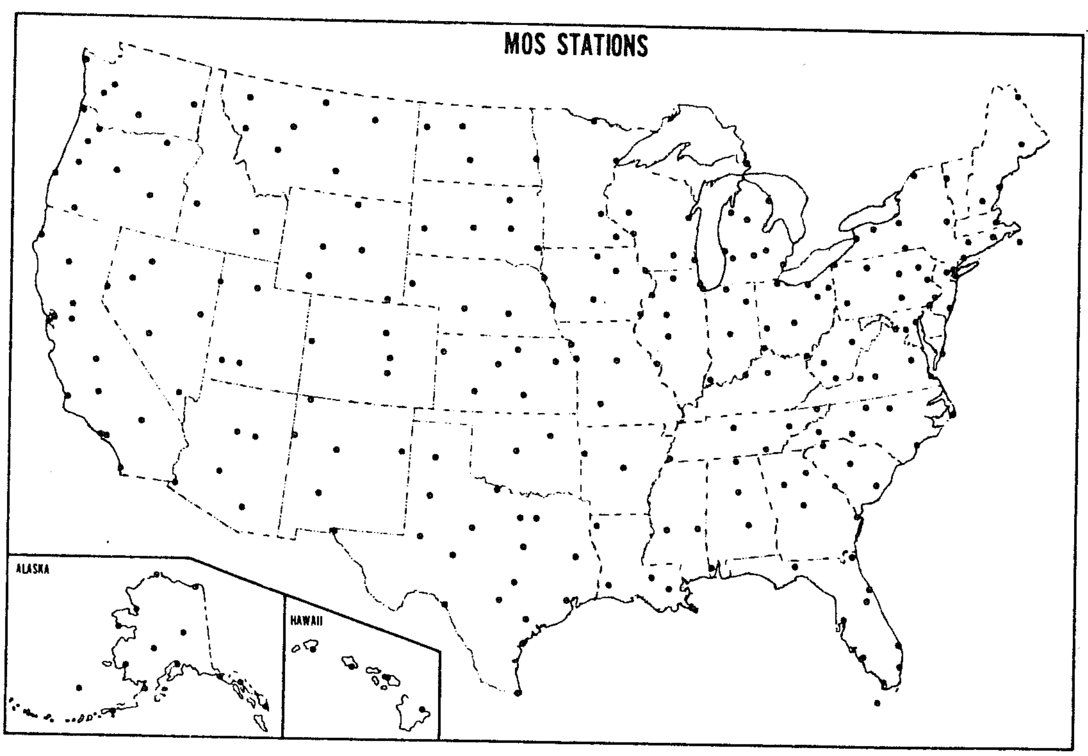

Figure 2. Locations of 254 stations in the 50 states for which routine weather observations are being collected for matching with numerical model output.

numerical output. Smoothing also introduces information from surrounding grid points to an otherwise local scheme. Considerable experimentation has indicated that smoothing of numerical model output should increase with increasing forecast projection, decreasing elevation of the predictors, and decreasing predictor scale.

In an attempt to account for the predominance of different weather regimes in winter and summer, the year is usually divided into two 6-month seasons: the cool months October through March, and the warm months April through September. The relatively short period of record precluded further stratification by time of year until recently when 3-month seasons proved successful in temperature forecasting (Hammons et al., 1976). To account for seasonal trend, the first two harmonics of the sine and cosine of the day of the year are normally included as potential predictors.

The MOS equations are applied twice daily on an operational basis at the National Meteorological Center (NMC) in Suitland, Md., immediately after each cycle of the numerical models (0000 and 1200 GMT). The resulting objective forecasts are then transmitted on facsimile and/or teletypewriter as guidance to NWS field forecasters. For each cycle, both "early" and "final" guidance forecasts are transmitted. The early guidance is based on the LFM model; the final guidance on the PE model. Surface observations reported at 0300 and 1500 GMT may also be used for early guidance, with 0600 and 1800 GMT reports used for final guidance. In addition, the TJ model is used as necessary.

3. Temperature

The application of MOS to develop forecast equations for maximum and minimum temperatures was described by Klein and Hammons (1975) and Hammons et al. (1976). The potential predictors were carefully selected from the output of PE and TJ models to include all available factors which might influence surface temperature such as height, thickness, temperature, wind, moisture, stability, vorticity, divergence, and vertical velocity at various levels and projections. For the first and second period forecasts, nine surface synoptic reports were included as possible predictors to give the latest observed conditions at the station. These reports were at 0600 and 1800 GMT, 6 hours after the initial time of the numerical models (0000 or 1200 GMT), but still early enough for operational use. The complete predictor list for the 0000 GMT cycle is given in Table 1.

Since temperature is a continuous nearly normally distributed variable, forecast equations were derived by a forward stepwise screening regression program (Miller, 1958). Separate equations were developed for each of 228 stations, four projections, two run times, and four seasons, for a total of 7296 multiple regression equations. All equations contained exactly 10 terms since previous research (Bocchieri and Glahn, 1972; Glahn and Lowry, 1972) indicated this is approximately the optimum number of predictors for continuous variables and for samples of this size.

A sample temperature equation used for final guidance is given in Table 2 for today's maximum at Las Vegas, Nev., during the three winter months of December, January, and February. The predictors are listed in the order of selection. As more predictors are added, they contribute irregularly diminishing increments to the reduction of variance (square of the multiple correlation coefficient). The

Table 1. Potential predictors of maximum and minimum surface temperature for MOS screening regression. Numbers indicate valid time of predictors in hours after 0000 GMT. Stars indicate the predictor was smoothed by 5 points (*), 9 points (**) or 25 points (***).

Predictor	Today Max	Tonight Min	Tomorrow Max	Tomorrow Night Min
		a) PE Model		
850-mb height	12,24	24,36	36,48	48,48*
500-mb height	12,24	24,36	36,48	36*,48,48*
1000-500-mb thickness	12,24	24,36	36,48	48,48*
1000-850 mb thickness	12,24	24,36	36,48	48,48*
850-500-mb thickness	12,24	24,36	36,48	48,48*
1000-mb temperature	12,24,24*,36*	24*,36,36*,48*	36*,48,48*,48**	48*,48**,48***
850-mb temperature	12,24,24*,36*	24*,36,36*,48*	36*,48,48*,48**	48*,48**,48***
700-mb temperature	24	24*	24**	-
Boundary layer potential temp	12,24,24*,36*	24*,36*,48*	36*,48*,48**	48*,48**,48***
Boundary layer U wind	12,24*	24*,36*	36*,48*	48*,48**,48***
Boundary layer V wind	12,24*	24*,36*	36*,48*	48*,48**,48***
Boundary layer wind speed	24	36	48	48*,48**
850-mb U wind	24	24*	24**	24***
850-mb V wind	24	24*	24**	24***
700-mb U wind	24	24*	24**	24***
700-mb V wind	24	24*	24**	24***
1000-mb relative vorticity	24*	36*	48**	48***
850-mb relative vorticity	24*	36*	48*	48**
500-mb relative vorticity	24*	36*	48*	48**,48***
850-mb vertical velocity	24	24*	24**	-
650-mb vertical velocity	24	24*	24**	-
Stability (1000-700-mb temp)	24	24*	24**	-
Stability (850-500-mb temp)	24	24*	24**	-
400-1000 mean rel hum	12*,24*,36*	24*,36*,48*	36**,48**	48**,48***
Precipitable water	18*,30*	30*,42*	42*,42**	42**,42***
Boundary layer wind divergence	24*	36*	48*	48**,48***
		b) Trajectory Model		
Surface temperature	24,24*	24*,24**	24*,24**	24**,24***
850-mb temperature	24,24*	24*,24**	24*,24**	24**,24***
700-mb temperature	24	24*	24**	24**,24***
Surface dew point	24,24*	24*,24**	24*,24**	24***
850-mb dew point	24*	24*	24**	24***
700-mb dew point	24	24*	24**	24***
700 mb-surface mean rel hum	24	24*	24**	24***
850-mb 12-hr net vert displ	24	24*	24**	24***
850-mb 24-hr net vert displ	24	24*	24**	24***
700-mb 12-hr net vert displ	24	24*	24**	24***
700-mb 24-hr net vert displ	24	24*	24**	24***
Surface 12-hr horiz conv	24,24*	24*,24**	24*,24**	24***
850-mb 12-hr horiz conv	24	24*	24**	24***
George's K index	24	24*	24**	24***
		c) Other Variables		
Sine day of year	00	00	00	00
Cosine day of year	00	00	00	00
Sine of twice day	00	00	00	00
Cosine of twice day	00	00	00	00
Latest surface temperature	06	06	-	-
Latest surface dew point	06	06	-	-
Latest cloud cover	06	06	-	-
Latest surface U wind	06	06	-	-
Latest surface V wind	06	06	-	-
Latest surface wind speed	06	06	-	-
Latest ceiling	06	06	-	-
Previous maximum	06	-	-	-
Previous minimum	-	06	-	-

Table 2. Predictors in order of selection for temperature forecast equation for winter maximum at Las Vegas, Nev., approximately 24 hr after 0000 GMT.

Order	Predictor	Projection	Cumulative RV (%)
1.	Yesterday's max temp	-	71.7
2.	Boundary layer pot temp (PE)	24*	79.4
3.	Cosine twice day of year	-	81.9
4.	Latest surface temp (SS)	6	82.8
5.	500-1000 mb thickness (PE)	12	83.3
6.	Mean relative humidity (PE)	36*	83.9
7.	Surface dewpoint (TJ)	24	84.8
8.	Surface convergence (TJ)	24*	85.1
9.	850-mb temperature (TJ)	24*	85.5
10.	850-mb zonal wind (PE)	24	85.9

Final standard error of estimate = 3.26 ^{o}F

SS - surface synoptic observation; PE - primitive equation model; TJ - 3-dimensional trajectory model; * indicates 5-point smoothing operator was applied; projection is valid time of predictor in hours after 0000 GMT; RV is reduction of variance.

first and fourth terms selected are observed surface temperatures, while the third predictor reflects the seasonal trend of normal temperatures. The remaining seven terms of the equation are numerical predictors, with four coming from the PE model and three from the TJ model.

In order to evaluate the utility of the MOS temperature forecasts, their accuracy was compared to that of the official forecasts issued to the public at the local level for the 87 stations across the United States which are routinely verified (Zurndorfer et al., 1978). Both early and final guidance forecasts for April through September of 1977 were generated from regression equations which had been developed by stratifying archived PE model output into 3-month seasons. Operationally, the early guidance forecasts are obtained by substituting LFM fields in PE-based multiple regression equations. Observed weather elements from surface reports are not used as predictors. In contrast, the final guidance is produced a few hours later each day using PE model forecasts in PE-derived equations. Surface observations 5 to 6 hours later than the model input data are also used as predictors for the first two projections. In addition, the sine and cosine of the day of the year are involved in producing both sets of forecasts.

The guidance forecasts are expressed as calendar day maximum (max) and minimum (min) temperatures. In contrast, the local forecasts are predicted for the following 12-hr periods: max's between 1200 GMT and 0000 GMT, and min's between 0000 GMT and 1200 GMT. Using max/min observations from our Asheville data collection, we verified forecasts for projections of approximately 24 (max), 36 (min), 48 (max), and 60 (min) hours from 0000 GMT. Mean algebraic errors (mean forecast minus mean observed temperatures), mean absolute errors, and the number (or percent) of absolute errors of 10°F or more were computed for each case where all the guidance and local forecasts were available. Since the verifying observations did not correspond directly to the valid periods for the local forecasts, the magnitude of each of the verification scores should be viewed with some caution. However, general trends and relative differences between the guidance and local forecasts are still meaningful.

A comparison of the average scores for the 87 stations combined is given in Table 3. The mean algebraic errors indicate that the local forecasts are less biased (i.e., the errors are closer to zero) than both sets of guidance forecasts for the initial (24-hr) projection. This may be a reflection of the advantage the local forecaster obtains from using observed data about 3 hours later than that contained in the final guidance. In contrast, the early guidance and locals tend to be equally biased for the other three (longer-range) projections. These scores also show that the final guidance has a tendency to underforecast both the max and min temperatures; the early guidance and local forecasts are somewhat better in this respect.

The mean absolute errors in Table 3 indicate that, after the first projection, there is very little difference in the overall quality of the three types of forecasts. In fact, the early guidance, which was handicapped by lack of observed input for the first two projections, has the best mean absolute error for the 48-hr max. Conversely, the final guidance is clearly superior to both the early guidance and local forecasts in regard to having fewer absolute errors of 10°F or more (i.e., big busts) for all four projections. For the guidance, this is

Table 3. Comparative verification of early and final guidance and local max/min temperature forecasts for 87 stations, 0000 GMT cycle, April-Sept. 1977).

Forecast Projection (Hours)	Type of Forecast	Mean Algebraic Error (°F)	Mean Absolute Error (°F)	Number (%) of Absolute Errors ≥10°	Number of Cases
24 (Max)	Early	-0.8	3.3	577 (4.0)	14467
	Final	-0.6	3.1	375 (2.6)	
	Local	-0.0	2.9	461 (3.2)	
36 (Min)	Early	0.2	3.0	345 (2.4)	14490
	Final	-0.2	2.9	301 (2.4)	
	Local	0.3	3.1	419 (2.9)	
48 (Max)	Early	-0.8	3.9	969 (6.7)	14459
	Final	-1.2	4.0	962 (6.7)	
	Local	-0.9	4.1	1074 (7.4)	
60 (Min)	Early	0.1	3.7	827 (5.7)	14491
	Final	-0.4	3.6	678 (4.7)	
	Local	-0.0	3.6	743 (5.1)	

probably an indication of the increased stability associated with using PE forecasts in PE-derived equations.

4. Surface wind

MOS has also been used successfully to forecast surface wind, defined as the one-minute average direction and speed for a specific time. Ten-term single-station equations were derived by Carter (1975) at each of 233 stations in the United States by applying screening regression to PE model predictors. As with temperature, surface synoptic reports available 6 hr after numerical model input time were screened for the initial projection. Separate equations were derived for zonal (U) and meridional (V) wind components and for wind speed (S) for seven projections at 6-hr intervals from 12 to 48 hr.

Some constraints were imposed on the selection of predictors. For any given station and projection, the three equations for U, V, and S all contain the same 10 predictors, but with different regression coefficients. Further, the first three predictors were forced to be the boundary layer forecasts of U, V, and S for the valid time of the wind predictand. The remaining seven predictors were selected one at a time by picking at each step the meteorological variable which reduced the variance of any of the three predictands by the largest fractional amount.

As an example, the cool season (6 months Oct.-Mar.) equations valid 12 hr after 0000 GMT at Las Vegas, Nev., are shown in Table 4. Column 1 gives the selected predictors and columns 6, 7, and 8 give the coefficients. For these particular equations, the three PE boundary layer predictors U, V, and S resulted in reductions of

Table 4. Sample equations for estimating the U and V wind components and the wind speed, S, 12 hr after 0000 GMT at Las Vegas, Nevada, during the cool season from PE forecasts and surface observations.

Predictor	Forecast Projection (hr)	Cumulative reduction of variance			Coefficients			Units
		U	V	S	U	V	S	
Regression Constant					-39.950	-16.880	-6.333	kt
1. Boundary layer U	12	0.028	0.000	0.081	0.073	0.220	0.142	$m\ s^{-1}$
2. Boundary layer V	12	0.035	0.262	0.104	-0.074	0.041	-0.001	$m\ s^{-1}$
3. Boundary layer S	12	0.036	0.270	0.201	0.174	0.166	0.113	$m\ s^{-1}$
4. Observed S	6	0.046	0.270	0.355	0.081	-0.004	0.448	kt
5. Observed V	6	0.048	0.395	0.358	-0.017	0.422	-0.044	kt
6. 850-mb geostrophic U	18	0.067	0.434	0.362	0.119	0.293	-0.007	$m\ s^{-1}$
7. Observed U	6	0.100	0.435	0.362	0.166	-0.066	0.006	kt
8. 850-mb geostrophic S	18	0.108	0.450	0.389	-0.090	-0.214	0.200	$m\ s^{-1}$
9. 500-mb height	12	0.124	0.452	0.389	0.007	0.003	0.001	m
10. 850-mb relative vorticity x 10^5	12	0.136	0.452	0.397	0.316	-0.026	0.302	s^{-1}
Total standard error of estimate		3.40	4.69	3.40				

variance of 4, 27, and 20 percent for the U, V, and S predictands respectively. Next, as was the case at most stations for the 12-hr prediction equations, the 0600 GMT observed winds were selected. These predictors, along with four others from the PE model, produced an additional reduction of variance of approximately 10-20 percent for each predictand. Similar equations were later derived from the LFM model.

MOS forecasts of surface wind have been available for use as guidance by NWS forecasters since May 1973. The guidance and local forecasts prepared at approximately 90 stations for projections of 18, 30, and 42 hours have been verified since the 1973-74 cool season. The results were recently summarized by Glahn et al. (1978) as follows:

Mean absolute errors (MAE's) of direction were computed for all cases when the local forecasts of speed were 8 knots or greater. Figures 3 and 4 show the MAE's for direction for the cool and warm seasons. The following summary can be made: (1) the 18-hr forecasts were better than the 42-hr forecasts; (2) the guidance was definitely better than the locals; (3) both the guidance and locals improved over the 4-year period; (4) even though the locals improved considerably over the period, there was only a slight tendency for them to "gain" on the guidance; and (5) the early guidance was only slightly better than the final for the year both were available.

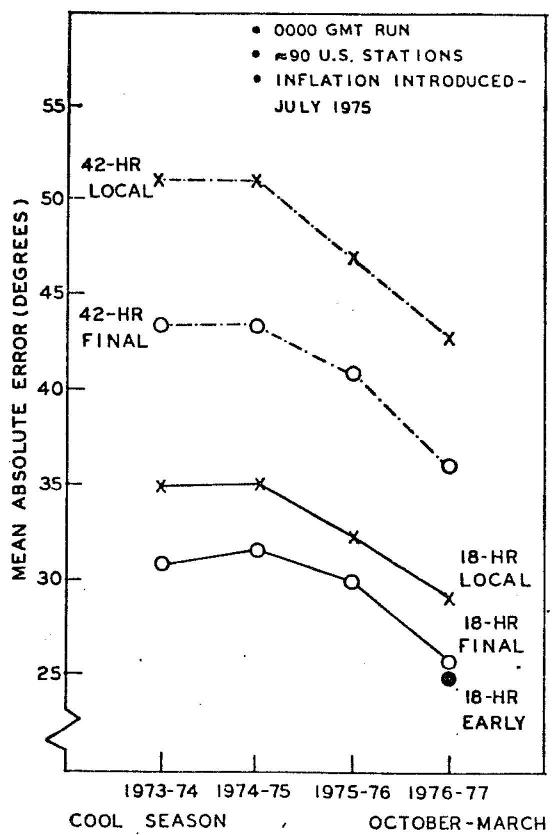

Figure 3. Mean absolute error for local and guidance surface wind direction forecasts for the cool season (October-March).

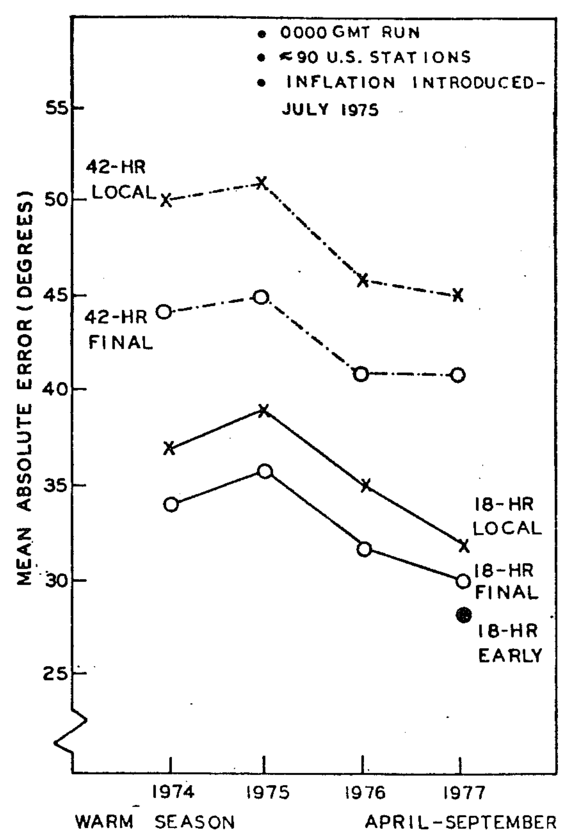

Figure 4. Same as Fig. 3 except for the warm season (April-September).

Contingency tables of forecast and observed wind speed were prepared with categories of ≤7, 8-12, 13-17, 18-22, and ≥23 knots; from these, skill scores were computed. Figures 5 and 6 show these scores for the cool and warm seasons. It should be noted that starting in July 1975 all guidance forecasts of wind speed were inflated (Klein et al., 1959).

Figures 5 and 6 can be summarized as follows: (1) the 18-hr forecasts were better than the 42-hr forecasts; (2) the guidance was definitely better than the locals; (3) both the guidance and locals improved over the 4-year period for the cool season but not for the warm season; (4) there was less difference in the cool season between the guidance and locals at the end of the 4-year period than at the beginning; and (5) the early guidance was considerably better than the final for the year both were available.

5. Probability of precipitation

Forecasts of the point probability of precipitation (PoP) during 12-hr periods have been issued by the National Weather Service since 1965, and nationwide MOS guidance for those forecasts has been produced operationally since 1972 (Lowry and Glahn, 1976). Since measureable precipitation (≥ .01 inches) does not occur often enough in

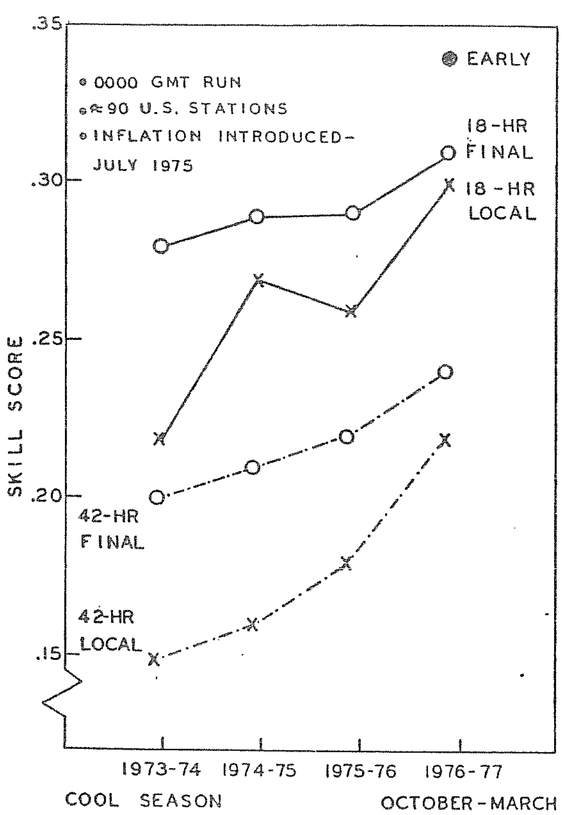

Figure 5. Skill score for local and guidance surface wind speed forecasts for the cool season (October-March).

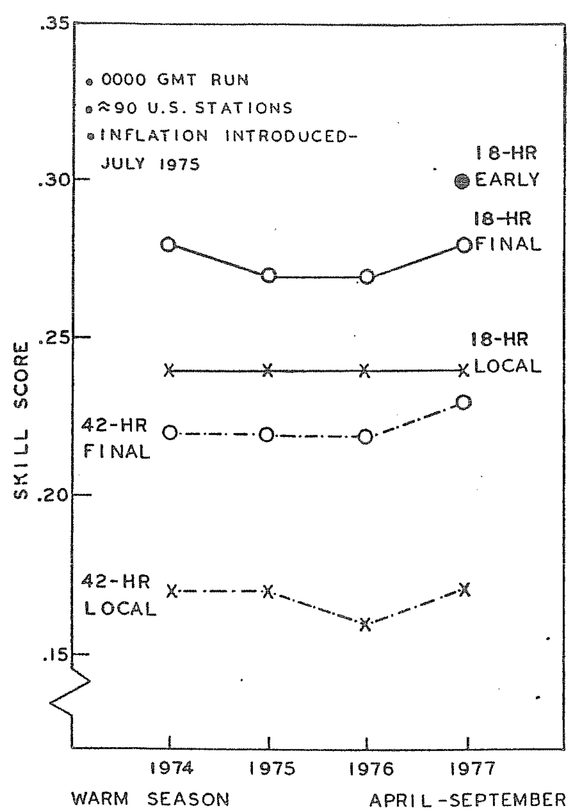

Figure 6. Same as Fig. 5 except for the warm season (April-September).

a small data sample to allow derivation of reliable single-station equations, TDL combines data from a number of stations within each of several homogeneous regions and then derives a single equation for each region. In application, the equation is used at each station within the region with input data appropriate to that particular station--a generalized operator technique (Harris et al., 1963).

For example, Fig. 7 illustrates 22 regions used to forecast first period PoP from LFM data during the warm season (Apr.-Sept.) of 1978 (Gilhousen, 1978). The boundaries were determined subjectively by analyzing the relative frequency of precipitation when the LFM model predicted $\geq$ 65 percent mean relative humidity or a precipitation amount $\geq$ .01 inches. Some of the boundaries were drawn to prevent a region from covering too large a geographical area, rather than to differentiate areas with different relative frequencies. Also, the boundaries placed between areas 10 and 15, 11 and 15, and 11 and 12 were made to separate stations that have a nocturnal rainfall maximum from those that have an afternoon maximum.

Within each region a standard set of the 70 most valuable binary and continuous LFM predictors was offered for screening, but different predictor sets were required for each of four projections. In addition,

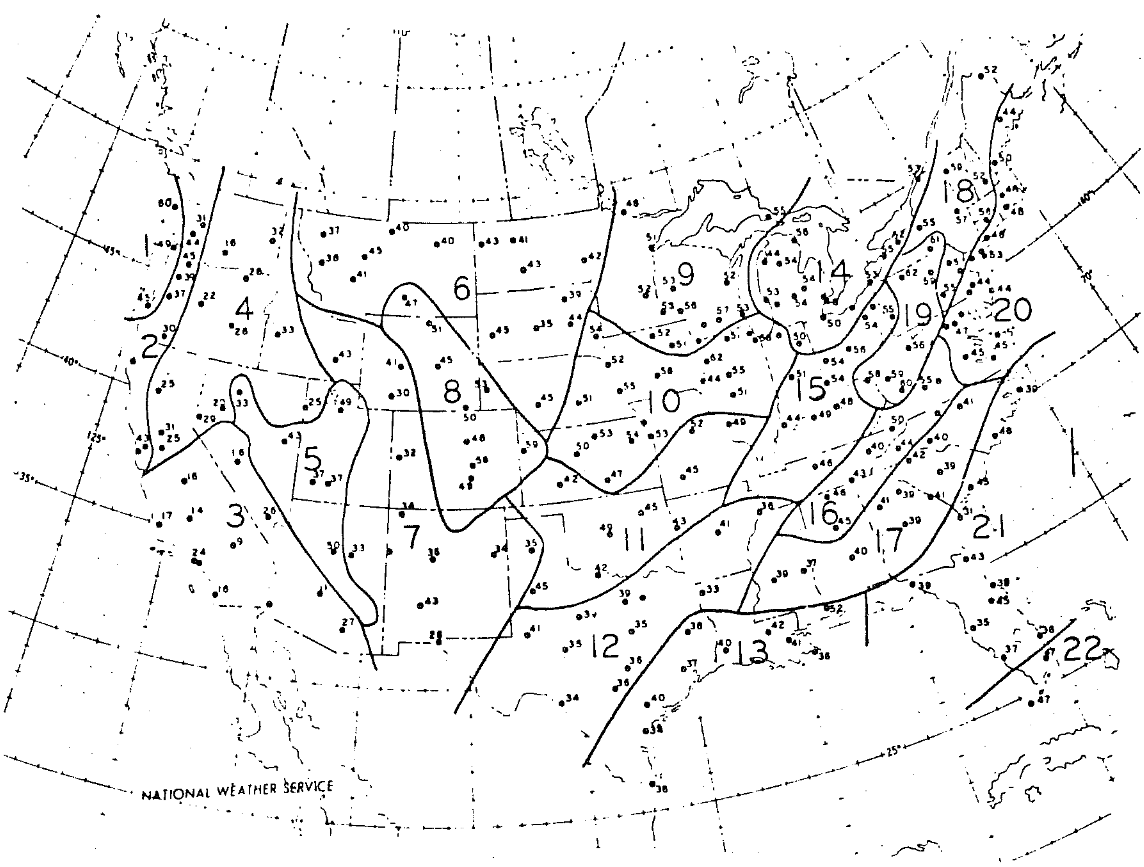

Figure 7. The 22 regions used for first period PoP during the summer of 1978. The plotted values are the relative frequency of precipitation ($\geq$.254 cm.) for cases when the 18-hr forecast of LFM mean relative humidity was $\geq$65 percent.

surface observations valid three hours after initial data time (0000 GMT or 1200 GMT), harmonics of the day of the year, and climatic relative frequencies of precipitation were added to the standard list of predictors to contribute information not provided by the LFM. Since the predictand is binary (i.e., precipitation has or has not occurred), this application of regression (Miller, 1964) produces equations which give probability forecasts of precipitation at the forecast site.

Table 5 shows the first six predictors picked by the screening process for four widely separated regions. The table applies to the "today" period (from 12 to 24 hours after 0000 GMT) during the warm

Table 5. The first six predictors picked for four regions (see Fig. 7) for the first 12-hr period PoP for 0000 GMT, summer 1978. Total R.V. is the reduction of variance for the 12-predictor equation. The number of cases is the total amount of developmental data when all stations within a region are grouped together.

Predictor	Projection	Smoothing	Binary	Additional R.V.
		Region 1 (Northwest)		
1349 cases			Total R.V. = .4837	
LFM Mean Rel. Hum.	12	5	Continuous	.3693
LFM Precip. Amt.	24	5	Continuous	.0344
LFM Mean Rel. Hum.	24	5	≤ 65%	.0232
LFM Precip. Amt.	18	5	=.0000cm	.0169
LFM 850 Mb U Wind	18	1	≤ 6 m/s	.0054
LFM Bound. Lay. Hum.	18	5	≤ 85%	.0049
		Region 5 (Mountainous Southwest)		
2696 cases			Total R.V. = .2684	
LFM Mean Rel. Hum.	12	5	Continuous	.2055
LFM Mean Rel. Hum.	18	5	≤ 65%	.0120
LFM Precip. Amt.	18	5	≤ .2540cm	.0120
LFM Total Totals Index	24	5	≤ 50	.0076
Observed Sky Cover	03	-	Scattered	.0111
LFM Bound. Lay. Hum.	18	5	≤ 85%	.0001
		Region 15 (Ohio Valley)		
4946 cases			Total R.V. = .3428	
LFM Mean Rel. Hum.	24	5	Continuous	.2826
LFM Precip. Amt.	24	5	Continuous	.0232
LFM Mean Rel. Hum.	18	5	≤ 80%	.0039
LFM 850-Mb Ht.	18	5	≤ 1510m	.0076
LFM Total Totals Index	24	5	≤ 50	.0085
LFM Mean Rel. Hum.	18	1	≤ 65%	.0057
		Region 22 (South Florida)		
900 cases			Total R.V. = .1755	
LFM Mean Rel. Hum.	12	5	Continuous	.1309
Rel. Freq. ≥.0254cm 12-00 GMT	--	-	Continuous	.0140
6-hr Δ LFM Mean Rel. Hum.	06-12	5	Continuous	.0140
LFM Mean Rel. Hum.	24	5	≤ 65%	.0092
LFM Precip. Amt.	18	5	≤.0254cm	.0001
Cosine Day of Year	--	-	Continuous	.0001

season. Note that many of the predictors selected are in binary form. For instance, a humidity predictor indicated by the symbol "$\leq$ 65%" in the binary column means that the predictor selected was set equal to one if the humidity was less than or equal to 65 percent and set equal to zero otherwise. Table 5 indicates that the mean relative humidity (from surface to 400 mb.) is the most important LFM predictor of PoP and the forecast precipitation amount is the second most important. Other predictors selected occasionally are winds and humidity in the boundary layer, height and winds at 850 mb, observed values of clouds and weather, measures of stability (like the Total Totals index), climatological frequencies of precipitation, and cosine of the day of the year.

At the same time the PoP equations were derived for 12-hr periods, TDL also derived MOS PoP equations for the two 6-hr periods within each 12-hr period. With this procedure, all three equations have the same predictors, but with different coefficients. This tends to insure consistency between the 6-hr and 12-hr forecasts; however, consistency is not guaranteed.

Although the equations of Table 5 explain less than half the variability of PoP (R.V. below 50 percent), they produce objective forecasts which are superior to climatological forecasts and competitive with the best subjective estimates. This is illustrated by Table 6, which verifies PoP forecasts produced by MOS, local offices, and climatology at 87 stations during the summer of 1977 (Zurndorfer et al., 1978) and 86 stations during the winter of 1976-77 (Bocchieri

Table 6. Verification scores for subjective (local) and objective (MOS) PoP forecasts for the 1977 summer and 1976-77 winter seasons at 87 stations in the United States.

Projection (Hours)	Type of Forecast	Brier Score	Improvement Over Climatology (%)	Number of Cases
	a) Summer (Apr. - Sept. 1977):			
12-24 (1st period)	MOS Local	.113 .110	24.0 27.0	27943
24-36 (2nd period)	MOS Local	.124 .122	18.0 19.7	27879
36-48 (3rd period)	MOS Local	.135 .132	10.9 13.2	27959
	b) Winter (Oct. 1976 - Mar. 1977):			
12-24 (1st period)	MOS Local	.082 .077	22.2 27.6	15177
24-36 (2nd period)	MOS Local	.099 .097	17.4 20.5	15196
36-48 (3rd period)	MOS Local	.104 .103	13.8 14.3	15205

et al., 1977). The forecasts were evaluated by computing the Brier Score, which is defined as one-half the P-score proposed by Brier (1950). In terms of improvement over climatology, the local forecasts were 3 to 5 percent better than MOS during the first period, but only 1 to 2 percent better during the third period. These results indicate that the local forecaster can significantly improve the MOS PoP in the first period by making intelligent use of later surface, radar, and satellite data, but improvement in the third period is very small. Table 6 also shows that improvements over climatology are smaller for the warm season than for the cool season. This is related to the fact that it is more difficult to forecast convective precipitation in summer than in winter when synoptic-scale precipitation patterns are better organized.

Figure 8 is a plot of the relative frequency of precipitation for each forecast value when all PoP data were combined on a nationwide basis (Bocchieri et al., 1977). This graph shows excellent reliability for both the local and final PoP forecasts. On the average, the frequency of observed precipitation was close to the probability of precipitation forecast by both MOS and the local forecasters, except for a slight tendency to overforecast at very high probabilities.

6. Cloud amount

Another weather element for which probability forecasts have been derived by the MOS technique is opaque sky cover, commonly known as cloud amount. Initially, separate equations were derived for each of 233 stations to estimate the probability of clear, scattered, broken,

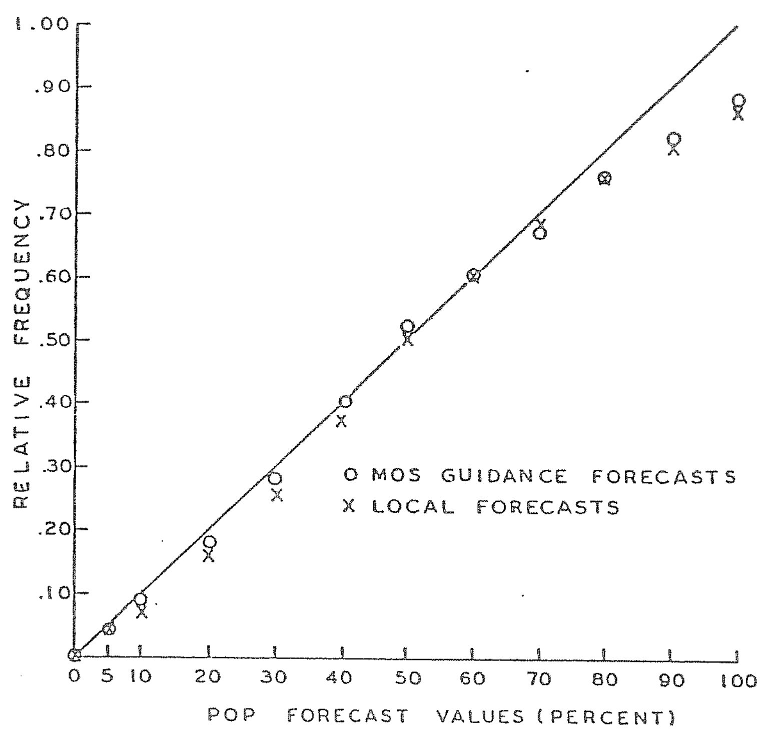

Figure 8. Reliability of PoP forecasts during cool season: Oct. 1976-March 1977.

and overcast sky conditions from numerical models and observed surface reports (Carter and Glahn, 1976). Later, a new set of equations was derived for 21 regions by applying the generalized operator technique simultaneously for both cloud amount and ceiling (Crisci, 1977). The new set of equations proved to be as accurate as the old set, while providing greater consistency between MOS cloud and ceiling forecasts.

An example of the generalized equations is shown in Table 7 for a region (not shown) located south and west of Las Vegas including parts of Nevada, Arizona, and California. As in Table 5, the predictors are expressed in both binary and continuous form and are taken from both the LFM model and surface observations. However, Table 7 has four binary predictands, instead of one, and therefore contains four separate equations which give the probability of clear, scattered, broken, and overcast, respectively. The equation for each category has the same 15 predictors, but with different coefficients, to insure that the four probability estimates always sum to unity (Miller, 1964). In operation, after the probability of each cloud category is determined, the "best" single category is obtained by inflating the probabilities and minimizing the bias of the resultant categorical forecast (Carter and Glahn, 1976). Of the 15 predictors listed in Table 7, 9 are taken from the LFM model, 5 from surface observations, and 1 from the station elevation. Most of the predictors, such as temperature-dew point spread, relative humidity, and sky cover, are directly related to cloud amount; while others, such as winds, visibility, stability (G index), and elevation, are indirectly related. Thus the MOS technique results in physically reasonable equations.

In order to evaluate the utility of the MOS cloud equations, the "best" category forecast was compared to a matched sample of local (subjective forecasts and verified by Glahn et al. (1978). The results

Table 7. Winter MOS equations for estimating cloud amount categories at 1800 GMT in a region near Las Vegas from the 0000 GMT run of the LFM model during the cool season (Oct.-Mar.).

Predictor	Limit	Tau	Smoothing	Coefficients Clear	Scattered	Broken	Overcast
LFM mean rel. hum.	Continuous	24	5 Pt.	-.01511	.00709	.00567	.00235
LFM temp-dewpoint (1000 mb)	4^{o} K	24	5 Pt.	-.04228	-.01523	.03185	.02565
Obs sky cover	8.5	06	None	-.06829	.09775	-.03008	-.1360
Obs ceiling	2950 (ft.)	06	None	-.1164	.1409	-.04497	.02044
LFM mean rel. hum.	90%	18	5 Pt.	-.03477	.2078	-.04924	-.1238
Obs sky cover	0.5	06	None	.1400	-.06210	-.05529	-.02263
LFM G index	-2725 (m.)	24	5 Pt.	.00940	.05254	.004241	-.06618
LFM rel. hum. (L1)	70%	24	5 Pt.	.03536	.04383	.01542	-.09460
LFM rel. hum. (L1)	Continuous	24	5 Pt.	.00878	-.00401	-.00343	-.00134
LFM bound. lay. rel. hum.	Continuous	18	5 Pt.	.00317	-.00111	-.00133	-.00073
LFM mean rel. hum.	Continuous	18	5 Pt.	-.00694	.00204	.00392	.00098
Station elevation	Continuous (ft.)	0	None	-.00001	.00001	-.00001	.000001
Obs vision obstruct.	0.0	06	None	-.1730	.1450	.08109	-.05305
Obs U wind	Continuous (kts.)	06	None	-.00151	.00315	.00132	-.00297
LFM mean rel. hum.	90%	24	5 Pt.	-.09921	.00671	.1529	-.06044
			Constants	1.174	-.4842	-.2103	.5206

in terms of skill score are given in Figs. 9 and 10 for 94 stations, two different forecast projections, and three different years. In summary: (1) the 18-hr forecasts were better than the 42-hr forecasts; (2) the guidance was definitely better than the locals at 42 hours; (3) the locals were generally better than the final guidance at 18 hours, but the early guidance was the best of the three for the year all were available; (4) the cool season scores were generally better than the warm season scores; and (5) trends in skill varied considerably by projection and season.

7. Precipitation type

Given the occurrence of precipitation, what is the probability it will be frozen (i.e., snow or sleet)? MOS has been applied to estimate this conditional probability of frozen precipitation (PoF) by Glahn and Bocchieri (1975) and Bocchieri and Glahn (1976). They used the logit model of Brelsford and Jones (1967), which fits an S-shaped curve to a categorical predictand as a function of continuous predictors. Since the logit model is asymptotic to zero and one, it is ideal for this type of forecasting problem.

The development proceeded in two steps. First, for each of 186 stations, a "50%" value was found for each of three variables

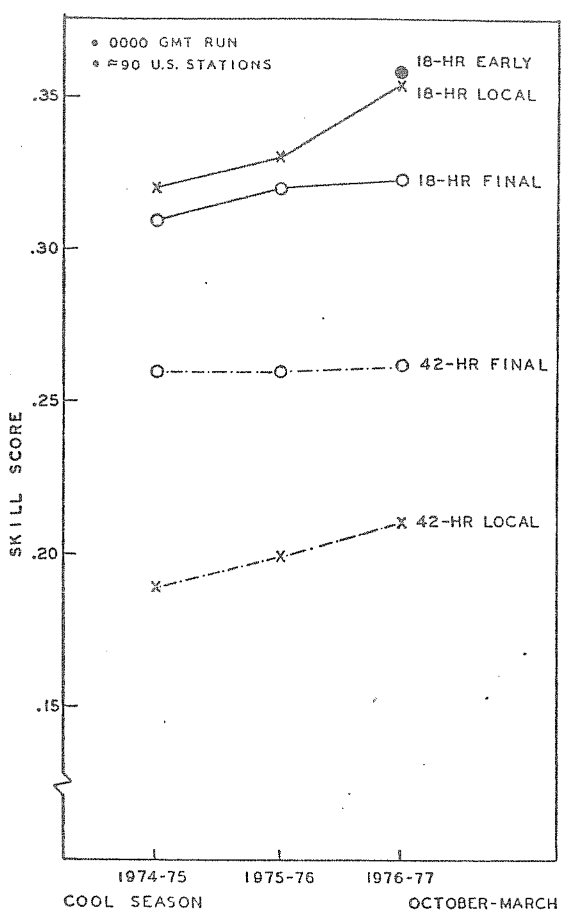

Figure 9. Skill score for local and guidance cloud amount forecasts for the cool season.

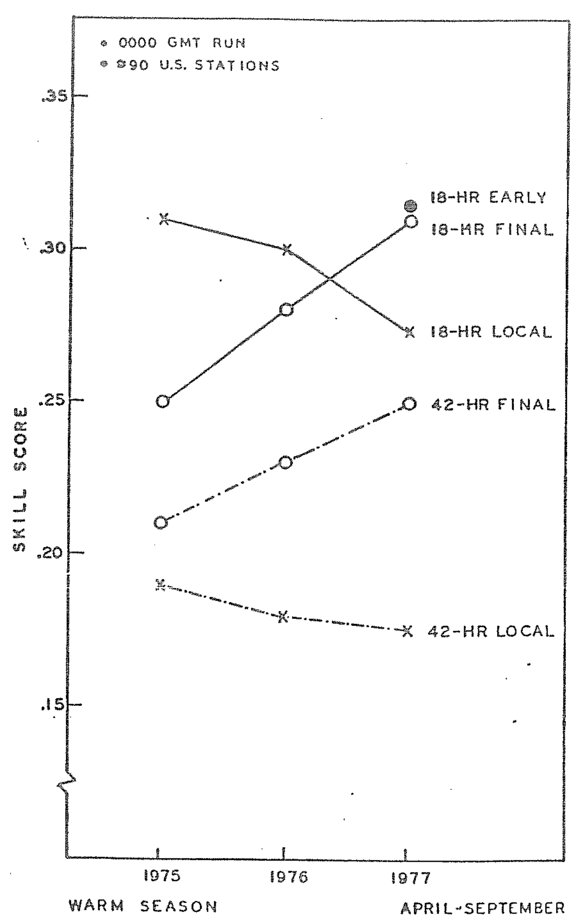

Figure 10. Same as Fig. 9 except for the warm season.

predicted by the PE model--850-mb temperature, 1000-500 mb thickness, and boundary layer potential temperature. For instance, the value of the 850-mb temperature which indicates a 50-50 chance of frozen precipitation at a particular station (provided precipitation occurs) was found. These 50% values were determined by using the logit model to fit data from three winter seasons (single station approach).

Secondly, the deviations from the 50% values were determined for each station for each variable; the relative frequency (for those cases when precipitation occurred) of frozen precipitation was then computed, again with the logit model, as a function of these new variables. In order to get stable results in this last step, data for all stations were combined (generalized operator). In addition to the meteorological variables, station elevation and the first harmonic of the day of year were used as predictors.

The predictors actually put into the logit model for four different times are shown in Table 8. The objective screening of a large number of variables was not practical, since the logit model doesn't lend itself to this technique. Therefore, the predictors were chosen subjectively, with care being taken to be sure they are reasonable. Notice once again that forecast values for other times than the verifying time are included in an attempt to allow for the biases of the numerical model.

How good are the forecasts from the logit model? The first year a comparative verification was made between the locals and the guidance forecasts was the cool season 1973-74. In the NWS verification, local categorical forecasts of precipitation type made at about 1000 GMT are recorded for the valid times 1800 GMT (today), 0600 GMT (tonight), and 1800 GMT (tomorrow). Note that this is a conditional forecast; i.e., it is a forecast of type of precipitation if precipitation occurs. Therefore, it is available whether or not precipitation occurs. In this verification, a guidance forecast of frozen precipitation is defined as a PoF >50 percent.

The results for four seasons are shown in Fig. 11 (Glahn et al., 1978). It should be noted that some changes in the verification procedure took place during these four years. First, the number of stations changed from approximately 90 for the first two years to approximately 60 for the last two years. Second, starting in the

Table 8. Predictors in the PoF equations for each of four projections when predictors are taken from the PE model.

12 hr Predictor	Projection	24 hr Predictor	Projection	36 hr Predictor	Projection	48 hr Predictor	Projection
Station elevation	—	Station elevation	—	Station elevation	—	Station elevation	—
Sin DOY	—	Sin DOY	—	Sin DOY	—	Sin DOY	—
Cos DOY	—	Cos DOY	—	Cos DOY	—	Cos DOY	—
1000-500 mb thickness	12	850-mb temperature	12	850-mb temperature	24	850-mb temperature	36
850-mb temperature	12	B.L.P. temperature	12	B.L.P. temperature	24	B.L.P. temperature	36
B.L.P. temperature	12	1000-500 mb thickness	24	1000-500 mb thickness	36	1000-500 mb thickness	48
850-mb temperature	24	850-mb temperature	24	850-mb temperature	36	850-mb temperature	48
B.L.P. temperature	24	B.L.P. temperature	24	B.L.P. temperature	36	B.L.P. temperature	48
		850-mb temperature	36	850-mb temperature	48		
		B.L.P. temperature	36	B.L.P. temperature	48		

1975-76 season, TDL verified only those cases when the local PoP was 30 percent or greater because of concern that the forecasters might not have put much effort into making the conditional PoF forecasts when they thought precipitation to be unlikely. Also, for the 1976-77 season, TDL verified the early PoF guidance for the 18-hr projection (based on the LFM) as well as the final guidance (based on the PE).

Figure 11 indicates: (1) the 18-hr forecasts were better than the 42-hr forecasts; (2) the guidance was consistently better than the local forecasts in terms of skill score; (3) there has been improvement in both the guidance and locals over the four-year period, especially the locals; and (4) there was little difference in the early and final guidance in terms of skill score.

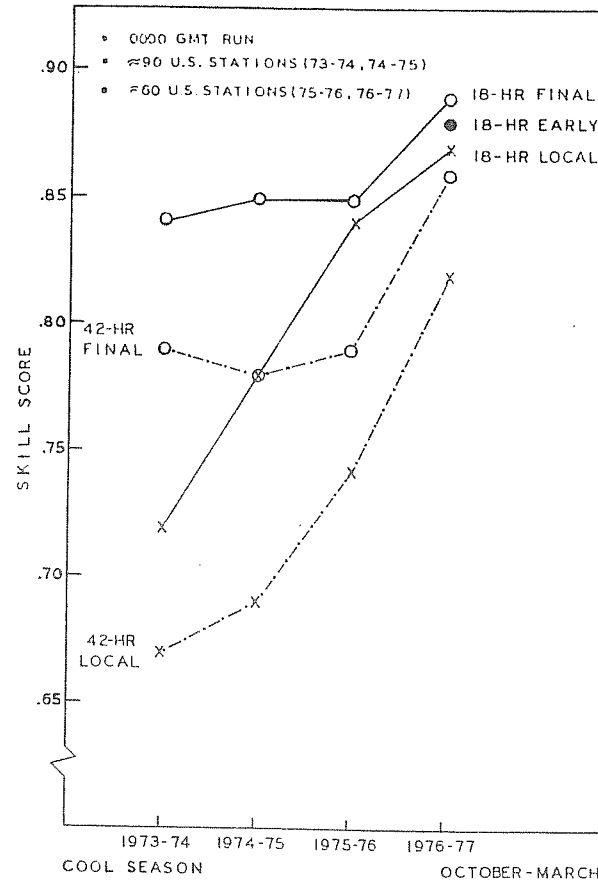

Figure 11. Skill score for local and guidance PoF forecasts.

It is possible that part of the improvement of the locals during the last two seasons was due to verifying only those cases when the local PoP's were 30 percent or greater. We also believe the forecasters are using the guidance more now than when it first started.

Recently Bocchieri (1978) developed a new system, called PoPT, which gives conditional probability forecasts for three precipitation type categories: snow or sleet, freezing rain, and rain or mixed types. Freezing rain forecasts weren't available in the previous system (PoF). Also, the probability forecasts are transformed into categorical forecasts so that a "best category" is provided. The MOS technique was used with output from the LFM model to develop logit forecast equations. Forecasts are valid at every sixth hour from 12 through 48 hours after both the 0000 GMT and 1200 GMT cycle times.

In addition to new precipitation type categories, the PoPT system differs from PoF in a number of other ways. First, to help account for the evaporational cooling effect, new predictors such as BLWBT, 850 WBT, OBS SFC T, and OBS SFC Td (see Table 9) are included. Secondly, new predictors are designed to help model predictor interactions. Thirdly, generalized operator forecast equations are developed for each of seven regions in the conterminous United States; in the PoF system, regionalization was not used.

Table 9. The potential predictors included in the development of the PoPT system.

Abbreviation	Definition
a. Model output Predictors	
BLPT	Boundary layer potential temperature
BLWBT	Boundary layer wet-bulb 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
10-8.5 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	850-mb temperature and boundary layer potential temperature
10-5 Th + BLPT	1000-500 mb thickness and boundary layer potential temperature
8.5-5Th + 10-8.5 Th	850-500 mb thickness and 1000-850 mb thickness
850 T + BLWBT	850-mb temperature and boundary layer wet-bulb temperature
c. Observed surface and miscellaneous predictors	
OBS SFC T	Observed surface temperature
OBS SFC Td	Observed surface dew-point temperature
COS DOY	Cosine of the day of year

Bocchieri used linear screening regression to determine the predictors to include in logit forecast equations. Results from the screening vary by region but generally indicate that the 850 T + BLWBT joint predictor is the most important for the snow category. For the freezing rain category, the 8.5-5 Th + 10-8.5 Th, 10-5 Th + BLPT, and the 850 T + BLPT joint predictors are picked relatively early in the screening process, along with the OBS SFC T and OBS SFC Td.

A comparative verification between the PoPT and PoF systems on independent data indicates that PoPT is generally 4 to 8 percent more accurate than PoF in terms of the probability forecasts for the snow category. Verification of the PoPT system for the 24-hr projection on both developmental and independent data samples shows that the scores for the rain and snow categories are generally quite good and stable. However, the scores for the freezing rain category deteriorated somewhat on the independent sample. Therefore, to improve the stability, Bocchieri rederived the forecast equations with data from the developmental and independent sample periods combined.

The new system will be implemented in Sept. 1978. Conditional probabilities of precipitation type will be transmitted over teletypewriter. The categorical forecasts will be available on a new early guidance, 4-panel facsimile chart which gives probability of precipitation (PoP) forecasts for four 12-hr periods and precipitation type forecasts valid at the beginning of each period. Figure 12 shows a sample panel from the new chart; the PoP forecasts (solid lines) are analyzed at 15 percent intervals, and the precipitation type forecasts (hypothetical) are represented by symbols at stations where the PoP is ≥ 25 percent. Since the PoP values are smoothed by the analysis, there may be stations showing precipitation type forecasts in areas where the analysis shows PoP to be < 25 percent, or no precipitation type forecasts where PoP is analyzed to be ≥ 25 percent. The facsimile chart will be available at about 0646 GMT and 1921 GMT.

8. Probability of precipitation amount

The next weather element to be discussed is the probability of precipitation amount (PoPA). MOS has been applied to this element by Bermowitz and Zurndorfer (1978) to forecast the probabilities of occurrence in five categories (Table 10). For the country as a whole, the first category (less than 1/4 inch) accounts for 95 percent of the precipitation events, while the last category (2 inches or greater) accounts for only 0.1 percent of the events. This distribution makes the large amounts very difficult to forecast.

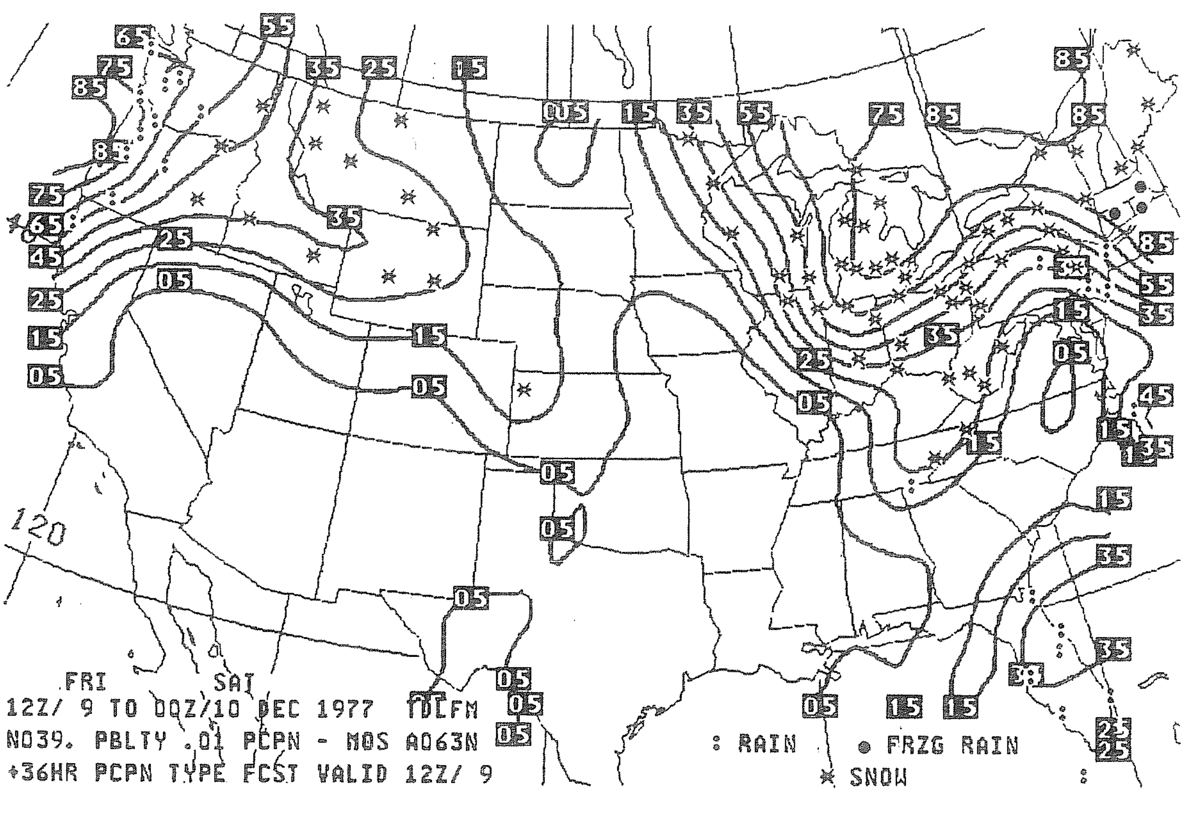

Figure 12. A sample panel from the early guidance, 4-panel facsimile chart that shows analyzed PoP forecasts (solid lines) and categorical precipitation type forecasts at stations where PoP is ≥ 25 percent. The chart is available at about 0646 GMT and 1921 GMT.

Table 10. Categories used for forecasting probability of precipitation amount.

Category	Amount of Precipitation (inches)
1	< .25"
2	≥ .25"
3	≥ .50"
4	≥ 1.00"
5	≥ 2.00"

The developmental data consisted of 6-hr observed precipitation amounts at 233 stations in the conterminous U.S. Five seasons of data were used to develop the final (PE and TJ models) equations for both the cool and warm seasons; three seasons each were used to develop the early (LFM) equations for both cool and warm seasons.

The regional approach was used for development of equations. The data were combined from a number of stations within a relatively homogeneous region; one equation was then derived for that region. This technique was used because the data would not support development of stable single station equations; it is especially useful for prediction of rare events such as heavy precipitation. Different regions were used for early and final equations and for the warm and cool seasons.

The regions were determined by a subjective analysis of the relative frequency of occurrence of observed precipitation amounts for various amounts forecast by either the PE (for final equations) or LFM (for early equations) models. They were drawn in such a way as to keep the relative frequency fairly constant in each region. Figure 13 shows the final cool season regions. Regions for early warm and cool season equations and for final warm season equations are similar. Unfortunately, as shown in Fig. 13, the regions are fairly large; however, if they were smaller, there would not be enough data in each region to derive stable equations.

Predictors, in binary and continuous form, were obtained from appropriate model forecast fields interpolated to each of the 233 stations for which TDL has developmental data (see Fig. 2). The predictors offered were those that would be expected to have a physical relationship with precipitation. From the LFM model, predictors included forecasts out to 24 hours of precipitation amount, humidity in layers, precipitable water, vertical velocity, wind components, heights of constant pressure surfaces, moisture divergence in the boundary layer, and stability indices. From the PE model, predictors included were similar to those from the LFM model; however, forecasts out to projections of 48 hours were available. From the TJ model, predictors included forecasts out to 24 hours of net vertical displacement, precipitation amount, humidity, boundary layer moisture divergence, and stability indices. A few other predictors screened were station elevation and sine and

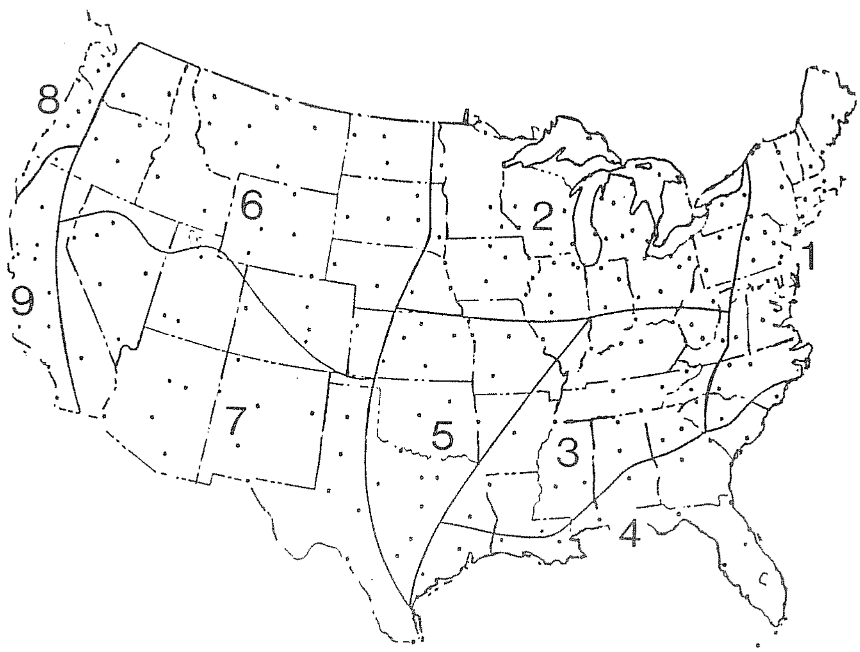

Figure 13. Regions used to develop cool season late guidance PoPA equations. Dots show the locations of the 233 stations used in the development.

cosine of the day of the year; the latter two are designed to capture a seasonal trend of precipitation amount, if any exists. In general, the most important predictors were found to be precipitation amount and surface to 490-mb mean relative humidity from the PE and LFM models.

The PoPA equations for all categories for a given projection and region were derived simultaneously. This means that the resulting equations for each of the categories have the same predictors in the same order. Of course, the coefficients preceding each of the predictors usually differ in each equation. Twelve-term equations were developed in all cases since recent evidence has indicated that twelve is about the optimum number of predictors to use in the PoPA system.

Since we want to make categorical forecasts of precipitation amount, the probability forecasts were transformed to categorical forecasts by maximizing the threat score, the primary statistic used for verification of precipitation amount forecasts at NMC. This was done by deriving threshold probabilities for all categories for each region and projection.

To determine the quality of the PoPA categorical forecasts over a relatively extended period, TDL established a comparative verification program with the assistance of NMC. Operational early and final categorical forecasts were compared against those produced subjectively at NMC (SUBJ) and by the LFM and PE models. Threat scores and biases were computed for all forecast systems at 215 cities for the categories $\geq$ .25, $\geq$ .50, $\geq$ 1.0, and $\geq$ 2.0 inches for projections 12-36 and 36-60 hr after 0000 GMT and 24-48 hr after 1200 GMT.

They were also computed for the categories $\geq$ .25, $\geq$ .50, and $\geq$ 1.0 inch for projections 18-24 hr after 0000 GMT and 12-18 hr after 1200 GMT. Subjectively prepared forecasts for these 6-hr projections were verified only for the category $\geq$ .25 inch, since NMC did not record categorical forecasts greater than that. The verification period consisted of ten months, June 1976 to March 1977, and was divided into a warm and cool season consisting of about 100 and 175 days, respectively.

Results of the comparative verification have been analyzed in detail by Bermowitz and Zurndorfer (1978). Their comparison of the warm season PoPA forecasts against those of the PE and LFM models shows that the PoPA forecasts had better threat scores with only one exception--the 2.0 inch category for the 12-36 hr projection where the LFM had a higher threat score. Overall, the warm season PoPA forecasts were slightly better than those prepared subjectively at NMC, although considerable variation was evident. Both PoPA and SUBJ tended to overforecast precipitation amount. On the other hand, the LFM and PE tended to underforecast it, especially the higher amounts.

Results of the cool season verification show that the PoPA forecasts had better threat scores than those of the PE and LFM models. This is especially so when PoPA is compared to the PE. However, PoPA generally had lower threat scores than SUBJ. When compared to the results obtained for the warm season, there appears to be some degradation of the cool season PoPA forecasts with respect to those of SUBJ, PE, and LFM. The bias indicates that PoPA and SUBJ tended to overforecast precipitation amount, while the PE and LFM underforecasted the higher amounts. These results are similar to those obtained for the warm season.

To summarize, the results of the verification at 215 cities over a ten-month period indicate that the PoPA forecasts are, indeed, useful guidance. They appear to be superior to the precipitation amount forecasts made by the PE and LFM models and almost as good as those prepared subjectively at NMC.

Twice per day, early and final guidance PoPA forecasts are supplied as guidance to forecasters at NMC and in the field. NMC forecasters receive computer printouts of probability and categorical forecasts in tabular form, as well as categorical forecasts in map form.

Field forecasters receive the PoPA forecasts by request through the Federal Aviation Administration's Weather Message Switching Center in Kansas City. The probability forecasts sent on teletypewriter are rounded to the nearest 10 percent; therefore, a forecast that appears as 5321/3 means that the probabilities of the categories $\geq$ .25, $\geq$ .50, $\geq$ 1.0, and $\geq$ 2.0 inches, which precede the solidus, are 50, 30, 20, and 10 percent, respectively. The number following the solidus is the categorical forecast, in this case, .50-.99 inch.

9. Thunderstorms and severe local storms

Reap and Foster (1977) applied the MOS technique to generate 12 to 36 hr probability forecasts of thunderstorms and severe local storms. The equations were derived from a 3-year predictand sample consisting of data for 517 days from March 15 to September 15, for the years 1974-76. The predictand sample for thunderstorms consisted of manually-digitized radar (MDR) data collected from hourly teletypewriter reports and archived on magnetic tape. The MDR data were tabulated for blocks approximately 75-80 km on a side; the area covered by these blocks is shown in Figure 14. Both the echo intensity and coverage within each block are considered in accordance with a code shown in Table 11. Code values of 4 or greater within the period ± 12 hours from 0000 GMT were used to identify thunderstorm occurrences for the general thunderstorm equation. In effect, the thunderstorm probability equation predicts the occurrence of radar echoes with an intensity of VIP3 or greater within a grid block during the 12-36 hr valid period.

The predictand data for localized severe storms consisted of reports of tornadoes, surface hail ≥ 3/4 in (~2 cm) in diameter, and wind gusts ≥ 50 kts (~93 km/hr) and/or wind damage. These reports were extracted from archive tapes edited at the National Severe Storms Forecast Center (NSSFC) to eliminate any identifiable sources of error such as redundant, mispolotted, or false reports.

The predictor sample was based on forecast fields generated by the PE and TJ models. Of the 173 predictors tested, most were 24-hr forecasts based on 0000 GMT initial data, but a few 12-, 18-, and 36-hr forecast fields were included.

The most important innovation in the work was the development of an interactive predictor for simulating large seasonal variations in

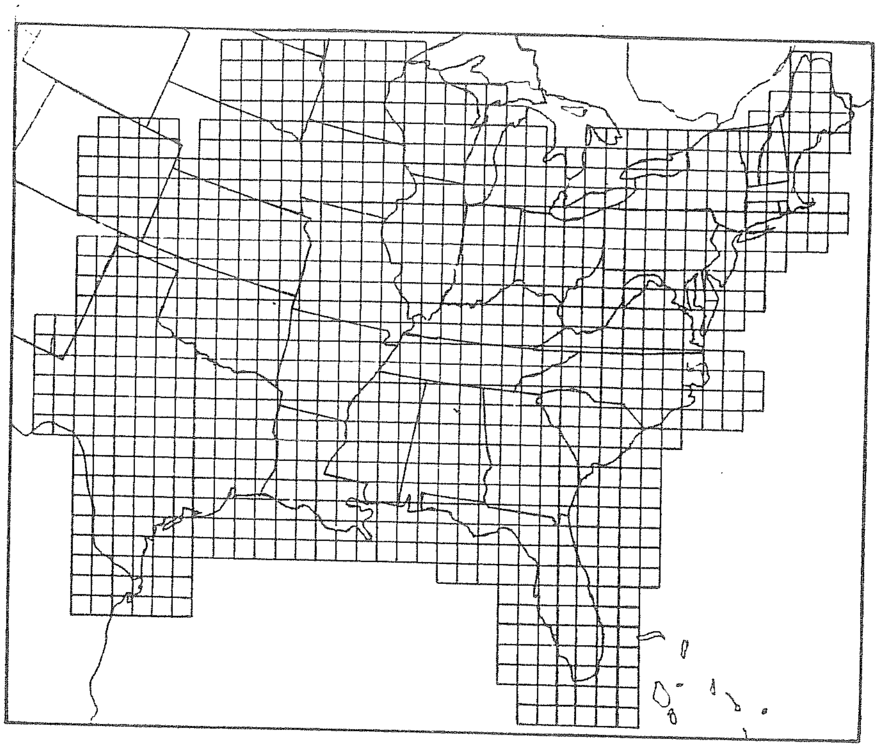

Figure 14. MDR grid region showing blocks that are approximately 75-80 km on a side.

Table 11. Manually Digitized Radar (MDR) code. VIP1, VIP2, etc. are time-averaged echo reflectivity values obtained from the operational video integrator and processor unit which is standard in the NWS radar network.

Code No.	Coverage in Box	Intensity Category
0		
1	any VIP1	Weak
2	≤1/2 of VIP2	Moderate
3	>1/2 of VIP2	
4	≤1/2 of VIP3	Strong
5	>1/2 of VIP3	
6	≤1/2 of VIP3 and 4	Very Strong
7	>1/2 of VIP3 and 4	
8	≤1/2 of VIP3, 4, 5, and 6	Intense or Extreme
9	>1/2 of VIP3 4, 5, and 6	Intense or Extreme

thunderstorm occurrence. Such variations are often related to sub-synoptic-scale processes not adequately resolved by the large-scale model predictors.

The interactive predictor (KF) is formed by multiplying the large-scale K stability index of J. J. George by the daily thunderstorm relative frequencies obtained from MDR data. This combination forces the climatology, as represented by the thunderstorm frequencies, to be more responsive to the daily synoptic situation. For example, a cold frontal passage during summer in the southeastern United States will often sharply depress the K index. As a result, the contribution of the interactive predictor to the probability forecast will be minimized, even though the thunderstorm frequency may be quite high for that particular region and time of year. Thus the interactive term modulates the climatic contribution.

To enhance the effect of local variability in the thunderstorm probability forecasts, Reap and Foster (1977) developed statistical relationships between the interactive predictor (KF) and thunderstorm occurrence separately for each of the MDR grid blocks shown in Fig. 14. Probability estimates $Y_{i,j}$ were obtained for each block by a simple linear regression equation. They then treated $Y_{i,j}$ as a new candidate predictor and submitted it along with all other predictors to the screening regression program to develop a single generalized operator equation over the entire MDR grid region. This procedure maximizes the local variability in the thunderstorm probability forecasts and eliminates any need for stratifying the sample by geographical regions. The effectiveness of the new predictor $Y_{i,j}$ is shown by the fact that it was selected first in the final multiple regression equations (Table 12), which gives 24-hr probabilities for

Table 12. Thunderstorm probability equation for period March 15 to September 15. 24-hr probabilities are valid within the interval 12-36 hours after 0000 GMT initial time. Predictors are 24-hr primitive equation model (PE) or trajectory model (TJ) forecasts.

Predictor	Model	Coefficient	Reduction of Variance (%)
Equation constant	--	-223.0	--
$Y_{I,J}$	TJ	0.99	28.92
Total Totals index (modified)	TJ/PE	1.63	1.55
500 mb wind speed (m/s)	PE	0.48	0.59
Boundary-layer moisture divergence (10^5/sec)	TJ/PE	-0.39	0.42
Surface dew point (°K)	TJ	0.51	0.15
K index	TJ/PE	-0.33	0.27
1000-400 mb mean relative humidity (%)	PE	0.29	0.41
Surface-700 mb mean relative humidity (%)	TJ	-0.25	0.25
TOTAL			32.56

the convective season for the entire MDR grid region shown in Fig. 14. The eight predictors shown gave a total reduction of variance of 32.6 percent with a corresponding multiple correlation coefficient of 0.57.

As given by Table 12, the leading term in the thunderstorm equation was the interactive predictor $Y_{i,j}$. This predictor captures most of the widespread general convective activity. The next four predictors are normally associated with more localized severe convective thunderstorms and serve to identify such storms in the probability forecast.

An equation similar to Table 12 has been derived for severe local storms. This equation gives the conditional probability of tornadoes, large hail, or damaging winds, given the occurrence of an ordinary thunderstorm. Forecasts based upon these equations are transmitted once daily by facsimile and teletypewriter. A sample facsimile chart is shown in Fig. 15. Note that the solid lines are absolute probabilities, but the dashed lines are conditional.

Verification of the forecasts shown in Fig. 15 is difficult because no other forecasts of this nature are available. However, Foster and Reap (1978) recently transformed the probabilities into categorical forecasts and compared them to severe storm outlooks routinely prepared by NSSFC in Kansas City during 1976 and 1977. Their statistics show positive skill for both sets of forecasts and no significant difference in overall accuracy.

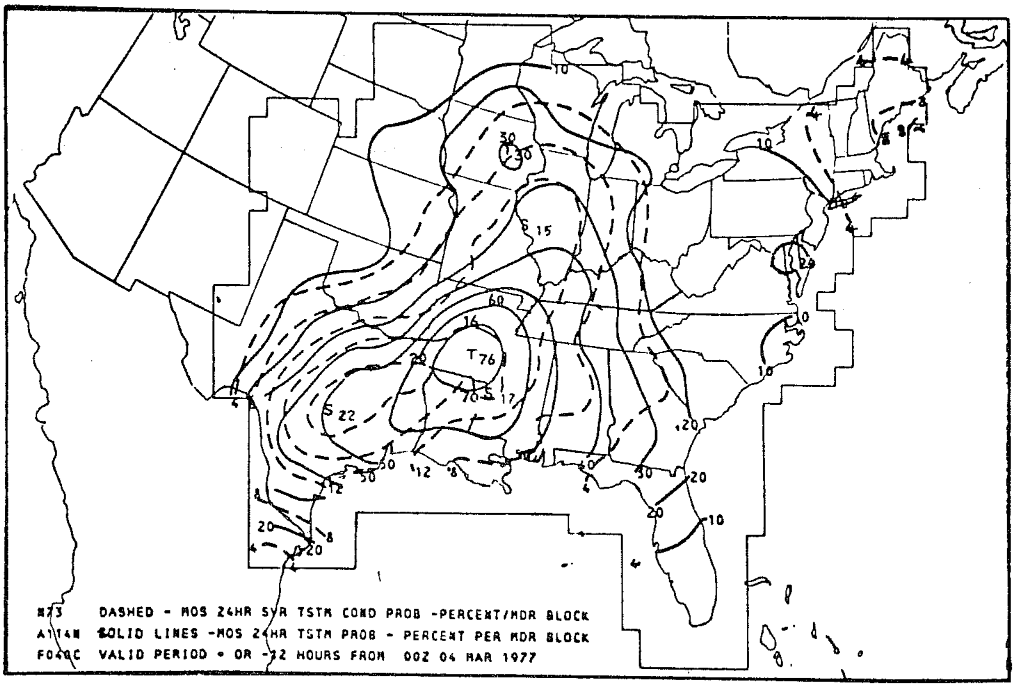

Figure 15. Computer-drawn map of thunderstorm probability (solid) and conditional probability of tornadoes, large hail, or damaging winds (dashed). The probabilities are valid for each MDR block during the 12-36 hr interval following 0000 GMT initial time or ± 12 hrs from 0000 GMT the next day.

10. Operational aspects

The MOS equations are applied twice daily on the large NOAA computer in Suitland, Md., immediately after the NMC numerical models are run. The MOS forecasts are then distributed to field stations by means of various teletypewriter bulletins and facsimile maps. One of the most complete bulletins is illustrated in Fig. 16. It gives the MOS forecasts prepared for Reno from the 1200 GMT cycle on June 2, 1978. Line 1 contains PoP forecasts for 6-hr periods ending at the date and time shown on the top line. Line 2 gives 12-hr PoP forecasts out to 60 hr from initial time (section 5). Thus the probabilities of precipitation for 12-hr periods ending 1200 GMT June 3, 0000 GMT June 4, and 1200 GMT June 4 are 0, 0, and 2 percent, respectively. The next two lines show forecasts of quantitative precipitation in both probabilistic and categorical form for 6- and 24-hr periods (section 8). Line 5 gives the probability that precipitation, if any, will be frozen (snow or sleet) (section 7). Line 6 gives forecasts of maximum and minimum temperatures for June 3 and 4 (section 3). Line 7 gives MOS forecasts of surface temperature every 3 hours from 6 to 51 hours in advance (Carter et al., 1978).

Line 8 gives the surface wind forecasts described in section 4. The forecasts are valid every 6 hr from 12 to 48 hr in advance. For example, the wind is expected to blow from 310 degrees with a speed of 10 knots at 0000 GMT on June 3. Line 9 illustrates the cloud amount forecasts described in section 6. The forecasts are given in increments of 6 hr from 12 to 48 hr after run time. The first

HDNG FOUS12 MOS FCSTS EARLY GUIDANCE 6/Ø2/78 12ØØ GMT

DATE/GMT		02/18	Ø3/ØØ	Ø3/Ø6	Ø3/12	Ø3/18	Ø4/ØØ	Ø4/Ø6	Ø4/12
RNO	POPØ6		Ø	Ø	Ø	Ø	Ø		
	POP12				Ø		Ø		2
	QPFØ6			ØØØ/1	ØØØ/1	ØØX/1			
	QPF24								ØØØØ/1
	POF	Ø	Ø	Ø	Ø	Ø	Ø	Ø	Ø
	MN/MX				41		82		45
	TEMP	67 74	75 68	59 5Ø	45 55	7Ø 77	79 74	65 55	49 58
	WIND	11Ø3	311Ø	32Ø6	2ØØ3	Ø3Ø2	3111	33Ø4	2ØØ2
	CLDS	91ØØ/1	433Ø/2	811Ø/1	811Ø/1	7111/1	622Ø/1	711Ø/1	821Ø/1
	CIG	XXØØØ9	XXXØ28	XXØØØ9	XØØØØ9	XXØØ19	XXXØ19	XXØØØ9	XØØØØ9
	VIS	XXØØØ9	XXØØØ9	XXXXØ9	ØXØØØ9	XXXØØ9	XXØØØ9	XXXXØ9	ØXØØØ9
	C/V	6/6	6/6	6/6	6/6	6/6	6/6	6/6	6/6

Figure 16. Example of MOS teletypewriter bulletin issued on June 2, 1978 from the 1200 GMT cycle. Forecasts for Reno are given on 12 different lines for 6 to 48 hr in advance, valid at the date/ times shown on line 2.

four numbers indicate the probability (in tens of percent) of clear, scattered, broken, and overcast, while the fifth number gives the "best" category. The next two lines give the probability of each of six categories of ceiling and visibility at 6-hr intervals from 12 to 48 hr after the 0000 GMT cycle, while the last line shows the "best" category (Crisci, 1977).

11. Conclusion

MOS forecast bulletins of the type illustrated in Fig. 16 are now available twice a day for approximately 360 civilian and military stations in the United States. They provide good guidance for almost all weather elements needed by the public. The objective, centralized MOS forecasts are about as skillful as subjective manual predictions produced by experienced forecasters at the local level, at least for projections of 24 to 60 hr on the synoptic scale. However for shorter time periods and smaller space scales, the human forecaster can still improve over MOS guidance.

Experience over the past few years has shown that the MOS forecasts are not unduly sensitive to the precise numerical model in operation. For example, MOS equations derived from the PE model have been applied to the LFM model to produce early guidance for several weather elements without significant deterioration in the quality of the forecasts. Additionally, changes were made by NMC to reduce the mesh length of both the LFM model (in Aug. 1977) and the PE model (in Jan. 1978), but the MOS forecasts showed no noticeable change in accuracy. Apparently the improved quality of the fine mesh models compensates for any change in their bias characteristics incorporated in the MOS equations. Of course, we will eventually rederive the MOS equations from the new models and probably produce better weather forecasts. But meanwhile, weather services should not delay starting a MOS system just because a change in numerical models is planned for the future.

Acknowledgment

The author is sincerely grateful to the following members of the Techniques Development Laboratory, Systems Development Office, National Weather Service, who have provided most of the material on which this article is based: H. R. Glahn, Joseph Bocchieri, Robert Bermowitz, Edward Zurndorfer, Gary Carter, David Gilhousen, Paul Dallavalle, David Vercelli, Ronald Reap, Donald Foster, and Dale Lowry.

Appendix

This appendix gives definitions, in mathematical terms, of various statistical quantities and verification scores used in the test.

a) Statistical quantities

The simple correlation coefficient may be expressed as:

$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}} \qquad (1)$$

where x_i and y_i are the ith values of the two variables being correlated, $\bar{x}$ and $\bar{y}$ are their mean values over the sample, and the summation is performed over n cases.

A simple linear regression equation is:

$$\hat{y}=a+bx \qquad (2)$$

where a and b are regression coefficients, y is the forecast value of the predictand or dependent variable, and x is the predictor or independent variable. The values of the constants a and b are determined by the method of least squares, which minimizes the squares of the differences between forecast values $\hat{y}$ and observed values y. The coefficient b is the slope of the line of regression and is given by the formula:

$$b=r\frac{s_y}{s_x} \qquad (3)$$

while a depends on the mean values of the two variables in accordance with the formula:

$$a=\bar{y}-b\bar{x} \qquad (4)$$

wyere $\bar{y}$ is the mean of the predictand, $\bar{x}$ is the mean of the predictor, r is the correlation coefficient defined in equation (1), s_y is the standard deviation of the predictand, and s_x is the standard deviation of the predictor. An unbiased estimate of s_x is given by the formula

$$s_x=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} \qquad (5)$$

where x_i represents any individual value of x from the first to the nth case, and $\bar{x}$ is the sample mean. The square of the standard deviation is called the variance.

Multiple linear regression relates one variable Y, called the dependent variable or predictand, to k other variables X_i, called the independent variables or predictors. The result is an equation which can be used for estimating the predictand as a linear combination of the predictors:

$$\hat{Y}=a_0+a_1X_1+a_2X_2+\cdots+a_kX_k. \qquad (6)$$

The carat indicates an estimate, and the a_i's are the regression constant and coefficients. The a_i's are determined such that the sum of the squares of the estimation errors is a minimum on the developmental (or dependent) sample of size n, i.e.,

$$\sum_{j=1}^{n}(y_j-\hat{y}_j)^2=\text{minimum}. \qquad (7)$$

A measure of the goodness of the equation for estimating Y is the reduction of variance RV, where

$$RV=\frac{\frac{1}{n}\sum_{j=1}^{n}(y_j-\bar{y})^2-\frac{1}{n}\sum_{j=1}^{n}(y_j-\hat{y}_j)^2}{\frac{1}{n}\sum_{j=1}^{n}(y_j-\bar{y})^2}. \qquad (8)$$

This is the fractional part of the variance of Y about its mean $\bar{Y}$, measured by the variance

$$\sigma_y^2=\frac{1}{n}\sum_{j=1}^{n}(y_j-\bar{y})^2 \qquad (9)$$

that is "explained" by the regression equation. RV is the square of the multiple correlation coefficient, i.e.,

$$RV=R^2_{Y.X_1,X_2,\ldots,X_k}. \qquad (10)$$

It is clear from the above equations that decreasing the sum of squares of the estimation errors is tantamount to increasing the reduction of variance RV and to decreasing the root mean square error (or standard error of estimate), where

$$RMSE=\left[\frac{1}{n}\sum_{j=1}^{n}(y_j-\hat{y}_j)^2\right]^{\frac{1}{2}}. \qquad (11)$$

Inflation is a technique to equalize the variability of forecast and observed quantities. The inflated estimate $\hat{y}'$ is defined by

$$\hat{y}'=\frac{\hat{y}-\bar{y}}{R}+\bar{y}\ , \qquad (12)$$

where $\hat{y}$ is the regression estimate, $\bar{y}$ the mean of the variable in the dependent sample, and R the multiple correlation associated with the regression equation. This procedure increases the root mean square error but may give a more desirable distribution of forecasts and minimize bias.

The logit model is asymptotic at zero and one and therefore lends itself to forecasting probabilities. The form of the equation is as follows.

$$P\{y/x\} = \frac{1}{1+e^{-(a+b_1x_1+b_2x_2+\ \ldots)}} \qquad (13)$$

where P{y/x} means the probability of y given x. The shape of the curve is given in Fig. 17, where it is compared to a linear regression model. The mean disadvantage of the method is that the determination of the coefficients is an iterative process and is therefore time consuming.

b) Verification scores

The P-score, P, is given by

$$P = \frac{1}{N} \sum_{i=1}^{r} \sum_{i=1}^{N} (f_{ij} - E_{ij})^2 \qquad (14)$$

where on each of N occasions an event can happen in only one of r possible classes, and f_{i1}, f_{i2}, ..., f_{ir} represent the forecast probabilities that the event will occur in classes 1, 2, ..., r, respectively. The E_{ij} take on the values 0 or 1, respectively, according to whether the event occurred in class j or not.

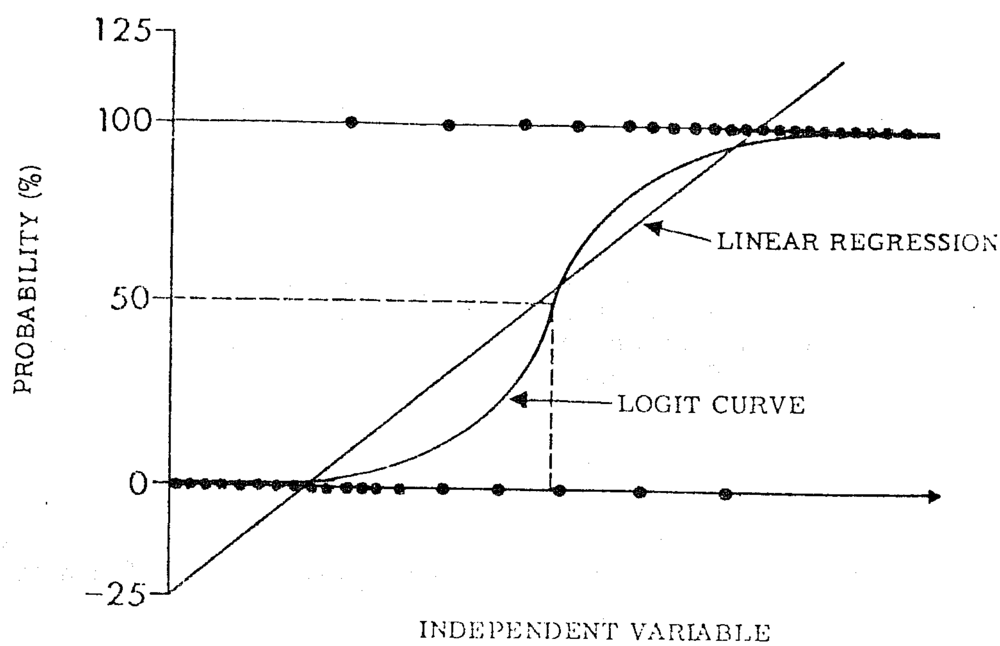

Figure 17. Comparison of the logit and linear regression models.

The threat score, skill score, and bias can be explained by considering the following two-category contingency table:

Observed Category	Forecast Category		
	1	2	Total
1	A	D	G
2	B	E	H
Total	C	F	I

The threat score, TS, of the event comprising category 2 is

$$TS = \frac{E}{F + H - E} \tag{15}$$

The skill score SS is computed from

$$SS = \frac{A + E - J}{I - J} \tag{16}$$

where J, the number of forecasts expected to be correct by chance, is given by

$$J = \frac{(C \times G) + (F \times H)}{I} \tag{17}$$

The bias, BS, of category 2 is

$$BS = \frac{F}{H} \tag{18}$$

References

Bermowitz, R. J., and E. A. Zurndorfer, 1978: Automated guidance for predicting quantitative precipitation. To be published in Mon. Wea. Rev., 106.

Bocchieri, J. R., 1978: A new automated system for forecasting precipitation type. Preprints, Conference on Weather Forecasting and Analysis and Aviation Meteorology, Silver Spring, Md., Oct. 1978, 8 pp.

_____, and H. R. Glahn, 1972: The use of model output statistics for predicting ceiling height. Mon. Wea. Rev., 100, 869-879.

_____, 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.

______, G. M. Carter, R. L. Crisci, D. B. Gilhousen, K. F. Hebenstreit, G. W. Hollenbaugh, and D. J. Vercelli, 1977: Comparative verification of guidance and local aviation/public weather forecasts - No. 3 (Oct. 1976-Mar. 1977). TDL Office Note No. 77-14, National Weather Service, NOAA, 49 pp. (To be published in Mon. Wea. Rev.)

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.

Carter, G. M., 1975: Automated prediction of surface wind from numerical model output. Mon. Wea. Rev., 103, 866-873.

______, and H. R. Glahn, 1976: Objective prediction of cloud amount based on model output statistics. Mon. Wea. Rev., 104, 1565-1572.

Carter, G. M., A. L. Forst, W. H. Klein, and J. P. Dallavalle, 1978: Improved automated forecasts of maximum/minimum and 3-hourly temperatures. Preprints, Conference on Weather Forecasting and Analysis and Aviation Meteorology, October 16-19, 1978, Silver Spring, Md., American Meteorological Society.

Crisci, R. L., 1977: The use of model output statistics for predicting ceiling, visibility, and cloud amount. Technical Procedures Bulletin No. 193, National Weather Service, NOAA, U.S. Department of Commerce, Silver Spring, Md., 15 pp.

Foster, D. S. and R. M. Reap, 1978: Comparative verification of the operational 24-h convective outlooks with the objective severe local storm guidance based on model output statistics. TDL Office Note 78-7, National Weather Service, Silver Spring, Md., 17 pp.

Gerrity, J. F., Jr., 1977: The LFM model - 1976: A documentation. NOAA Technical Memorandum NWS NMC 60, U.S. Department of Commerce, Camp Springs, Md., 68 pp.

Gilhousen, D. B., 1978: The use of model output statistics for predicting probability of precipitation. Technical Procedures Bulletin No. 233, National Weather Service, NOAA, U.S. Department of Commerce, Silver Spring, Md., 13 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.

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., E. A. Zurndorfer, J. R. Bocchieri, G. M. Carter, D. J. Vercelli, D. B. Gilhousen, P. J. Dallavalle, and K. F. Hebenstreit, 1978: The role of statistical weather forecasts in the National Weather Service's operational system. Preprints, Conference on Weather Forecasting and Analysis and Aviation Meteorology, Silver Spring, Md., Oct. 1978, 8 pp.

Harris, R. G., J. G. Bryan, and J. E. MacMonegle, 1963: Terminal weather prediction studies. Technical Note 3, Contract No. AD19(626)-16, The Travelers Research Center, Inc., Hartford, Conn., 103 pp.

Hammons, G. A., J. P. Dallavalle, and W. H. Klein, 1976: Automated temperature guidance based on three-month seasons. Mon. Wea. Rev., 104, 1557-1564.

Howcroft, J., and A. Desmaris, 1971: The limited-area fine mesh (LFM) model. Technical Procedures Bulletin No. 67, National Weather Service, 11 pp.

______, B. M. Lewis, and I. Enger, 1959: Objective prediction of five-day mean temperatures during winter. J. Meteor., 16, 672-682.

Klein, W. H., and H. R. Glahn, 1974: Forecasting local weather by means of model output statistics. Bull. Amer. Meteor. Soc., 55, 1217-1227.

______, and G. A. Hammons, 1975: Maximum/minimum temperature forecasts based on model output statistics. Mon. Wea. Rev., 103, 796-806.

Lowry, D. A., and H. R. Glahn, 1976: An operational model for forecasting probability of precipitation--PEATMOS PoP. Mon. Wea. Rev., 104, 221-232.

Miller, R. G., 1958: A statistical procedure for screening predictors in multiple regression. Final report, Contract No. AF19(604)-1580, The Travelers Weather Research Center, Inc., Hartford, Conn., 238 pp.

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

Pore, N. A., W. S. Richardson, and H. P. Perrotti, 1974: Forecasting extratropical storm surges for the Northeast coast of the United States. NOAA Technical Memorandum, NWS TDL-50, 70 pp.

Reap, R. M., 1972: An operational three-dimensional trajectory model. J. Appl. Meteor., 11, 1193-1202.

Reap, R. M. and D. S. Foster, 1977: Operational thunderstorm and severe local storm probability forecasts based on model output statistics. Preprints, Tenth Conference on Severe Local Storms, Omaha, Neb. Published by Amer. Meteor. Soc., Boston, Mass., pp. 376-381.

Richardson, W. S., and N. A. Pore, 1969: A Lake Erie storm surge forecasting technique. ESSA Technical Memorandum WBTM TDL-24, 23 pp.

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

Zurndorfer, E. A., G. M. Carter, P. J. Dallavalle, D. B. Gilhousen, K. F. Hebenstreit, G. W. Hollenbaugh, J. E. Janowiak, and D. K. Vercelli, 1978: Comparative verification of guidance and local aviation/public weather forecasts - No. 4 (Apr.-Sept. 1977). TDL Office Note 78-3, National Weather Service, NOAA, 49 pp.