The Use of Model Output Statistics (MOS) in Objective Weather Forecasting

HARRY R. GLAHN AND DALE A. LOWRY

Techniques Development Laboratory, National Weather Service, NOAA, Silver Spring, Md. 20910

(Manuscript received 16 March 1972, in revised form 27 July 1972)

ABSTRACT

Model Output Statistics (MOS) is an objective weather forecasting technique which consists of determining a statistical relationship between a predictand and variables forecast by a numerical model at some projection time(s). It is, in effect, the determination of the "weather related" statistics of a numerical model. This technique, together with screening regression, has been applied to the prediction of surface wind, probability of precipitation, maximum temperature, cloud amount, and conditional probability of frozen precipitation. Predictors used include surface observations at initial time and predictions from the Subsynoptic Advection Model (SAM) and the Primitive Equation model used operationally by the National Weather Service. Verification scores have been computed, and, where possible, compared to scores for forecasts from other objective techniques and for the official forecasts. MOS forecasts of surface wind, probability of precipitation, and conditional probability of frozen precipitation are being disseminated by the National Weather Service over teletype and facsimile. It is concluded that MOS is a useful technique in objective weather forecasting.

1. Introduction

Until rather recently, objective forecasting methods[1] have been considered as falling into one of two categories—dynamical and statistical. Now, the relatively new field of stochastic-dynamic prediction is being explored and is beginning to show promise for operational use sometime in the future (Epstein, 1969; Fleming, 1971). However, until stochastic-dynamic prediction is developed much further and more powerful computers are available, we must use some combination of dynamical and statistical methods for practical forecasting.

There has been little success in the prediction of such variables as surface wind, probability and form of precipitation, maximum and minimum temperature, cloudiness, ceiling, and visibility with dynamic models, and indeed, most models do not even forecast these variables directly. There are two general ways in which statistics can be used and the results applied to predictions from numerical models to yield estimates of those elements not successfully forecast directly by the numerical models.

The first, used initially by Klein *et al.* (1959), is usually called the perfect prog method. A concurrent statistical relationship is developed between the variable to be estimated and selected variables which can be forecast by a dynamic model. Both predictand and predictors are observed quantities in the developmental sample. In application, this relationship is applied to numerical model output at, say, a projection of 36 hr to get an estimate of the predictand 36 hr after the data input time for the numerical model.

The other method, which we call Model Output Statistics (MOS), consists of determining a statistical relationship between the predictand and variables from the numerical model at some projection time(s). Application is made in exactly the same way as with the perfect prog method.

The MOS technique is, in effect, the determination of the "weather-related" statistics of a numerical model. For instance, we may want to know what percent of the time rain occurs when the model predicts 80% relative humidity, or, what the best estimate is of the surface wind at an airport when a model predicts a particular 1000-mb geostrophic wind at that point in time and space.

The development and operational use of a Subsynoptic Advection Model (SAM) has been reported by Glahn and Lowry (1972). As an integral part of this same project, the output of SAM and the Primitive Equation (PE) model used by the National Weather Service (Shuman and Hovermale, 1968) has been used to derive regression equations for objectively forecasting surface wind, probability of precipitation (PoP), maximum temperature, cloud amount, and conditional probability of frozen precipitation. Forecasts of several of these weather variables are made operationally twice a day by the National Weather Service (NWS) and

[1] An objective forecasting system has been defined by Allen and Vernon (1951) as "strictly speaking . . . one which can produce one and only one forecast from a specific set of data". It " . . . does not depend for its accuracy upon the forecasting experience or the subjective judgment of the meteorologist using it." Subjective judgment is, of course, used in the *development* of the system.

transmitted over facsimile and teletype for use at field stations.

Input to SAM are the hourly observations made at 0700 and 1900 (all times GMT). These data, as well as the SAM and PE model outputs, have been saved for statistical analysis since April 1967. SAM produces forecasts for the period 0700 to 2400 (and the corresponding period 12 hr later); therefore, the forecasts cover the "first" forecast period, "today" in the case of the 0700 run and "tonight" for the 1900 run. The area for which SAM produces forecasts is generally the United States east of the Mississippi River.

In this paper, the applications of MOS and screening regression to forecasting the surface weather variables named above are presented. Verification scores on independent data are given and, where possible, compared to scores for forecasts from other objective techniques and for the official forecasts.

2. The screening regression procedure

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:

$$\hat{Y}=a_0+a_1X_1+a_2X_2+\cdots+a_kX_k.$$

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.,

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

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

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

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

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

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

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

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

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

Many times it is not known which or how many predictors to include in a regression equation. Even though the predictand may be correlated with hundreds of variables, a regression equation containing only a few of them usually explains nearly as much of the variance as an equation containing many. This is due to the high intercorrelations among the variables. Also, if many predictors are included, the predictand may be estimated extremely well in the dependent data sample, but the equation may be showing not only the real physical relationships but also the chance relationships in the dependent data that will not be present in other samples. Therefore, the equation with many terms may perform more poorly on independent data than the one with fewer terms.

A technique for selecting predictors to include in an equation, called screening (or stepwise) regression, was used in meteorology as early as 1944 by Bryan. Since being popularized by Miller (1958), it has had many applications in meteorology. [For instance, see Klein *et al.* (1959) and Pore (1964); other applications are discussed by Glahn (1965)]. Actually, several variations of this general technique have been used. The one explained below, sometimes called the forward stepwise method, is perhaps the simplest.

The first step in the procedure is to select the variable which correlates most highly (in either a positive or negative sense) with the predictand. This is the variable which explains a greater fraction of the predictand variance than any other of those available. Then, the next variable selected is the one that together with the first increases the reduction of variance the most. Selection can continue in this way until some specified cutoff criterion is met. Usually the cutoff criterion is some function of the additional reduction of variance afforded by the next best predictor. A discussion of the screening technique and the necessary matrix operations is given by Efroymson (1960).

Screening regression, as a mathematical technique, can be used no matter what the joint distribution of the predictand and predictors. (However, this distribution is important in the application of significance tests and the interpretation of results.) In fact, any or all of the variables involved can be binary (i.e., take on only one of two possible values, 0 and 1).

If a predictand can assume only one of two states, it can still be estimated by giving it the value of zero for the first state and one for the second state. The estimate provided by the regression equation can then be considered as the probability of the second state for the particular combination of predictor values on which the estimate was calculated. [See Mook (1948) and Lund (1955) for early uses of this particular procedure.]

If a predictand can assume only one of several, say q, states, it can be transformed into q binary predictands and each treated as discussed above. Miller (1964) used this technique for $q>2$ when all predictors were binary and called it Regression Estimation of Event Probabilities (REEP). Even though the individual estimates are not bounded by zero and one, their sum over all q states is always unity, provided exactly the same predictors are included in each of the q regression equations. Screening algorithms can also be specified for this application of regression.

The equations for estimating probabilities can include continuous as well as binary variables. However, except in special circumstances, the probability estimates may not be as well behaved as when all predictors are binary. It is also worthy of note that minimizing the *RMSE* is the same as minimizing the *P*-Score defined by Brier (1950) and generally used today in PoP verification.

It is usually better to develop an objective procedure for each station separately. However, when the data sample is small, it may be necessary to group several stations together to get a stable system. This is particularly true when the predictand is binary and especially when the climatological probability of the event is far from 0.5. An example of this "generalized operator" approach is contained in Russo *et al.* (1966).

3. Probability of precipitation (PoP)

Perhaps the first major use of MOS was in the estimation of probability of precipitation. For this binary predictand, all stations (about 100) for which data were available were grouped together in the generalized operator concept. Seasonal equations were developed and updated with more data twice a year (summer: April–September; winter: October–March). Predictors were all binary and were derived from PE relative humidity in the layer from the surface to approximately 400 mb, SAM saturation deficit [see Glahn and Lowry (1972) for a definition of saturation deficit], PE precipitation amount, and SAM sea level pressure. Precipitation was defined as the event when observed precipitation was $\geqslant 0.01$ inch.

TABLE 1. Regression equation for forecasting 12-hr PoP (1200–2400) used during the winter of 1968–69. S_d is 3-hr saturation deficit (m), SLP sea level pressure (mb), PE precipitation amount (inches), and relative humidity is in percent. If the variable is $\leqslant$ the indicated value, the contribution to PoP is the amount indicated; otherwise the contribution by that variable is zero. All times GMT.

Predictor	Contribution to PoP (percent)	Cumulative reduction of variance
1. Constant	42.67	
2. SAM $S_d \leqslant 0$ at 1800	10.73	0.3606
3. PE 12-hr precipitation $\leqslant 0.05$ at 2400	− 7.68	0.4334
4. SAM $S_d \leqslant 75$ at 2100	10.19	0.4529
5. SAM $S_d \leqslant -5$ at 1500	12.68	0.4669
6. PE mean relative humidity $\leqslant 70$ at 1800	− 6.33	0.4761
7. PE 12-hr precipitation $\leqslant .20$ at 2400	−12.97	0.4817
8. SAM SLP $\leqslant 1015$ at 1800	6.44	0.4863
9. SAM $S_d \leqslant 45$ at 1500	8.63	0.4889
10. PE mean relative humidity $\leqslant 90$ at 2400	− 7.36	0.4912
11. SAM $S_d \leqslant -15$ at 1500	8.62	0.4925
12. PE 13-hr precipitation = 0 at 2400	− 6.42	0.4937
(Probability range is 2% to 100%)		

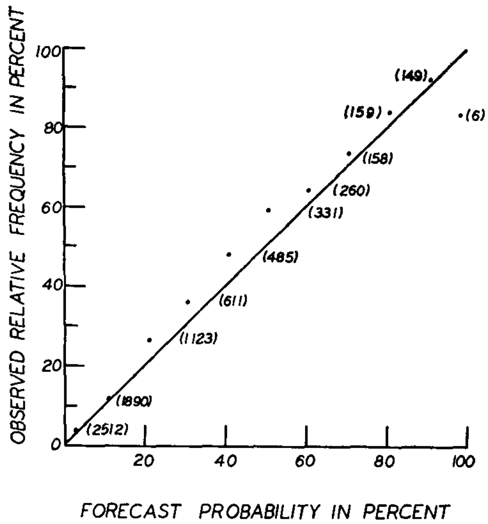

FIG. 1. Reliability of PoP forecasts for the period 1 July 1970 through 31 October 1971. The number of cases represented by each dot is shown in parentheses. The line represents perfect reliability.

Separate equations were developed for the two 6-hr periods 1200–1800 and 1800–2400 and for the 12-hr period 1200–2400 from the 0700 run. Equations for corresponding times were developed from the 1900 run. Since forecasts from the latter equations have not been verified on independent data, and for brevity, only the 0700 run equations will be discussed.

The equation for forecasting 12-hr PoP is given in Table 1. It is typical of all PoP equations. Since each predictor is either zero or one, each corresponding coefficient will contribute zero or its full value. Also, the theoretical minimum and maximum forecasts possible from this equation can be computed and are found to be 2% and 100%, respectively. Generally, the lower limit of all the PoP equations developed was near zero; the upper limit for the 12-hr equations was near 100% and for the 6-hr equations was near 90%.

It is possible that the theoretical limiting values would never occur. That is, the particular combinations of predictors that would give the limiting values do not occur. It is our experience, however, that the full range of theoretically possible PoP values are actually forecast.

One desirable characteristic of probability forecasts is reliability; for all of the forecasts of 20%, say, the relative frequency of the event should be as close to 20% as possible. Fig. 1 shows that the PoP forecasts were slightly low during the July 1970–October 1971 period.

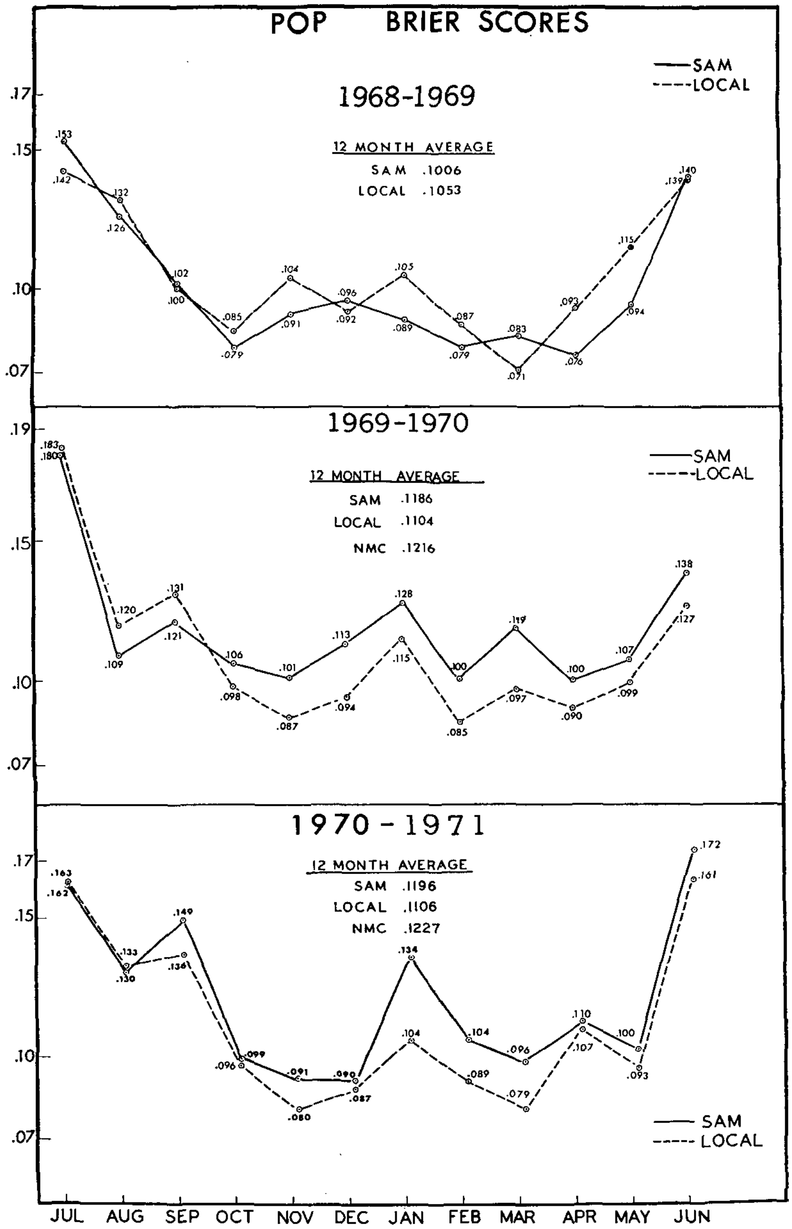

FIG. 2. The Brier score for the objective and local forecasts over a 3-year period. The yearly scores are also shown for the NMC guidance forecasts.

Since SAM remained virtually unchanged during its period of operation, the bias is probably due to changes made in the moisture portion of the PE model. A similar graph covering the period July 1968–June 1969 showed very little bias (see Glahn and Lowry, 1969).

We have compared the MOS forecasts with those made at local stations since July 1968 and with those made subjectively at NMC since July 1969; Fig. 2 gives the monthly and yearly Brier scores.[2] This figure indicates:

1) The MOS forecasts were better than the locals for the first 15 months of comparison. The reverse is true for the last 21 months.
2) The MOS forecasts have been better, on the average, than the NMC guidance for the 24-month period of comparison.

The sudden drop in skill of the MOS forecasts relative to the locals in the fall of 1969 was due in part to a change made in the moisture portion of the PE model in late October. Because of this change, PE predictors were not used from December 1969–August 1970. Since

[2] The Brier Score is defined by the National Weather Service to be one-half the P-Score formulated by Brier (1950).

that time, PE predictors have been included, following another PE model change in September 1970.

The fact that the forecasts made locally are now better than the MOS forecasts is really not surprising. In addition to having 1 or 2 hr later data available, most forecasters also have these MOS forecasts available for guidance. Quite likely, the accuracy of the MOS forecasts has decreased slightly due to PE model changes. It is also likely the locals have improved slightly, although the record we have is not sufficient to establish absolute (rather than relative) skill.

Additional details, too lengthy to be included here, are contained in Glahn and Lowry (1969).

4. Surface wind

Specifying a two-dimensional wind vector presents some interesting problems; these problems and possible solutions are discussed in some detail by Court (1958), Lenhard *et al.* (1963), Lewis (1968), and Glahn (1970a).

One prediction model consists of a separate regression equation for each of the two wind components. Since the mean square error of each component is minimized by the regressions, the mean square vector error is also minimized (Glahn, 1970a). However, the speed of the wind is, in the mean, underforecast by this model (Glahn, 1970a; Barrientos, 1970). If minimum *RMSE* of speed is desired, a separate equation can be derived for speed itself. Then the component estimates can be used for direction and the speed equation for speed.

Minimizing the mean square error of the individual component estimates does not minimize the mean square error of the direction computed from those estimates. Regression estimation of wind direction directly poses a special problem because of the circular nature of the variable. If the predictors include a vector that is rather well related to the predictand, the direction difference between that vector and the predictand can be used to define a new predictand with a scale of -180 to $+180$. The same basic problem exists with this new predictand as with the original with a scale of 0 to 360. However, the new predictand may usually lie in the range -90 to $+90$ and if so, perhaps omitting the few truant cases would produce a good result.

As part of the SAM project, separate regression equations were developed for estimating the U and V wind components and the wind speed valid at 1200 and 1800 for each of 10 stations in the eastern United States. Data were used for the period April–September, 1967 and 1968. The stations were Albany, Atlanta, Baltimore, Cleveland, Cincinnati, Washington, New York, New Orleans, Chicago, and St. Louis. Each sample was of approximately 200 cases in size. The equations for the U and V components were developed by forcing the computer program to select as the first two predictors the 1000-mb geostrophic wind components forecast by SAM valid at the same time as the predictand and then screening several other variables.

The equations for U and V valid at 1200 selected for testing usually, but not always, contained one or both of the observed 0700 wind components in addition to the 1000-mb winds valid at 1200; those used for testing valid at 1800 contained a selection of 1000-mb winds at 1200, 500-mb winds at 1200 and 1800, and 1000-mb temperatures at 1200 and 1800, in addition to the 1000-mb winds valid at 1800. The 1200 equations contained from two to four predictors; the 1800 equations contained from two to five predictors. The decision of which equation to test (how many predictors to include) was made subjectively.

The equations for estimating speed directly were derived by forcing the 1000-mb geostrophic wind speed forecast by SAM valid at the same time as the predictand and then screening several other variables. The equations used for testing contained from four to six predictors similar to those in the U and V equations, the only significant difference being that wind speeds were used as predictors, whereas in the U and V equations only wind components were used as predictors.

Sample equations for St. Louis are shown below:

$$\hat{U}^{12}=0.482+0.185U_0^{12}-0.333V_0^{12}+0.276V^{07}$$

$$\hat{V}^{12}=0.194+0.164U_0^{12}+0.175V_0^{12}-0.005U^{07}+0.170V^{07}$$

$$\hat{S}^{12}=1.576+0.239S_0^{12}+0.175S^{07}-0.040V_0^{18}+0.027U_0^{12}$$

where U, V and S are the U- and V-wind components, and the wind speed, respectively (knots); the subscript 0 indicates 1000-mb geostrophic values predicted by SAM; and the superscript indicates the valid time in GMT.

Thus, the estimate of wind speed at 1200 ($\hat{S}^{12}$) at St. Louis depends on the 1200 GMT 1000-mb geostrophic wind speed forecast by SAM (S_0^{12}), the observed 0700 surface wind speed at St. Louis (S^{07}), the 1800 GMT 1000-mb V-wind component forecast by SAM (V_0^{18}), and the 1200 GMT 1000-mb U-wind component forecast by SAM (U_0^{12}).

The equations for the 10 stations were evaluated for each day in April and May 1969 for which SAM data were available. The wind forecasts in the aviation bulletins (FT's) made at NWS offices were used for comparison. Since the FT's do not mention wind if the speed is expected to be less than 10 kt, the comparison was made in two ways.

For all those cases where the FT's included wind, and objective forecasts were available, the *RMSE* of direction (computed from the U and V equations) and speed (direct from the speed equation) and the bias (mean forecast minus mean observed) of speed (both direct from the speed equation and calculated from the U and V equations) were computed. Also, for all cases when the FT's and objective forecasts were available, contingency tables for speed were prepared by considering the FT forecast of wind to be under 10 kt when wind was not mentioned. From these contingency tables, which had categories <10, 10–12, 13–17, 18–22 and

VALID TIME (GMT)	PROJECTION (HR)	FORECAST	DIRECTION RMSE (DEG)	SPEED (kts)					
				RMSE	SKILL SCORE	PERCENT CORRECT	MEAN FORECAST	MEAN OBSERVED	BIAS
12	5	OBJECTIVE U,V EQUATIONS	35				9.4	10.4	-1.0
	5	OBJECTIVE SPEED EQUATION		3.5	.37	76	9.8		-.6
	3	FT	33	3.6	.36	71	12.0		1.6
18	11	OBJECTIVE U,V EQUATIONS	47				9.2	11.1	-1.9
	11	OBJECTIVE SPEED EQUATION		3.5	.29	54	11.3		.2
	9	FT	50	4.3	.24	49	12.6		1.5

FIG. 3. Comparison of official FT and objective wind forecasts for 10 stations in the eastern United States for April and May 1969. The number of cases used in calculating each statistic varies from 166 to 545.

>22 kt, skill scores and percent correct were computed. These scores are shown in Fig. 3.

Fig. 3 indicates that the directions from the objective forecasts were as good as those from the FT's and that the speeds from the objective were better than those from the FT's. The projections of the objective forecasts (5 and 11 hr) refer to the latest data used (0700). Actually, the forecasts could be available to the field forecasters before 0900. The FT's were prepared with 0900 and perhaps 1000 data, if available; transmission time for the forecasts is 1045. The bias in the speed computed from the U and V equations is noticeable.

Although the verification on independent data is not prodigious, it is sufficient to demonstrate the usefulness of this technique. Good results have also been obtained by Barrientos (1970). Alternative models for wind prediction are given by Glahn (1970a).

5. Maximum temperature

MOS max temperature equations have been derived for each of 16 stations every six months beginning April 1969 for the forthcoming summer or winter season. Each new set of summer (winter) equations was developed on a dependent data sample which included the available data from all the previous summers (winters) combined.

The predictand was the daily (midnight to midnight) observed max temperature. The predictors included forecasts from the SAM and PE model valid between 1200 and 2400 and initial 0700 observations. The predictand lead time was such that a forecast of max temperature for "today" could be made at about 0900 (0400 EST).

Initial data included dew point, weather, cloud amount, temperature, and surface wind components. Variables from the PE model included the following at 6-hr intervals: column mean relative humidity, 1000-mb temperature, precipitation amount, and 500-mb height. Forecast variables from SAM were the 3-hr saturation deficit, sea level pressure, and 1000-mb geostrophic wind components.

The variables selected most often by the screening are, in order: PE 1000-mb temperature, initial surface temperature, initial cloud amount, PE 500-mb height, and SAM saturation deficit. Since the equation for each station is unique, the order of these predictors and their respective contributions to the total reduction of variance differs for each station. A typical equation is shown in Table 2.

When equations developed on a seasonal basis are applied to independent data, there should be little overall bias in the forecasts. However, there may be a monthly bias even in the dependent data. The bias computed for each month during the winter season of 1970–71 revealed that there was indeed a monthly bias; as indicated in Fig. 4, the forecasts were slightly too low at the beginning (October) and end (March) of the period and too high during the intervening months. In an attempt to correct this periodic bias, the sine and cosine of the day of the year were included as possible predictors.

Each of the 16 single station equations which were rederived contained the cosine of the day of the year and 4 equations also contained the sine. Verification on winter 1970–71 data showed that the periodic bias was indeed reduced but not eliminated; also, the mean absolute error (MAE) was reduced by $0.29F$ (7%).

By verifying forecasts based on equations with 2, 4, 6, 8 and 10 predictors, for each of two seasons, it was determined that 10-predictor equations were better than those with a lesser number of predictors. This may be a little surprising since 85 predictors were screened and the sample size was only about 500. For the summer of 1971, the MAE from the 10-predictor equation was 0.04F (1%) less than the MAE from the 8-predictor equation.

The perfect prog technique has been used with much success by Klein and Lewis (1970) for maximum temperature prediction, and the NWS distributes forecasts made by this method for guidance at local offices. Forecasts from the MOS system, Klein-Lewis system, and

TABLE 2. Equation for estimating maximum temperature (°F) at Washington, D. C., in winter.

Predictor	Coefficient	Cumulative reduction of variance
Constant	21.62	
1000-mb temperature (°C) at 1200	0.310	0.742
Mean relative humidity (%) at 1200	−0.032	0.796
Observed surface temperature (°F) at 0700	0.570	0.829
Observed cloud cover at 0700*	−1.093	0.855
1000-mb U-wind component (kt) at 1200	0.084	0.865
Saturation deficit (m) at 1500	0.014	0.873
1000-mb temperature (°C) at 1800	0.488	0.876
1000-mb V-wind component (kt) at 2400	−0.044	0.880
Observed weather at 0700**	2.776	0.883
Precip amt $\leqslant 0.10$ inch at 2400 (binary)	1.999	0.885

* Observed total cloud cover in coded form: 0 = none or partial obscuration, 1 = scattered, 2 = broken, 3 = overcast, 4 = obscured.

** Observed weather at 0700 in coded form: 0 = none of the following, 1 = frozen precipitation, 2 = drizzle or freezing drizzle, 3 = rain or freezing rain.

1969	APR	MAY	JUNE	JULY	AUG	SEPT	OCT	NOV	DEC
MEAN ABSOLUTE ERROR (°F)									
MOS	3.9	3.2	3.6	3.2	2.8	2.9	3.8	3.9	4.9
K-L	4.4	3.7	3.8	2.7	3.3	3.3	4.1	3.7	5.4
LOCAL	3.5	2.8	2.8	2.4	2.1	2.4	3.5	3.3	3.5
BIAS (°F)									
MOS	-.4	-.4	-1.4	-1.4	-1.3	.2	-1.6	1.1	3.5
K-L	.8	-.1	.0	.2	-.1	1.5	1.7	-.8	3.7
LOCAL	.2	.3	.3	.8	.5	.6	.3	-.2	1.2
ABSOLUTE ERROR ≥ 10 °F (NO.)									
MOS	NA	NA	NA	NA	NA	NA	NA	NA	NA
K-L	NA	NA	NA	NA	NA	NA	NA	NA	NA
LOCAL	NA	NA	NA	NA	NA	NA	NA	NA	NA
	MOS EQUATIONS WITH 6 PREDICTORS						MOS EQUATIONS WITH 10 PREDICTORS		
	17 STATIONS						16 STATIONS		
	KLEIN-LEWIS (BAROTROPIC)								

1970	JAN	FEB	MAR	APR	MAY	JUNE	JULY	AUG	SEPT	OCT	NOV	DEC
MEAN ABSOLUTE ERROR (°F)												
MOS	4.9	4.5	4.1	3.9	3.9	3.7	4.0	3.5	3.4	3.2	3.6	3.9
K-L	5.1	5.1	5.3	4.2	3.7	2.8	2.9	2.2	2.8	3.9	3.6	4.2
LOCAL	3.5	3.0	3.7	3.8	2.9	2.6	2.5	2.0	2.3	3.5	3.0	3.2
BIAS (°F)												
MOS	2.9	1.2	-.8	-.6	-1.9	-2.6	-3.2	-2.8	-2.2	-.4	.3	1.0
K-L	2.6	1.0	1.4	-.7	1.0	.4	.4	.3	.2	2.4	.5	1.8
LOCAL	.6	-.7	-.4	.1	.8	.3	.2	.3	.3	1.9	.8	1.0
ABSOLUTE ERROR ≥ 10 °F (NO.)												
MOS	NA	33	35	18	16	10	18	8	5	3	19	20
K-L	NA	48	74	30	30	10	12	3	2	8	14	27
LOCAL	NA	9	28	24	13	10	7	4	2	8	9	8
	MOS EQUATIONS WITH 10 PREDICTORS									MOS EQUATIONS WITH 10 PREDICTORS (SINE & COSINE)		
	16 STATIONS											
	KLEIN-LEWIS (BAROTROPIC)			KLEIN-LEWIS (PE)								

1971	JAN	FEB	MAR	APR	MAY	JUNE	JULY	AUG	SEPT
MEAN ABSOLUTE ERROR (°F)									
MOS	4.6	4.5	4.0	4.4	3.5	2.6	2.5	2.7	3.1
K-L	4.7	4.5	4.2	4.5	3.7	2.6	2.4	2.4	3.0
LOCAL	3.3	3.2	3.3	3.8	3.2	2.3	2.3	2.3	2.6
BIAS (°F)									
MOS	2.0	1.6	.9	.0	-1.0	-1.2	-.8	-1.4	-.6
K-L	2.7	.8	1.3	-.0	.6	.8	.5	.6	.4
LOCAL	.9	.4	.3	-.5	.3	.8	.8	.8	.3
ABSOLUTE ERROR ≥ 10 °F (NO.)									
MOS	40	39	32	30	18	2	6	7	8
K-L	50	32	30	32	19	9	9	7	12
LOCAL	17	16	11	19	10	3	11	7	6
	MOS EQUATIONS WITH 10 PREDICTORS (SINE & COSINE)								
	16 STATIONS								
	KLEIN-LEWIS (PE)								

FIG. 4. Monthly verification scores for forecasts of today's maximum temperature at 16 eastern United States stations. The number of large errors was not computed prior to February 1970. K-L means the Klein-Lewis operational perfect prog system.

local offices have been compared over the 30-month period April 1969–September 1971. Verification statistics are shown in Fig. 4.

From April–September 1969 only 6-predictor MOS equations were used. During this period, the MOS forecasts had about 0.5F less *MAE* than the Klein-Lewis forecasts, and the local forecasts were about 0.5F better than the MOS forecasts. For the following six-month period (October 1969–March 1970) the *MAE* of the MOS forecasts continued to be less than the *MAE* of the Klein-Lewis forecasts and about 1.0F greater than the locals.

In April 1970, the Klein-Lewis system, which had been based on a barotropic numerical model, was replaced by a system based on the PE model (see Klein *et al.*, 1971). Fig. 4 indicates that the *MAE* of the Klein-Lewis system was less than the *MAE* for MOS for the period April–September 1970. During the winter of 1970–71 MOS, with the sine and cosine terms, was better than Klein-Lewis; during the summer of 1971 they had about equal *MAE*. The locals continued to be about 0.5F better in the summer and 1.0F better in the winter than the two objective systems.

Fig. 4 also shows that for the period February–September 1971 the number of local forecasts with errors ⩾10F was only 60% that of the MOS system. The number of Klein-Lewis large errors was 25% greater than for MOS. Additional details are given by Annett *et al.* (1972).

6. Cloud amount

Regression equations have been developed for estimating cloud amount at each of four stations for the times 1200, 1500, 2100 and 2400. The predictand is a coded variable such that 0 = clear, 1 = partial obscuration, 2 = thin scattered, 3 = scattered, 4 = thin broken, 5 = broken, 6 = thin overcast, 7 = overcast, and 8 = obscured. The possible predictors included observed weather and cloud amount at 0700, PE layer relative humidity and precipitation amount, and SAM 1000-mb wind components and saturation deficit. A sample prediction equation is shown in Table 3; it was developed on a sample of 298 cases from the winters of 1967–68 and 1968–69.

It is somewhat surprising that the relative humidity and saturation deficit predictors chosen by screening were valid at 2400 rather than 1200 and 1500, since the predictand was observed at 1200. This time shift, which was also noted for the other three stations St. Louis, Mo., New York, N. Y., and Atlanta, Ga., suggests a tendency on the part of each numerical model to be slow in the movement of its moisture variable.

The cloud forecasts have not been verified formally on independent data. However, forecasts have been computed daily for about two years and appear to be very good.

7. Conditional probability of frozen precipitation [PoFP(P)]

The predictand in this case is the probability of frozen precipitation given that precipitation occurs. To estimate this, we used a subsample of our total data con-

TABLE 3. Equation for estimating cloud amount at Washington, D. C., in winter.

Predictor	Coefficient	Cumulative reduction of variance
Constant	3.115	
Observed cloud cover at 0700*	0.775	0.403
Mean relative humidity (%) at 2400	0.013	0.503
Saturation deficit (m) at 2400	−0.006	0.525
1000-mb U-wind component (kt) at 2100	−0.022	0.543
1000-mb V-wind component (kt) at 1200	0.040	0.556
1000-mb U-wind component (kt) at 1800	−0.029	0.571

* Observed total cloud cover in coded form: 0 = none or partial obscuration, 1 = Scattered, 2 = broken, 3 = overcast, 4 = obscured.

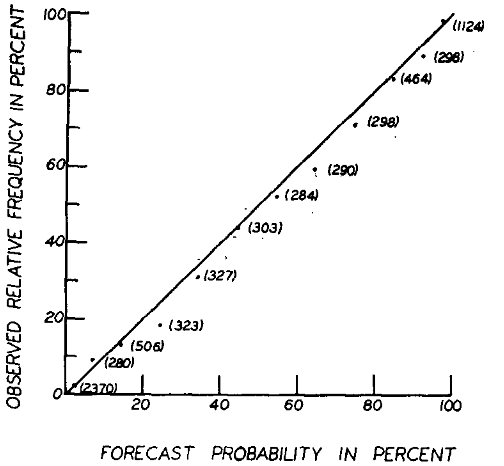

FIG. 5. The reliability of the PoFP(P) forecasts for the period 1 October 1969 through 15 May 1970. The number of cases represented by each dot is shown in parentheses. The line represents perfect reliability.

```
ZCZC
FOUS4 KWBC 170800
SAM FORECASTS
     12Z      15Z        18Z        21Z        00Z       POP 12 6 6  POFP B E
CAR  2408  176 2110   -12 2112   -32 2012   -42 1809   033 06 33   100 100
BTV  1810  -21 1813   -09 1913   -16 2011   124 2012   068 48 34   100 092
PWM  2205  138 2109   095 2012   -06 2110   -07 2108   025 07 25   094 083
BOS  2513  159 2114   191 2216   250 2316   245 2215   009 02 08   093 063
PVD  2307  179 2211   240 2313   282 2314   279 2210   002 01 02   092 047
BDL  2005  170 2209   258 2412   278 2413   257 2212   010 04 09   091 049
LGA  2312  294 2414   309 2515   326 2415   307 2415   002 01 02   088 037
ALB  1607  166 1912   236 2115   217 2214   201 2111   019 10 11   098 067
BGM  2011  255 2116   250 2218   233 2117   214 2216   023 16 09   097 052
```

FIG. 6. A portion of a typical SAM teletype bulletin transmitted 17 January 1971. For each station, the regression estimation of surface wind (4 digits) valid at 12Z (1200 GMT) is given. Then, for each of four valid times, the saturation deficit (three digits) and regression estimation of surface wind (four digits) is given. Next, the 12-hr PoP (three digits) and two 6-hr PoP's (two digits each) are given which cover the 1200–0000 period. At the far right are the conditional probabilities of frozen precipitation valid at 1200 (beginning of period) and 0000 (end of period).

sisting of only those cases when precipitation did occur at the predictand valid time. The main predictor was the "Wagner index" which was computed according to the work of Wagner (1957). This predictor is itself a conditional probability of frozen precipitation determined from the 1000–500 mb thickness. The other possible predictors screened were binary and were derived from the PE 1000-mb temperature (T_0). The generalized operator concept was used; data from nearly 100 stations were combined to derive the regression equations. However, the Wagner index carries with it pertinent climatological information. The equation derived on two winter seasons of data which produces forecasts valid at the end of the "today" period (2400) is shown in Table 4. This equation will produce values in the range -4 to 104 inclusive.

The reliability of PoFP(P) forecasts for the 1969–70 winter is given in Fig. 5. This figure shows a slight bias toward high forecast probabilities. The bias is probably due to changes that have been made in the initialization of the PE model affecting the 1000-mb temperature. A similar graph prepared before the changes were made did not show this bias. The fact that changes in the numerical model can cause problems in the objective interpretation of the model forecasts is, of course, one of the major problems with the MOS approach. However, the same problem, perhaps to a lesser degree, also exists with the subjective interpretation of numerical model forecasts.

TABLE 4. Equation for estimating conditional probability of frozen precipitation (percent) at 2400 in the eastern United States. PE T_0 indicates 1000-mb temperature forecast by primitive equation model.

Predictor	Coefficient	Cumulative reduction of variance
Constant	-3.91	
Wagner index at 0000	73.04	0.664
PE T_0 at 0000 $\leqslant$ 3C (binary)	9.63	0.690
PE T_0 at 1200 $\leqslant$ 1C (binary)	10.78	0.699
PE T_0 at 0000 $\leqslant$ 8C (binary)	7.27	0.700
PE T_0 at 0000 $\leqslant -2$C (binary)	6.85	0.701

8. Operational transmissions

SAM forecasts and some of the derived products discussed above are transmitted twice daily over teletype

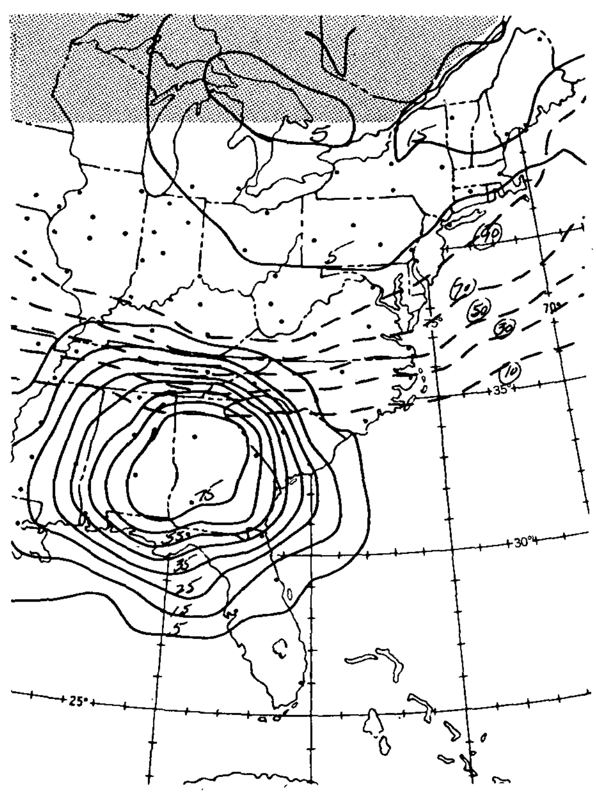

FIG. 7. A portion of a typical SAM facsimile chart. It shows the 6-hr probability of precipitation (solid lines) for the period 1200 to 1800 and the conditional probability of frozen precipitation (dashed lines) valid at 1200 GMT 29 January 1970.

and facsimile by the NWS. Samples of these transmissions are shown in Figs. 6 and 7. These figures indicate that the probability forecasts can vary rapidly in space and time, yet give consistent patterns. For instance, the probability of precipitation changes from 75% to 5% over a distance of about 150 mi.

9. Summary and conclusions

Our work with the MOS technique has shown it to be very useful in forecasting surface weather variables. This is especially true when probabilities are being forecast, since the errors of the numerical models are considered in determining the forecast equations.

MOS has been used successfully for probability of precipitation, surface wind, conditional probability of frozen precipitation, maximum temperature, and cloud amount. Forecasts for the first three of these from the MOS system are being transmitted twice daily by the NWS over teletype and facsimile. They are all being used daily in experimental, computer-produced worded forecasts (Glahn, 1970b).

Progress in objective weather forecasting within the next few years will come through the combining of numerical and statistical models. Due to the development of new, and the modification of old, numerical models, data samples containing numerical model output are a perishable commodity. Therefore, considerable prior planning and organization will be necessary in the operational implementation of MOS products. This is one of the main concerns of the Techniques Development Laboratory.

Acknowledgments. The authors were assisted in this work by many persons in the Techniques Development Laboratory. We wish to especially thank Mr. George Hollenbaugh, Lt. (jg) John Annett, Mrs. Evelyn Boston, Mrs. Elizabeth Booth, Mrs. Jackie Hughes, Mr. Dean Costantinou, Mr. Eugene Legg, Mr. Peter Howe, and Miss Judy Pegues.

REFERENCES

Allen, R. A., and E. M. Vernon, 1951: Objective weather forecasting. *Compendium of Meteorology*, T. F. Malone, Ed., Boston, Mass., Amer. Meteor. Soc., 796–801.

Annett, J. R., H. R. Glahn and D. A. Lowry, 1972: The use of model output statistics (MOS) to estimate daily maximum temperatures. NOAA Tech. Memo. NWS TDL 45, 14 pp.

Barrientos, C. S., 1970: An objective method for forecasting winds over Lake Erie and Lake Ontario. ESSA Tech. Memo. WBTM TDL 34, 20 pp.

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

Bryan, J. G., 1944: Special techniques in multiple regression (unpublished manuscript).

Court, A., 1958: Wind correlation and regression. Sci. Rept. No. 3, Contract AF19(604)-2060, Cooperative Research Foundation, San Francisco, Calif., 16 pp.

Efroymson, M. A., 1960: Multiple regression analysis. *Mathematical Methods for Digital Computers*, A. Ralston and H. S. Wilf, Eds., New York, Wiley, 191–203.

Epstein, E. S., 1969: Stochastic dynamic prediction. *Tellus*, **21**, 739–757.

Fleming, R. J., 1971: On stochastic dynamic prediction. *Mon. Wea. Rev.*, **99**, 851–872.

Glahn, H. R., 1965: Objective weather forecasting by statistical methods. *The Statistician*, **15**, No. 5, 111–142.

——, 1970a: A method for predicting surface winds. ESSA Tech. Memo. WBTM TDL 29, 18 pp.

——, 1970b: Computer-produced worded forecasts. *Bull. Amer. Meteor. Soc.*, **51**, 1126–1131.

——, and D. A. Lowry, 1969: An operational method for objectively forecasting probability of precipitation. ESSA Tech. Memo. WBTM TDL 27, 24 pp.

——, and ——, 1972: An operational subsynoptic advection model. *J. Appl. Meteor.*, **11**, 578–585.

Klein, W. H., and F. Lewis, 1970: Computer forecasts of maximum and minimum temperature. *J. Appl. Meteor.*, **9**, 350–359.

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

——, F. Lewis and G. A. Hammons, 1971: Recent developments in automated max/min temperature forecasting. *J. Appl. Meteor.*, **10**, 916–920.

Lenhard, R. W., Jr., A. Court and H. Salmela, 1963: Reply. *J. Appl. Meteor.*, **2**, 812–815.

Lewis, F., 1968: Regression of complex variables. *Preprints, First Statistical Meteor. Conf.*, Amer. Meteor. Soc., 83–88.

Lund, I. A., 1955: Estimating the probability of a future event from dichotomously classified predictors. *Bull. Amer. Meteor. Soc.*, **36**, 325–328.

Miller, R. G., 1958: The screening procedure. Studies in Statistical Weather Prediction, Final Rept., Contract AF19(604)-1590 (B. Shorr, Ed.) The Travelers Research Center, Inc., Hartford, Conn., 86–95.

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

Mook, C. P., 1948: An objective method of forecasting thunderstorms for Washington, D. C. in May (unpublished manuscript).

Pore, N. A., 1964: The relation of wind and pressure to extratropical storm surges at Atlantic City. *J. Appl. Meteor.*, **3**, 155–163.

Russo, J. A., I. Enger and G. T. Merriman, 1966: A statistical approach to the 12–48-hr prediction of precipitation probability. Final Rept., Contract Cwb-11100, The Travelers Research Center, Inc., Hartford, Conn. 107 pp.

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

Wagner, J. A., 1957: Mean temperature from 1000 mb to 500 mb as a predictor of precipitation type. Sci. Rept. No. 2, Contract AF19(604)-1305, Dept. of Meteorology, M.I.T.