An Application of Model Output Statistics to Forecasting Quantitative Precipitation

ROBERT J. BERMOWITZ

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

(Manuscript received 5 August 1974)

ABSTRACT

The Model Output Statistics (MOS) technique has been applied to the prediction of quantitative precipitation. Data at 233 stations for two winter seasons are pooled to develop generalized operator equations for prediction of the probability of precipitation amount (PoPA) in 5 categories for a 12–24 hr forecast projection. Predictors subjected to screening regression are obtained from the National Meteorological Center's (NMC) Primitive Equation Model and the Techniques Development Laboratory's Trajectory Model. The equations are developed by means of two approaches. The first, referred to as unconditional, uses both precipitation and no precipitation cases in the developmental sample. The second, referred to as conditional, uses only precipitation cases.

To test the system on independent data, forecasts at 58 cities in the United States are verified for October 1972. Probability forecasts are transformed to categorical forecasts by (1) maximizing the percent correct, (2) maximizing a quantitative precipitation forecast (QPF) score used to verify categorical forecasts at NMC, (3) maximizing a utility score, and (4) minimizing the categorical bias. The categorical forecasts are then compared to those obtained from (1) the Limited Area Fine Mesh Model, (2) subjective preparation at NMC, and (3) climatology. Verification scores include the percent correct, a QPF score used at NMC, and a utility score. In addition, the categorical bias is computed for each forecast method. The results indicate that the PoPA categorical forecasts obtained by minimizing the bias are, in general, slightly better than all the other forecasts. Also, "unconditional" forecasts are about as good as "conditional" forecasts.

1. Introduction

Knowledge of quantitative precipitation is important for hydrological purposes. Excessive precipitation amounts over relatively short time periods can be of vital importance for the public safety. Unfortunately, accurate prediction of quantitative precipitation, especially the larger amounts, is difficult and remains, for the most part, an unsolved meteorological problem.

Statistical attempts at solving the quantitative precipitation forecast (QPF) problem have included the "perfect prog" approach (e.g., Klein, 1968) and synoptic climatology (e.g., Spiegler and Fisher, 1971; Korte *et al.*, 1972). Recently, Model Output Statistics (MOS) (Glahn and Lowry, 1972) has been used with considerable success for prediction of a number of weather elements including probability of precipitation (PoP) (Glahn and Lowry, 1969), conditional probability of frozen precipitation (Glahn and Lowry, 1972), surface wind (Glahn, 1970; Barrientos, 1970; Carter, 1973), ceiling and visibility (Bocchieri and Glahn, 1972), and maximum temperature (Annett *et al.*, 1972; Klein and Hammons, 1973). Wasserman (1972) has used objectively produced PoP forecasts prepared at the National Meteorological Center (NMC) in a MOS type approach to predict the frequency of occurrence of precipitation amounts exceeding specified values; his study was for 13 cities over the eastern United States.

The purpose of this study was to apply the MOS technique in the development of an objective method for producing probabilistic and categorical forecasts of precipitation amount for 233 cities over the conterminous U. S. for the period 12–24 hr after 0000 GMT.

2. Development of equations

The equations for estimating probability of precipitation amount (PoPA) were developed for five categories of quantitative precipitation. These categories, defined in Table 1, are the same ones used in preparing subjective forecasts at NMC. The developmental sample consisted of data at 233 stations for the cold seasons of 1970–1971 and 1971–1972. The cold season is defined as the period October–March.

Table 1 also contains the relative frequency of occurrence of each category in the dependent data sample. Note the low frequencies of occurrence for categories 2–5. This illustrates the difficulty of the QPF problem; prediction of the more important higher categories reduces to prediction of the rare event. In order to increase the size of the developmental data sample, data from all 233 stations were pooled and a generalized

TABLE 1. Five predictand categories of quantitative precipitation and relative frequency of 1200–2400 GMT amounts based on the developmental sample of 70,131 cases.

Category	Precipitation amount (in)	Relative frequency
1	≤0.24	0.951
2	0.25–0.49	0.028
3	0.50–0.99	0.016
4	1.00–1.99	0.004
5	≥2.0	0.001

operator equation, applicable at all stations, was developed. The equation was developed by means of two approaches. The first, referred to as unconditional, used both precipitation and no precipitation cases in the developmental sample. The second, referred to as conditional, used only precipitation cases.

Potential predictors subjected to screening regression were subjectively chosen from the output of two numerical models run operationally at NMC—the Primitive Equation (PE) (Shuman and Hovermale, 1968) and Trajectory (TJ) (Reap, 1972) models. Predictors, all binary variables, could also be chosen from forecast fields smoothed by taking simple 5-, 9-, and 25-point averages. In this way some attempt was made to take into account inaccuracies in the prognostic fields. Output from the PE model was available for projections beyond 24 hr while output from the TJ model was available only at 24 hr. Predictors included forecasts of precipitation amount, precipitable water, relative humidity in layers; vertical velocities, net vertical displacements, heights of constant pressure surfaces, and boundary layer u and v wind components.

Equations, one for each of the predictand categories and each containing the same 12 predictors, were developed. The choice of the number of predictors was based on experience in developing automated MOS PoP forecasts at the Techniques Development Laboratory. Of interest, is that of the 12 selected predictors, 8 were either forecasts of precipitation amount or precipitable water. Forecast probabilities obtained from the equations for categories 2–5, especially those of categories 4 and 5, are rather low. Indeed, the maximum forecast probability of category 4 is 15.9% and that of category 5 is only 3.0%.

3. Verification

In order to test the PoPA system, 12–24 hr forecasts at 58 U. S. cities were verified on one month, October 1972, of independent data. In addition, 12–24 hr forecasts at the same 58 cities were verified on a subset of the dependent data consisting of 78 days during the period January–March 1972. The 58 cities, uniformly distributed over the U. S., are the same ones NMC has used to verify categorical forecasts of precipitation amount.

In order to perform the verification, probability forecasts were transformed to categorical forecasts by (1) maximizing the percent correct, (2) maximizing a QPF score that has been used at NMC and to be called the NMC QPF score in this paper (actually, the score used is 4 minus the NMC QPF score), (3) maximizing a utility score, and (4) minimizing the bias.[1] A discussion of maximizing a score is given by Bocchieri and Glahn (1972).

The matrices used to transform the probability forecasts by maximizing the three verification scores are shown in Table 2. They are, of course, the scoring matrices used to compute the respective verification scores. When the matrices are used for verification, the

TABLE 2. Matrices for maximizing percent correct (A), 4-NMC QPF score (B), and utility score (C), and for minimizing the bias (D).

(A)

Observed category	Forecast category 1	2	3	4	5
1	1	0	0	0	0
2	0	1	0	0	0
3	0	0	1	0	0
4	0	0	0	1	0
5	0	0	0	0	1

(B)

Observed category	Forecast category 1	2	3	4	5
1	4	3	2	1	0
2	3	4	3	2	1
3	2	3	4	3	2
4	1	2	3	4	3
5	0	1	2	3	4

(C)

Observed category	Forecast category 1	2	3	4	5
1	0.5	0.2	0.0	0.0	0.0
2	0.2	0.6	0.3	0.0	0.0
3	0.0	0.2	0.7	0.4	0.2
4	0.0	0.0	0.3	0.8	0.6
5	0.0	0.0	0.0	0.4	1.0

(D)

Observed category	Forecast category 1	2	3	4	5
1	0.08	0.05	0.0	0.0	0.0
2	0.03	0.14	0.10	0.0	0.0
3	0.0	0.05	0.25	0.17	0.13
4	0.0	0.0	0.11	0.47	0.38
5	0.0	0.0	0.0	0.30	1.0

---

[1] Bias is the number of forecasts of a category divided by the number of observations of that category. A categorical bias equal to 1 means unbiased forecasts of that category.

TABLE 3. Comparative verification on independent data of unconditional (U) and conditional (C) PoPA forecasts transformed to categorical forecasts by maximizing (1) percent correct, (2) NMC QPF score, and (3) utility score, and (4) minimizing the bias.

Verification score	PoPA forecasts transformed by							
	Maximizing						Minimizing bias	
	Percent correct		NMC QPF score		Utility score			
	U	C	U	C	U	C	U	C
Percent correct	95.2	95.1	94.6	94.4	94.9	94.8	93.0	93.3
NMC QPF score	3.9161	3.9138	3.9149	3.9132	3.9126	3.9109	3.8885	3.8862
Utility score	0.4814	0.4805	0.4805	0.4792	0.4804	0.4796	0.4798	0.4764

percent correct gives the same credit for correct forecasts of different categories but no credit for misses. The NMC QPF score also gives the same credit for forecasts of different categories; however, partial credit is given for misses. The utility score gives more credit for correct forecasts of higher categories than correct forecasts of lower categories. Also, credit is given for near misses, with more credit given for near misses of higher categories than lower categories.

The method for minimizing the bias is the same as that for maximizing a score, except that a minimum bias matrix is used instead of a scoring matrix. The transformation matrix for doing this was developed on the developmental data sample and is shown in Table 2(D). The technique for developing a minimum bias matrix is discussed by Bocchieri *et al.* (1973).

PoPA categorical forecasts were compared to categorical forecasts obtained from (1) the Limited Area Fine Mesh (LFM) model (Howcroft and Desmaris, 1971), (2) subjective preparation at NMC, and (3) climatology. Climatology is defined as a category 1 forecast. Verification scores included the percent correct, the NMC QPF score, and the utility score. Categorical bias was also computed.

A comparison was also made between unconditional and conditional PoPA categorical forecasts. Conditional probability forecasts of each category were transformed to unconditional forecasts by multiplying the conditional probability of each category by the probability of precipitation. The probability of no precipitation was then added to the probability of category 1 just obtained, since this category must now contain no precipitation cases as well as precipitation cases ≤ 0.24 in. The quality of the resulting unconditional probability forecasts is, of course, a function of the quality of the PoP forecasts. The latter were obtained from PoP equations that were in operation at NMC during the winter season of 1972–1973 (Lowry *et al.*, 1973).

4. Results

Table 3 summarizes the comparison between the unconditional and conditional PoPA categorical forecasts. It can be seen that the unconditional categorical forecasts were about as good as the conditional categorical forecasts. Further evidence of this result is contained in the Brier score (Brier, 1950). Unconditional probability forecasts had a Brier score of 0.0408 which compares to Brier score of 0.0410 for the conditional probability forecasts. Results of the verification on dependent data are not shown since they are similar to the independent data results.

Table 4 compares the unconditional PoPA categorical forecasts to categorical forecasts obtained from the LFM model, subjective preparation at NMC, and climatology. Again, dependent data results are not shown because of their similarity to the results of the independent data verification. The following conclusions are indicated in Table 4: (1) PoPA categorical forecasts obtained by maximizing the percent correct were the only forecasts with a slight edge over climatology, (2) PoPA categorical forecasts were slightly better than NMC subjective and LFM forecasts, no matter how the PoPA forecasts were transformed with the exception of minimizing the bias.

TABLE 4. Comparative verification on independent data of (1) transformed unconditional PoPA, (2) NMC subjective, (3) LFM model, and (4) climatology forecasts.

Verification score	PoPA forecasts transformed by				NMC subjective forecasts	LFM forecasts	Climatology
	Maximizing			Minimizing bias			
	Percent correct	NMC QPF score	Utility score				
Percent correct	95.2	94.6	94.9	93.0	93.0	91.0	95.2
NMC QPF score	3.9161	3.9149	3.9126	3.8885	3.9069	3.8722	3.9149
Utility score	0.4814	0.4798	0.4804	0.4798	0.4770	0.4692	0.4806

TABLE 5. Categorical bias for (1) transformed PoPA, (2) NMC subjective, and (3) LFM forecasts.

	PoPA forecasts transformed by					
	Maximizing			Minimizing bias		
Category	Percent correct	NMC QPF score	Utility score		NMC subjective forecasts	LFM forecasts
1	1.05	1.04	1.04	0.99	0.99	0.97
2	0.0	0.63	0.15	0.98	1.51	1.80
3	0.33	0.0	0.63	1.00	1.21	1.83
4	0.0	0.0	0.0	1.29	0.35	0.56
5	0.0	0.0	0.0	3.00	0.0	0.0

It should be noted, however, that the PoPA forecasts transformed by maximizing percent correct, NMC QPF score, and utility score are more biased than those of the LFM model or NMC subjective as shown in Table 5. As a matter of fact, none of these three methods of transforming the PoPA forecasts resulted in category 4 or 5 forecasts. (There were 17 observations of category 4 and 2 observations of category 5.) It would seem that any method for predicting categorical precipitation amounts should be able to produce forecasts of those categories for which a requirement exists. This is a reason why climatology, which can produce only category 1 forecasts, is not a suitable system for practical purposes. Table 5 shows that PoPA forecasts transformed by minimizing categorical bias have the least bias. Category 4 and 5 forecasts are produced by this method.

Another attempt to reduce the category bias was made with a technique often referred to as "inflation" (Klein *et al.*, 1959). A recent application of this technique to forecasting wintertime cloudiness (Glahn, 1974) resulted in forecasts that were less biased and had a better percent correct than those transformed by an empirical matrix devised to produce relatively unbiased forecasts. Unfortunately, the results of using the inflation technique on the independent data in this study were disappointing; no category 4 or 5 forecasts were produced by this transformation method. In addition, category 2 forecasts were more biased than those transformed by minimizing the bias.

It can be seen in Table 4 that the categorical PoPA forecasts obtained by minimizing the bias scored slightly worse than the other PoPA forecasts, but nearly the same as the NMC subjective and slightly

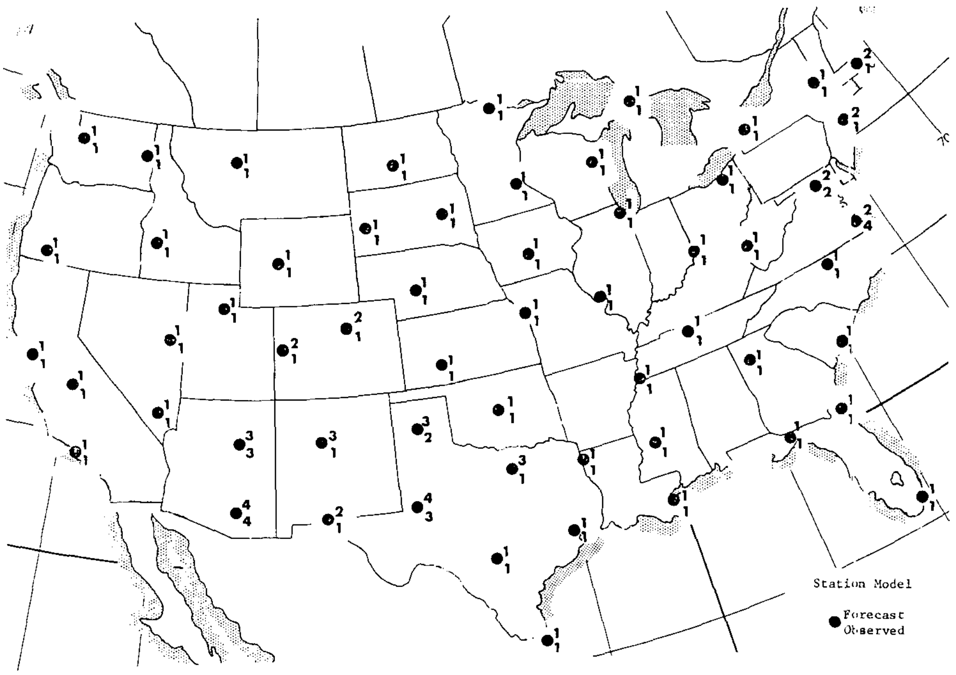

FIG. 1. PoPA 12–24 hour categorical forecasts transformed by minimizing the bias and observed 12–24 hour categorical amounts for the period ending 0000 GMT 20 October 1972. For corresponding precipitation amount categories, see Table 1.

better than the LFM forecasts. Therefore, consideration of bias as well as verification scores indicates that the PoPA categorical forecasts obtained by minimizing the bias are slightly better than the other forecasts.

A sample PoPA categorical forecast, transformed by minimizing the bias, is shown for the 58 cities in Fig. 1. Some rather good forecasts of larger amounts (e.g., Tucson and Winslow) are possible with this method. Of course, poor forecasts also occur (e.g., Norfolk and Albuquerque).

5. Summary and conclusions

Verification has indicated that PoPA categorical forecasts obtained by minimizing the bias were slightly better than all the other forecasts. They scored nearly the same as the NMC subjective and slightly better than the LFM forecasts and were the least biased of all the forecasts. Minimizing the bias was the only method of transforming PoPA forecasts which had the desirable characteristic of producing category 4 and 5 forecasts.

Verification has also indicated that unconditional PoPA categorical forecasts were about as good as conditional PoPA categorical forecasts.

On the basis of the results presented here, it is obvious that there remains considerable room for improvement in forecasting quantitative precipitation, especially amounts exceeding 1 inch. Improvements in the PoPA forecasts can probably be achieved by developing regionalized, generalized operator equations. In this way, more localized effects (e.g., topography and proximity to large bodies of water) can be taken into account.

Acknowledgments. I would like to express my gratitude to the following members of the Techniques Development Laboratory: Dr. Harry R. Glahn for many helpful discussions during the course of this work and for reviewing this paper, Capt. Frank T. Globokar (USAF) for developing the minimum bias matrix, and Mrs. Mary-Blue Battle for typing the paper.

REFERENCES

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.

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

——, ——, and F. T. Globokar, 1973: An application of model output statistics to the prediction of ceiling and visibility. Final Rept. No. FAA-RD 73-14, Techniques Development Laboratory, National Weather Service, 64 pp.

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

Carter, G. M., 1973: Use of model output statistics in automated prediction of surface winds. TDL Office Note No. 73–4, 9 pp.

Glahn, H. R., 1970: A method for predicting surface winds. ESSA Tech. Memo. WBTM TDL 29, 18 pp.

——, 1974: Problems in the use of probability forecasts. *Preprints of the Fifth Conf. on Weather Forecasting and Analysis,* Boston, Amer. Meteor. Soc., 32–35.

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

——, and ——, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211.

Howcroft, J., and A. Desmaris, 1971: The limited area fine mesh (LFM) model. NWS Tech. Proc. Bull. No. 67, 11 pp.

Klein, W. H., 1968: An objective method of predicting quantitative precipitation in the Tennessee and Cumberland Valleys. *Proc. of the First Statistical Conf.*, Boston, Amer. Meteor. Soc., 20–28.

——, and G. A. Hammons, 1973: Use of model output statistics for automated prediction of max/min temperatures. TDL Off. Note No. 73–3, 10 pp.

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

Korte, A. F., D. L. Jorgenson, and W. H. Klein, 1972: Synoptic climatological studies of precipitation in the plateau States from 850-, 700-, and 500-millibar lows during spring. NOAA Tech. Memo. NWS TDL 48, 130 pp.

Lowry, D. A., J. R. Bocchieri, and B. Rochlin, 1973: Operational forecasts derived from primitive equation and trajectory model output statistics (PEATMOS)—No. 4. NWS Tech. Proc. Bull. No. 83, 6 pp.

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

Spiegler, D. B., and G. E. Fisher, 1971: A snowfall prediction method for the Atlantic seaboard. *Mon. Wea. Rev.*, **99**, 311–325.

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

Wasserman, S. E., 1972: PEATMOS probability of precipitation as an aid in predicting precipitation amounts. NOAA Tech. Memo. NWS ER 50, 12 pp.