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

TDL OFFICE NOTE 80-3

A COMPARISON OF OBJECTIVE MODELS FOR DERIVING THRESHOLD PROBABILITIES TO MAXIMIZE THE THREAT SCORE

Edward A. Zurndorfer

April 1980

A COMPARISON OF OBJECTIVE MODELS FOR DERIVING THRESHOLD PROBABILITIES TO MAXIMIZE THE THREAT SCORE

Edward A. Zurndorfer

1. INTRODUCTION

Probability of precipitation amount (PoPA) and categorical forecasts of precipitation amount based on Model Output Statistics (MOS) (Glahn and Lowry, 1972) are currently produced twice daily from 0000 GMT and 1200 GMT data (Bermowitz and Zurndorfer, 1979). Forecasts of ≥ .25, ≥.50, ≥ 1.00 and ≥ 2.00 inches and a "best" categorical forecast are provided to forecasters both at the National Meteorological Center and in the field for various projections from both 0000 GMT and 1200 GMT.

In order to convert these probability forecasts to a "best" categorical forecast, we derive threshold probabilities[1] that maximize the threat score for each of the probability forecasts. Until recently, these threshold probabilities were determined subjectively using an iterative technique (Bermowitz and Zurndorfer, 1979). However, in recent papers, Bermowitz and Best (1978) and Miller and Best (1978) describe an objective technique for obtaining these threshold probabilities. Using this objective technique, we can determine the threshold probabilities as a function of the multiple correlation coefficients, R, (between predictors and predictand) and sometimes C, the climatological frequency of the predictand.

In this paper, we compare several of the models for objectively determining these threshold probabilities for the purpose of converting probability forecasts to categorical forecasts of precipitation amount. The research presented in this paper is a continuation of that reported by Bermowitz and Best (1978); the primary purpose of the work is to use the results for our operational PoPA equations.

2. METHOD

The three models that we used for testing were the: (1) R model, (2) R and C model, and (3) M & B model. For a description of these models, see Bermowitz and Best (1978).

To test the three models, we performed a verification in which precipitation amount forecasts made from thresholds obtained from the three models were compared to those obtained from the subjective, iterative technique. Four sets of independent data forecasts were available for the comparative verification, based on output from the Limited-area Fine Mesh (LFM) model: (1) 12-24 h from 0000 GMT warm season (April-September) equations, (2) 24-36 h from 0000 GMT warm season equations, (3) 18-24 h from 1200 GMT cool season (October-March) equations and (4) 12-24 h from 1200 GMT cool season equations. Each independent data set consisted of one year of data.

---

[1]The threshold probability for a category, say ≥ .25 inch of precipitation, is a value that if exceeded by a probability forecast for that category, would result in a categorical forecast of ≥ .25 inch. If the threshold value is not exceeded, the categorical forecast would be < .25 inch.

In all cases except the 18-24 h projection (only a 6-h period) threat score and biases were computed for forecasts of the precipitation amount categories ≥ .25, ≥ .50, ≥ 1.00, and ≥ 2.00 inches. For the 18-24 h projection, the ≥ 2.00 inch category was not used. Verification scores were computed at 233 cities over the conterminous U.S. for each projection.

3. RESULTS AND CONCLUSIONS

In Tables 1-4 we present the comparative verifications on the four data sets. The column headed SUBJ represents the results of verification of categorical forecasts obtained by using threshold probabilities derived with the iterative technique. However, there were no 1200 GMT 12-24 h cool season subjective forecasts available; before the 1979-80 cool season, no forecasts were made for that projection.

From Table 1, while it is apparent for the categories ≥ .25, ≥ .50, and ≥ 1.00 inches that the threat scores for the four systems are about the same, the R and RC models have the best biases. That is, the R and RC models have biases closest to 1.0. For the ≥ 2.0 category, the RC model has the largest threat score (.060) although its bias (1.61) is larger than the R model bias (1.25) whose threat score (.049) is second best among the four systems. Overall, therefore, the R and RC models are the best with perhaps a slight edge to the RC model.

For the results of the 0000 GMT 24-36 h warm season verification, we see from Table 2 that again the R and RC models performed the best. In particular, either the R or the RC model had the highest threat score and/or the bias closest to 1.0 for the four categories. Note also the rather large biases for the M and B model--especially for the categories ≥ 1.00 and ≥ 2.00 inches. In short, the R model should be classified as the best model here by virtue of having the biases closest to 1.0 and almost the largest threat scores for each of the four categories.

The results of verification for the cool season (October-March) shown in Tables 3 and 4 indicate: (1) That the R and RC models performed the best; (2) the M & B model, while producing high threat scores, has biases far greater than 1.0, especially for the higher amounts, and (3) if any model has to be selected as being the best, it appears that the RC would have a slight edge over the R model.

The results of this comparative verification indicate indeed that the objective models are superior to the subjective model. In particular, the RC model performed best for less than 24-h projections, and the R model did best for the one projection beyond 24 hours. Perhaps one explanation for these results is that the forecasts for projections less than 24 hours were made from equations derived on seven to eight years of dependent data. Therefore, the climatological frequencies, C, used in deriving threshold probabilities for the RC model tend to be stable and reliable. On the other hand, for the 24-36 h projection from 0000 GMT, forecasts were made from equations derived on only four years of dependent data. In this case, the climatological frequencies tend to be less stable and reliable.

4. OPERATIONAL CONSIDERATIONS

Starting with the 1980-81 cool season and continuing into the 1981 warm season, we will use the R and RC models to derive threshold probabilities (that maximize the threat score) for the purpose of converting PoPA forecasts to categorical forecasts of precipitation amount. For projections less than hour 24 from model run time we will use the RC model and for projections greater than hour 24 from model run time we will use the R model. At any rate, the results shown in this paper, together with results for other projections (not shown) indicate that the objective technique is indeed a useful technique for deriving threshold probabilities to maximize the threat score. One cannot overemphasize the savings--both from the standpoint of the computer and man hours--in using the objective method instead of the subjective method to derive the threshold probabilities. For example, one must derive a total of 36 threshold probabilities for a typical projection (one probability for each of 9 regions and 4 predictands). This involves perhaps 6 additional computer runs after the PoPA regression equations have been derived in order to subjectively determine the threshold probabilities. These additional computer runs require both a large amount of storage (core and disk space) and perhaps 4 minutes on the CPU to run. When one has to derive threshold probabilities for 20 projections for 2 seasons (as is done for the PoPA guidance package), one is talking about saving 6x4x20x2=960 minutes on the CPU per year. Therefore, we highly recommend that other users who must convert probability forecasts to a "best" categorical forecast, investigate the use of these objective methods for deriving threshold probabilities.

REFERENCES

Bermowitz, R. J. and D. L. Best, 1978: An objective method for maximizing threat score. TDL Office Note 78-15, National Weather Service, NOAA, U.S. Department of Commerce, 11 pp.

______, and E. A. Zurndorfer, 1979: Automated guidance for predicting quantitative precipitation. Mon. Wea. Rev., 107, 122-128.

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.

Miller, R. G., and D. L. Best, 1978: A model for converting probability forecasts to categorical forecasts. TDL Office Note 78-14, National Weather Service, NOAA, U.S. Department of Commerce, 10 pp.

Table 1. Comparative verification of 0000 GMT 12-24 h warm season (April-September) LFM based PoPA categorical forecasts made from threshold probabilities derived from the (1) subjective method (SUBJ), (2) R, (3) RC, and (4) M&B models. Independent data sample consists of one season of forecasts at 233 cities.

CATEGORY (INCH)	THREAT SCORE				BIAS			
	SUBJ	R	RC	M&B	SUBJ	R	RC	M&B
≥ .25	.229	.228	.235	.237	1.25	1.17	1.16	1.20
≥ .50	.153	.151	.151	.152	1.36	1.23	1.19	1.30
≥1.00	.096	.093	.094	.092	1.43	1.22	1.27	1.31
≥2.00	.036	.049	.060	.046	1.40	1.25	1.61	0.68

Table 2. Same as Table 1 except for the 0000 GMT 24-36 h warm season projection.

CATEGORY (INCH)	THREAT SCORE				BIAS			
	SUBJ	R	RC	M&B	SUBJ	R	RC	M&B
≥ .25	.177	.205	.206	.206	1.53	1.20	1.21	1.29
≥ .50	.120	.139	.141	.139	1.75	1.31	1.38	1.62
≥1.00	.064	.067	.073	.075	1.79	1.02	1.23	2.04
≥2.00	.014	.059	.057	.032	1.26	0.92	1.28	3.45

Table 3. Same as Table 1 except for the 1200 GMT 18-24 h cool season (October-March) projection.

CATEGORY (INCH)	THREAT SCORE				BIAS			
	SUBJ	R	RC	M&B	SUBJ	R	RC	M&B
≥ .25	.229	.225	.240	.240	1.15	1.03	1.25	1.66
≥ .50	.162	.161	.148	.148	1.48	1.13	1.40	2.11
≥1.00	.077	.068	.067	.072	1.54	0.54	0.98	3.02

Table 4. Same as Table 1 except for the 1200 GMT 12-24 h cool season projection. No results for the subjective method (SUBJ) were available.

CATEGORY (INCH)	THREAT SCORE				BIAS			
	SUBJ	R	RC	M&B	SUBJ	R	RC	M&B
≥ .25	--	.335	.341	.338	--	0.99	1.13	1.33
≥ .50	--	.273	.280	.269	--	1.01	1.20	1.57
≥1.00	--	.169	.173	.170	--	1.14	1.46	2.24
≥2.00	--	.024	.041	.042	--	0.19	0.42	1.75