PREPRINTS OF PAPERS PRESENTED AT THE

INTERNATIONAL SYMPOSIUM ON PROBABILITY AND STATISTICS IN THE ATMOSPHERIC SCIENCES

JUNE 1-4, 1971
HONOLULU, HAWAII
STATISTICS OF NUMERICAL PREDICTION MODELS

Harry R. Glahn, Dale A. Lowry, George W. Hollenbaugh, and John R. Annett

Techniques Development Laboratory
National Weather Service, NOAA
Silver Spring, Maryland

1. INTRODUCTION

Until rather recently, objective forecasting methods 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. 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 is usually called the perfect prog method. A concurrent statistical relationship is developed between the variable to be estimated and selected variables forecast by a dynamic model. Both predicated 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 hours to get an estimate of the predicated 36 hours 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 predicated and variables from the numerical model at the desired projection time. 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 wish to know what percent of the time it rains when the model predicts 80 percent 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 location in time and space. In the following sections we will describe some of our results with this technique.

2. PROBABILITY OF PRECIPITATION

Perhaps the first major use of MOS was in the estimation of probability of precipitation (PoP) (Glahn and Lowry, 1969). To do this, we used a combination of the variables predicted by two models being run operationally at the National Meteorological Center (NMC): the Primitive Equation (PE) model (Shuman and Hovemara, 1968) and the Subsynoptic Advection Model (SAM) (Glahn, Lowry, and Hollenbaugh, 1969). These PoP forecasts are made about 3:00 A.M. EST for the "today" period 7:00 A.M. to 7:00 P.M. The area covered is generally the United States east of the Mississippi River.

Screening Regression was used to select predictors which contributed most to the reduction of variance of the binary predicated. Data for nearly 100 stations were grouped together and the equations developed and updated twice a year (Summer: April-September; Winter: October-March). Predictors were all binary and were computed from PE relative humidity, SAM saturation deficit, PE precipitation amount, and SAM sea level pressure.

Figure 1 shows a portion of a facsimile chart transmitted January 18, 1971. The isopleths indicate PoP forecasts for the 12-hr period 1200 GMT January 18 to 0000 GMT January 19. In this particular chart, there is a variation from < 5% to > 55% over a distance of about 150 miles.

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; Figure 2 gives the monthly and yearly Brier scores.* This figure indicates:

a) The MOS forecasts were better than the local forecasts for the first 15 months of comparison. The reverse is true for the last 15 months. (The local forecasts are made with 1 or 2 hours later data available, and many stations receive the MOS forecasts before the local forecasts are finalized.)

b) The MOS forecasts have been better, on the average, than the NMC guidance for the 18-month period of comparison.

\[ \text{Brier Score} = \frac{1}{N} \sum_{i=1}^{N} (F_i - O_i)^2 \]

where \( N \) = sample size, \( F \) = forecast probability, and \( O \) = observed (0 = no precipitation, 1 = precipitation).
The sharp decrease in skill of the MOS forecasts relative to the locals was mainly due
to a major change in the moisture portion of the
PE model in late October 1969. Because of this
change, PE predictors were not used from December
1969 through August 1970. Since that time PE
predictors have been included, following another
PE change in September 1970.

The effort in PoP forecasting is being
extended to all three forecast periods by com-
bining the output from the TDL trajectory model
(Reap, 1968), which is also being run opera-
tionally at NMC, with that from the PE model.
These Primitive Equation and Trajectory (PEAT)
MOS forecasts have been compared with those made
by the perfect prog technique (Klein, 1971) as
well as with the official local office forecasts
and the NMC guidance.

Table 1 gives the Brier Scores for
October, November, and December for 86 stations
over the 48 states. Figure 3 shows the bias in
the four sets of forecasts. From this somewhat
limited verification, we conclude:

a) The PEATMOS forecasts are competitive
with NMC and, except for the first
period, with the local offices.

b) The PEATMOS forecasts are too low for
all forecast categories except 6-15
percent. Since a large number of
forecasts are in this category, the
overall bias is small. All other fore-
casts are too high on the average,
particularly the perfect prog system.

### Table 1. Brier Scores for local, NMC, PEATMOS, and perfect prog PoP forecasts for 86 stations in the 48 states. Climatology is long-term relative frequency by month and by station

<table>
<thead>
<tr>
<th>PERIOD 1 (TODAY)</th>
<th>BRIER</th>
<th>NMC</th>
<th>MOS</th>
<th>PERFECT PROG</th>
<th>CLIMATOLOGY</th>
<th>CASES</th>
</tr>
</thead>
<tbody>
<tr>
<td>OCTOBER</td>
<td>.0804</td>
<td>.1020</td>
<td>.0957</td>
<td>.1134</td>
<td>.1273</td>
<td>750</td>
</tr>
<tr>
<td>NOVEMBER</td>
<td>.0836</td>
<td>.1012</td>
<td>.1013</td>
<td>.1199</td>
<td>.1452</td>
<td>2316</td>
</tr>
<tr>
<td>DECEMBER</td>
<td>.0891</td>
<td>.1082</td>
<td>.1024</td>
<td>.1189</td>
<td>.1522</td>
<td>2220</td>
</tr>
<tr>
<td>AVERAGE</td>
<td>.0854</td>
<td>.1042</td>
<td>.1009</td>
<td>.1185</td>
<td>.1456</td>
<td>(5286)</td>
</tr>
<tr>
<td>IMPROVEMENT OVER CLIMATOLOGY</td>
<td>.413</td>
<td>.284</td>
<td>.307</td>
<td>.186</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIAS (PERCENT)</td>
<td>AVERAGE</td>
<td>2.6</td>
<td>4.6</td>
<td>- .2</td>
<td>6.9</td>
<td></td>
</tr>
<tr>
<td>PERIOD 2 (TODAY)</td>
<td>BRIER</td>
<td>NMC</td>
<td>MOS</td>
<td>PERFECT PROG</td>
<td>CLIMATOLOGY</td>
<td>CASES</td>
</tr>
<tr>
<td>OCTOBER</td>
<td>.1104</td>
<td>.1073</td>
<td>.0981</td>
<td>.1256</td>
<td>.1229</td>
<td>752</td>
</tr>
<tr>
<td>NOVEMBER</td>
<td>.1104</td>
<td>.1073</td>
<td>.1175</td>
<td>.1290</td>
<td>.1503</td>
<td>2282</td>
</tr>
<tr>
<td>DECEMBER</td>
<td>.1196</td>
<td>.1247</td>
<td>.1221</td>
<td>.1275</td>
<td>.1529</td>
<td>2218</td>
</tr>
<tr>
<td>AVERAGE</td>
<td>.1129</td>
<td>.1187</td>
<td>.1166</td>
<td>.1278</td>
<td>.1474</td>
<td>(5252)</td>
</tr>
<tr>
<td>IMPROVEMENT OVER CLIMATOLOGY</td>
<td>.234</td>
<td>.195</td>
<td>.209</td>
<td>.133</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIAS (PERCENT)</td>
<td>AVERAGE</td>
<td>2.2</td>
<td>4.4</td>
<td>- .7</td>
<td>6.6</td>
<td></td>
</tr>
</tbody>
</table>

| PERIOD 3 (TODAY) | BRIER | NMC | MOS | PERFECT PROG | CLIMATOLOGY | CASES |
| OCTOBER          | .1219 | .1225| .1212| .1312        | .1342       | 750   |
| NOVEMBER         | .1208 | .1226| .1316| .1373        | .1468       | 2289  |
| DECEMBER         | .1373 | .1341| .1375| .1341        | .1546       | 2215  |
| AVERAGE          | .1279 | .1274| .1326| .1350        | .1482       | (5254) |
| IMPROVEMENT OVER CLIMATOLOGY | .137 | .140 | .105 | .089 |
| BIAS (PERCENT)   | AVERAGE | 1.2 | 3.1 | -2.8 | 6.5 |
Figure 2. Monthly and average Brier Scores for local forecasts and MOS forecasts from the SAM teletype and facsimile products for July 1968 through December 1970. NMC Scores are given as averages only. Nineteen stations in the eastern U. S. are included.
Figure 4. Mean absolute errors for local, MOS (SAM), and perfect prog (Klein) forecasts. Sixteen stations in the eastern U.S. are included.

b) The perfect prog and MOS techniques gave forecasts of about equal skill.

c) The perfect prog system applied to PE forecasts gave slightly better results than the same system applied to barotropic forecasts in all 6 months for which the comparison was made.

It is clear that for this short range forecast, the forecasts made at local stations have less error, on the average, than our objective forecasts. Quite likely, to equal the local skill, we will have to include more information about the initial soundings and, perhaps, stratify according to synoptic situation. A complicating factor is that the maximum temperature does not occur at the same time each day. The poor relative performance of MOS during the 1970 summer was probably due to some change in the PE model that affected the PE temperature forecasts.

4. SURFACE WIND

Separate regression equations were developed for estimating the U and V wind components and the wind speed valid at 1200 and 1800 GMT for each of 10 stations in the eastern U.S.—Albany, Atlanta, Baltimore, Cleveland, Cincinnati, Washington, New York, New Orleans, Chicago, and St. Louis (Glahn, 1970). Data for April through September 1967 and 1968 were used. The most important predictors were geostrophic winds from SAM, although other variables such as 1000- and 500-mb winds and temperatures and initial observed winds also played a part.

The equations were evaluated for each day in April and May 1969 for which SAM data tapes were available. The wind forecasts in the terminal forecasts (FT's) made at the Weather Service Offices were used for comparison. Since the FT's do not mention wind if the speed is expected to be less than 10 kts, the comparison was made in two ways.
For all those cases were the FT's included wind and objective forecasts were available, the root mean square error (RMSE) of direction (computed from the U and V equations) and speed were computed. Also, contingency tables for speed were prepared by considering the FT forecast of wind to be under 10 kts when wind was not mentioned. From these contingency tables, skill scores and percent correct were computed. These scores are shown in Table 2.

Table 2. Comparison of official FT and objective forecasts for 10 stations in the eastern U. S. for April and May 1969.

<table>
<thead>
<tr>
<th>Valid Time (GMT)</th>
<th>Projection (hrs)</th>
<th>Forecast</th>
<th>Direction Error (deg)</th>
<th>Speed Error (kts)</th>
<th>Frequency Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>5</td>
<td>OBJECTIVE</td>
<td>35</td>
<td>2.5</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>SPEED EQUATION</td>
<td>35</td>
<td>2.5</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>FT</td>
<td>35</td>
<td>2.5</td>
<td>22</td>
</tr>
<tr>
<td>18</td>
<td>11</td>
<td>OBJECTIVE</td>
<td>47</td>
<td>3.3</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>SPEED EQUATION</td>
<td>47</td>
<td>3.3</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>FT</td>
<td>59</td>
<td>4.3</td>
<td>24</td>
</tr>
</tbody>
</table>

Table 2 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 hours) refer to the latest data used (0700 GMT). Actually, the forecasts are available to the field forecasters before 0900 GMT, although this was not true for the period of comparison. The FT's were prepared with 0900 and perhaps 1000 GMT data available; transmission time for the forecasts is 1045 GMT.

5. DISCUSSION AND CONCLUSIONS

In this paper we have described the MOS technique and some of our uses of it within TDL. The verifications completed to date are very encouraging in that the MOS forecasts are competitive with both local forecasters and NMC guidance, although in most cases the local forecasters are slightly better. Only within the last few years have objective forecasts of actual "weather" variables been mass-produced with an accuracy that can compete with trained forecasters. We expect this trend to continue so that objective forecasts can be used unchanged as a large portion of the time and modified only in special and difficult synoptic situations.

The main problems encountered with MOS are:

a) Only relatively small data samples are available.

b) Frequent updates of the system are necessary, or, at least, desirable.

c) Numerical model changes may render a system unusable or necessitate a change in application.

Even with the problems listed above, we feel the use of MOS is desirable, especially when a probability estimate is desired.

Acknowledgements. The authors wish to thank the many members of TDL, and especially Mr. Harry Akens, Mrs. Elizabeth Booth, Mrs. Evelyn Boston, Mr. Dean Costantinou, Mr. Darrel Foat, Mr. Peter Howe, Miss Loretta Thompson, and Mrs. Betty Wells, who have contributed to this work in various ways over the past several years.

REFERENCES


