A STUDY OF THE SIGNIFICANCE OF FORECASTER CHANGES TO MOS TEMPERATURE GUIDANCE

Robert B. Wassall
National Weather Service Forecast Office
Philadelphia, Pennsylvania 19106

1. INTRODUCTION

"We can improve upon MOS more than any other office in the Eastern Region!" That was the challenge extended by WSFO, Washington in Eastern Region STAFF MEETING NOTES dated March 15, 1976. The statement followed a summary of 1975 verification results for six locations in the Washington forecast area. The article proudly noted improvement over MOS for all three forecast periods and for both temperature and precipitation forecasts.

Seldom willing to back off from a challenge, the author decided to summarize 1975 verification data for forecasts issued by WSFO, Philadelphia, Pa. Since Philadelphia also keeps data for six locations (Harrisburg, Williamsport, Wilkes Barre-Scranton, Allentown, Atlantic City, and Philidelphia), and since the Philadelphia forecast area is similar to that of Washongton's with regard to climatology and topography, it seemed that a comparison of results would be a fair estimate of the relative forecast skill displayed by the two WSFOs.

While the results of the comparison (for temperature only), Table 1, proved extremely interesting, characteristics of the error pattern prompted the author to make a much more detailed analysis of the temperature forecasts. It became apparent in working with the data that only a small number of forecasts significantly improved on the MOS guidance. The vast majority of the forecasts appeared to be within a couple degrees of the MOS guidance values. From Table 2 it can be seen that the average improvement over MOS for all forecast periods was 0.37° ranging from 0.64° for the first period cold season forecasts to 0.13° for third period warm season forecasts. (The warm season was defined as May through October and the remaining six months were considered the cold season, henceforth called summer and winter in this report.)

It is the main objective of this study therefore to analyze temperature predictions (FP) made by forecasters at WSFO, Philadelphia, during the year 1975, in an attempt to determine the significance of the changes made in MOS guidance.

Before proceeding with the detailed analysis, the comparison between the Phildelphia and Washington results is shown below.

TABLE 1 - Average improvement over MOS for all temperature forecasts made in 1975.

	Period 1	Period 2	Period 3
Washington Forecasts	0.5°F	0.4°F	0.2°F
Philadelphia Forecasts	0.47	0.39	0.25

It is remarkable that the overall Washington and Philadelphia performances were so similar. Depending upon the rounding off technique, the average improvement over MOS for each of the forecast periods could be considered exactly the same.

In view of these results, Philadelphia is willing to call the contest a "stand-off" if Washington is willing to agree to the stalemate!

The author's first reaction to the above comparison was to say that Philadelphia had also done a commendable job in outperforming MOS. But, reflecting further on the small average improvement over MOS, a person could not help but wonder, what does it all mean? Does John Q. Public have the sensitivity to detect temperature differences which, on the average, are less than one degree? Would the deterioration in forecast accuracy have been noticed if MOS forecasts had been used without change? And finally, why would forecast performance have been without benefit of the MOS guidance? The analysis which follows addresses itself to those questions.

Table 2 - Average improvement over MOS guidance °F) for all temperature forecasts issued by WSFO, Philadelphia during the year 1975. Six forecast locations and both the 5:00 a.m. and 5:00 p.m. forecasts are not included.

Season	# of Fcsts	1	2	3	All
Winter	6456	0.64°F	0.57°F	0.38°F	0.53F
Summer	6537	0,30	0.22	0.13	0.22
Annual	12993	0.47	0.39	0.25	0.37

Tables 3, 4, and 5 show the percentage of FP forecasts that improved on the MOS guidance by specified amounts for each of the forecast periods. Included is a column showing cumulative percentages for specified improvements OR MORE. It can be seen from Table 3 for instance that 6% of second period summer forecasts improved on MOS by 3 degrees and that 12% of second period summer forecasts improved on MOS by 3 degrees or more.

Table 3 - Frequency of specified improvement over MOS for temperature forecasts issued by WSFO, Philadelphia for six locations during summer 1975. Negative improvements indicate a degradation of the MOS forecasts. *(less than 0.5, or, 99.5 or greater but less than 100).

SUMMER

	1st Period		2nd Period		3rd Period		All Fcsts	
Improvement over MOS (°F)	% of Fcsts	Cum % Fcsts	% of Fcsts	Cum % Fcsts	% of Fcsts	Cum % Fcsts	% of Fcsts	Cum % Fcsts
12	0*	0*	0	0	0	0	0	0
11	0	0*	0	0	0	0	0	0
10	0*	0*	0	0	0*	0*	0*	0*
9	0*	0*	0*	0*	0*	0*	0*	0*
8	0*	1	0*	0*	0*	0*	0*	0*
7	0*	1	0*	1	0*	1	0*	1
6	1	2	1	1	1	1	1	1
5	1	3	2	3	1	2	1	3
4	3	6	3	6	2	5	3	5
3	7	13	6	12	5	10	6	11
2	14	27	13	25	12	21	13	24
1	17	43	16	41	16	37	16	41
0	26	69	26	67	30	67	27	67
-1	14	83	15	82	16	82	15	82
-2	10	92	10	92	10	92	10	92
-3	4	96	4	96	4	96	4	96
-4	2	98	2	98	2	98	2	98
-5	1	99	1	99	1	99	1	99
-6	0*	99	1	100*	1	99	0*	99
-7	0*	100*	0*	100*	0*	100*	0*	100*
-8	0*	100*	0*	100*	0*	100*	0*	100*
-9	0*	100*	0	100*	0*	100*	0*	100*
-10	0*	100*	0*	100	0	100*	0*	100*
-11	0**	100	0	100	0*	100	0*	100

Table 4 - Frequency of specified improvement over MOS for temperature forecasts issued by WSFO, Philadelphia for six locations during winter 1975. Negative improvements indicate a degradation of the MOS forecasts. *(less than 0.5, or, 99.5 or greater but less than 100).

WINTER

	1st Period		2nd Period		3rd Period		All Fcsts	
Improvement over MOS (°F)	% of Fcsts	Cum % Fcsts	% of Fcsts	Cum % Fcsts	% of Fcsts	Cum % Fcsts	% of Fcsts	Cum % Fcsts
13	0	0	0*	0*	0	0	0*	0*
12	0	0	0	0*	0	0	0*	0*
11	0*	0*	0*	0*	0*	0*	0*	0*
10	0*	1	0*	0*	0*	0*	0*	0*
9	0*	1	0*	1	0*	1	0*	1
8	1	1	1	1	0*	1	0*	1
7	1	2	1	2	1	2	1	2
6	2	4	1	3	1	2	1	3
5	3	7	4	8	3	5	3	6
4	5	12	5	12	4	9	5	11
3	7	19	7	19	7	16	7	18
2	14	33	14	33	13	29	14	32
1	15	48	16	49	14	43	15	47
0	22	71	19	69	25	68	22	69
-1	12	83	12	81	13	81	12	82
-2	9	92	10	91	10	91	10	91
-3	4	96	4	95	4	95	4	95
-4	2	98	3	98	2	97	2	97
-5	1	99	1	98	1	98	1	99
-6	1	99	1	99	1	99	1	99
-7	0*	100*	0*	100*	1	100*	0*	100*
-8	0*	100*	0*	100*	0*	100*	0*	100*
-9	0*	100*	0*	100*	0*	100*	0*	100*
-10	0*	100	0*	100*	0*	100*	0*	100*
-11	0	100	0*	100*	0*	100	0*	100*
-17	0	100	0*	100	0	100	0*	100

The most frequent improvement over MOS in all forecast periods and in both seasons was "zero" improvement! Overall, 22% of the winter forecasts and 27% of the summer forecasts fell into that category. A "zero" improvement could be achieved by forecasting the same value as the MOS guidance or by forecasting a value that missed the observed temperature by the same amount as MOS but in the opposite direction.

For the entire year, 44% of the forecasts improved on MOS; 24% had the same accuracy; and 32% were less accurate. The largest single improvement was 13 degrees and the largest negative change (degredation) was 17 degrees. The latter was an isolated occurrence, the next largest negative change being 11 degrees.

The cumulative columns in Tables 3 through 5 show the percentage of forecasts that improved on MOS by a certain value or more. Only 3% of summer forecasts and 6% of winter forecasts improved on MOS by 5 degrees or more. On a year-round basis, only a little over one forecast in four (28%) improved on MOS by 2 degrees or more.

The length of the forecast seemed to have little effect on performance. It appeared to be just as difficult to make a large improvement over MOS in the first period as it did in the second or third forecast periods.

The data was further analyzed by averaging errors by forecast periods (six locations combined) and for entire forecasts (six locations and three forecast periods for a total of 18 forecasts). From Table 6 it becomes obvious that the more forecasts that are averaged, the more difficult it is to maintain a specific average improvement over MOS. While 5% of the individual forecasts (annually) improved on MOS by 5 degrees or more, the same improvement was noted in only 1.1% of the forecast periods (six forecasts combined) and in 0% of the complete forecasts (18 forecasts combined). The frequency of improvement of 2 degrees or more diminished from 28% for individual forecasts to 13% for forecast periods and to 6% for complete forecasts.

2. DISCUSSION OF RESULTS

Although certain users with weather sensitive operations use the forecasts issued by the National Weather Service, the primary user continues to be the general public. The forecasts are used by the masses to determine general comfort conditions, dress requirements, the need for heating or cooling, the practicality of conducting outdoor activities including vacations, the need to take certain protective actions, and as the

Table 5 - Frequency of specified improvement over MOS for temperature forecasts issued by WSFO, Philadelphia for six locations during 1975. Negative improvements indicate a degradation of the MOS forecasts. *(less than 0.5, or, 99.5 or greater but less than 100).

ANNUAL

	1st Period		2nd Period		3rd Period		All Fcsts	
Improvement over MOS (°F)	% of Fcsts	Cum % Fcsts	% of Fcsts	Cum % Fcsts	% of Fcsts	Cum % Fcsts	% of Fcsts	Cum % Fcsts
13	0	0	0*	0*	0	0	0*	0*
12	0*	0*	0	0*	0	0	0*	0*
11	0*	0*	0*	0*	0*	0*	0*	0*
10	0*	0*	0*	0*	0*	0*	0*	0*
9	0*	1	0*	0*	0*	0*	0*	0*
8	1	1	1	1	0*	1	1	1
7	1	2	1	1	1	1	1	1
6	1	3	1	2	1	2	1	2
5	2	5	3	5	2	4	3	5
4	4	9	4	9	3	7	3	8
3	7	16	6	15	6	13	7	15
2	14	30	14	29	12	25	13	28
1	16	46	16	45	15	40	16	44
0	24	70	23	68	28	68	24	68
-1	13	83	13	81	14	82	14	82
-2	9	92	10	91	10	92	10	92
-3	4	96	4	95	4	96	4	96
-4	2	98	3	98	2	98	2	98
-5	1	99	1	99	1	99	1	99
-6	1	100*	1	99	1	99	1	99
-7	0*	100*	0*	100*	0*	100*	0*	100*
-8	0*	100*	0*	100*	0*	100*	0*	100*
-9	0*	100*	0*	100*	0*	100*	0*	100*
-10	0*	100	0*	100*	0*	100*	0*	100*
-11	0	100	0*	100*	0*	100	0*	100*
-17	0	100	0*	100	0	100	0*	100

Cumulative Improvement Over MOS (°F)	SUMMER			WINTER			ANNUAL		
	Single Fcsts	Fcst Periods	Total Fcsts	Single Fcsts	Fcst Periods	Total Fcsts	Single Fcsts	Fcst Periods	Total Fcsts
5 or more	3%	0.5%	0%	6%	1.7%	0%	5%	1.1%	0%
4 or more	5	0.9	0	11	4.8	0.6	8	2.8	0.3
3 or more	11	3.5	0	18	9	1.9	15	6	1.0
2 or more	29	9	1.9	32	18	9	28	13	6
1 or more	41	25	17	47	37	33	44	31	25
0 or more	67	63	66	69	65	69	68	64	67

Table 6 - Percent of temperature forecasts improving on MOS by specified amounts in forecasts issued by WSFO, Philadelphia for six locations during 1975.

Single Fcsts - each forecast for a location and time period is included in this percentage distribution as an individual data point.

Fcst Periods - the improvements are averaged for the six locations.

Total Fcsts - the improvements are averaged for the six locations and three time periods, a total of 18 data points.

subject of idle conversation. For most of these considerations, a difference of a few degrees in a temperature forecast would probably make little or no difference in the

It would appear that a forecaster would have to improve on the accuracy of his temperature forecasts by at least 4 or 5 degrees before the general public would benefit.

It would seem therefore that users receive little benefit from the massaging of MOS temperature guidance by forecasters. If only 3% of individual summer forecasts and 6% of winter forecasts display an improvement of 5 degrees or more because or massaging, resultant deterioration would probably go unnoticed by the general user if MOS forecasts were used "as-is."

The impact of using MOS temperatures directly would be softened by the fact that forecasters normally give a range of temperatures in a forecast. The range may be as much as 5 to 8 degrees. The FP temperatures entered for verification purposes and the MOS temperatures are frequently in the same predicted range, although perhaps several degrees apart. In most instances, then, direct use of the MOS temperatures would not change the temperature range issued to the public. The improvement that is actually achieved by massaging the MOS temperatures is for the most part an internal matter that does not reach the user. It is difficult to estimate how frequently the forecast range would actually be changed if MOS was used "as-is" but it would appear to be very seldom.

3. CONCLUSIONS

When one considers all forecasts issued for a period (6) or in an entire forecast (18), it becomes even more apparent that the massaging is of doubtful value. As more temperatures are considered, it is increasingly difficult to maintain an average improvement over MOS that is significant. Instead, improvements that may appear to be significant for an individual location are often offset in part or in toto by other forecasts that are not as accurate as the MOS forecasts. The question should be asked, "Does it help the user in general if a forecast for one location is significantly more accurate than MOS if at the same time, the forecast for another location is significantly less accurate than MOS?" Would the user be just as well off in the long run if MOS temperatures had been used without change?

It would seem that resources are being misused if the practice of massaging MOS temperatures can achieve an average improvement of just 0.37 degrees over the period of a year. Perhaps forecasters should use MOS "as-is" except for those occasional situations when the MOS temperatures are obviously in error because of frontal timing or a persistent model error.

An important unanswered question however, is the level of performance that might have been achieved by experienced forecasters working without benefit of the MOS temperatures as guidance. Is it possible that forecasters,

Continued on Page 36.

WASSAL from page 30 --

unduly influenced by MOS, tend to compromise on opportunities to make large improvements over MOS? In the year studied, a total of 589 MOS forecasts were in error by 10 degrees or more. The average error of those 589 forecasts was 12.1 degrees. After massaging by forecasters, the average error of the forecasts was reduced by just 2.6 degrees. Is this small error reduction an indication of the forecaster's inability to detect potentially large guidance errors or is it the result of a growing dependence on MOS? It is the author's opinion that questions should be answered before yielding completely to the MOS forecasts.