ANALYSIS OF TEMPERATURE FORECASTS AT MILWAUKEE, WISCONSIN OVER A ONE-YEAR PERIOD

Timothy W. Oster
National Weather Service
Milwaukee, WI 53207

ABSTRACT

Computer guidance and forecaster-produced temperature forecasts issued for Milwaukee, Wisconsin were examined for one year. A tabulation of the forecast errors for each of five periods and for the four seasons show that the forecasters tended to over-forecast highs and underforecast lows. Guidance also demonstrated this tendency. In addition, periods during which guidance was consistently much too warm or too cold were found and are examined.

1. INTRODUCTION

Temperatures are one of the most important weather elements forecast routinely in public weather forecasts. The public's interest in the temperature outlook is probably second only to the precipitation forecast. And, of course, power companies have a vital interest in the temperature forecasts during both the heating and air-conditioning seasons.

A year's worth of temperature forecasts for Milwaukee, Wisconsin have been examined to see what kinds of errors are commonly made in the public weather forecasts. Also, automated temperature forecasts, obtained by use of the Model Output Statistics (MOS) technique (1), have been examined for the same period in an effort to find any systematic errors which may be useful to a forecaster using the guidance. Thirdly, the public forecasts and MOS guidance have been compared directly to see how much improvement the forecasters have shown over the machine.

2. DATA SOURCES

The coded form of a Travelers forecast -- called an FP4 -- is issued by Weather Service Forecast Officers (WSFO's) twice daily for various locations in the U.S. At WSFO Milwaukee, only one is issued and that is for Milwaukee itself.

The FP4's are issued at 0958 GMT and 2158 GMT each day. The temperature forecasts contained in the FP4's are for five 12-hour periods (see table one). So, for example, and FP4 issued at 0958 GMT Monday will give the forecast high for Monday (1st period), the low Monday night and Tuesday morning (2nd period), the high Tuesday (3rd period), the low Tuesday night and Wednesday morning (4th period) and the high Wednesday (5th period).

Temperature forecasts were recorded from the FP4's for Milwaukee for the period April 1978 thru March 1979. Some of these were not available for various reasons, so for some days, no 5th period forecasts were obtained, while forecasts from other days were missing entirely.

MOS final guidance temperatures were examined for the same period of time. To explain briefy, the MOS technique used in generating the guidance combines numerical and statistical data. Observed high and low temperatures have been correlated with various combinations of observed weather elements, numerical weather forecasts and climatic factors to derive forecast equations. The early guidance is based on forecasts from the Limited-area Fine Mesh (LFM) model. Final guidance is based on output from the hemispheric Primitive Equation (PE) model (2). (Editor's note: the PE model has since been replaced by the Spectral Model.)

The final guidance was chosen rather than the early guidance for two reasons. The first and most important was that the early guidance is available for only the first four periods. Second, the final guidance temperatures had been shown to be superior to early guidance until recently. For the period of time discussed in this paper, early guidance was shown to have been better than the final guidance for the country as a whole (3, 4). It's not known if this was also true for Milwaukee or what the magnitude of the difference between them was at Milwaukee.

Guidance forecasts derived from the 0000 GMT data correspond to 0958 GMT FP4, while the 1200 GMT MOS forecasts correspond to the 2158 GMT FP4. A few MOS forecasts were missing each month. For the sake of convenience and brevity, all of the guidance forecasts are occasionally referred to as "MOS forecasts", although it should be understood that the 5th period temperatures are not derived by the MOS technique. The 5th period forecasts are based on the "perfect prog" technique (5).

It should be noted that while the FP4's give forecasts for 12-hour periods, the MOS guidance temperatures are for 24-hour periods. In spite of this difference, the two sets of forecasts are quite comparable since the great majority of highs and lows at Milwaukee occur within the time frame designated by the FP4. Besides, the usefulness of MOS guidance in preparing the FP4 depends on how well it corresponds to the 12-hour periods.

Time Issued	1st Period	2nd Period	3rd Period	4th Period	5th Period
0958 GMT (00 GMT Data)	Max 12-00 GMT	Min 00-12 GMT	Max 12-00 GMT	Min 00-12 GMT	Max 12-00 GMT
2158 GMT (12 GMT Data)	Min 00-12 GMT	Max 12-00 GMT	Min 00-12 GMT	Max 12-00 GMT	Min 00-12 GMT

Table One

Time spans covered by FP4 forecasts.

The highs and lows used for verification were the official readings obtained at Milwaukee's General Mitchell Field for the same time periods as are used for the FP4's. Mitchell Field is located about three miles west of Lake Michigan and about eight miles to the south of downtown Milwaukee.

3. METHOD OF ANALYSIS

Each FP4 forecast and guidance forecast was compared against the official observed temperature and the errors recorded. In cases where both guidance and FP4's were available, the forecasters' improvement over guidance was calculated. An improvement is a positive number, while a negative value indicates that the forecaster did worse than the machine.

The number of cases falling into the different error or improvement categories for each period were tallied and recorded. The totals obtained were then assembled into Tables for each season (Tables 2 thru 13).

Error intervals of 5 degrees F were chosen as the most convenient and informative for displaying the data. Everyone strives for a zero error, but errors of 1 to 5 degrees are generally acceptable. In the range of 6 to 10 degrees, the errors become unacceptable and quite noticeable to the public using the forecast. A forecast that is off by 11 degrees or more is not very useful.

The seasons used here are: Spring - March, April, May; Summer - June, July, August; Fall - September, October, November; and Winter - December, January, February. The same months were used in deriving the four seasonal equations for the final MOS temperature guidance (6).

4. ANALYSIS OF THE TEMPERATURE FORECASTS BY SEASON

a. Spring

Table 2 show the error distribution of the FP4 maximum and minimum temperature forecasts for each period during the spring. The distribution for maximum temperature forecasts is quite even for the 1st period, but subsequent periods show a tendency to overforecast more often than to underforecast. This is particularly evident in the 4th and 5th periods.

Meanwhile, the minima are shown to have been underforecast (i.e. the forecasts have been too low) more often than overforecast in all periods - even the first. Again, the bias was slightly greater in the 4th and 5th periods than in the earlier ones.

The spring guidance forecasts (Table 3) showed the same tendency toward overforecasting the highs as the forecasters did. In fact, the bias was somewhat greater for the guidance temperatures. The one exception was the 4th period, where the high was underforecast more than it was overforecast.

Guidance sharply underforecast the low temperatures in the first period, and the subsequent periods showed a similar bias of smaller magnitude.

It should be pointed out here that the tables for the FP4 forecasts and the guidance forecasts are not exactly comparable. Some FP4's were missing when guidance was available, and vice versa, so the two groups of forecasts do not cover the same days exactly. While the differences are probably small, this should be kept in mind.

Table 4 shows the forecaster's improvement over guidance. The most consistent improvements on the maximum temperatures were made not in the 1st, but in the 2nd period. 58 percent of the 2nd period guidance forecasts were improved upon while only 24 percent were made worse. The other periods showed about 10 to 15 percent more forecasts were made better than were made worse. And the greatest number improved by 6 degrees or more were in the 5th period.

The minimum temperatures show similar improvements, although both the improvements and changes for the worse were greatly concentrated in the 1 to 5 degree range.

b. Summer

Table 5 shows the error distribution for the summer FP4 forecasts. The tendency to forecast maximum temperatures that are too high was even more apparent in the summer than in the spring. The overall trend was to overforecast most in the periods farthest out. In the 5th period, nearly 70 percent of the maximum temperatures forecast were too high, while only about 20 percent were too low. For all five periods combined, twice as many forecasts that were too high were issued as compared to those that were too low.

On the other hand, the minimum temperature forecasts tended to be on the low side, just as they were in the spring. In contrast to the highs, the most underforecasting of the lows was done not in the 5th, but in the 1st period. In fact, in the 4th and 5th periods, the number of lows underforecast and over-forecast were nearly equal.

The guidance forecasts (Table 6) for the summer highs show mixed results with the 1st, 3rd and 5th periods showing a trend toward overforecasting, while in the 2nd and 4th periods, the highs were underforecast. This sort of pattern seems a bit strange until one realizes that all of the 1st, 3rd and 5th period highs are from guidance based on 0000 GMT data, while the 2nd and 4th are based on 1200 GMT data. Thus, the 1200 GMT MOS underforecast the highs while the 0000 GMT guidance overforecast them - especially in the 5th period. The difference may be due to the fact that the forecasts derived from the 0000 GMT data are strongly influenced by the maximum temperature reported at that hour, while those made from the 1200 GMT data are not.

The low temperature guidance forecasts were more consistent, with a tendency to under-forecast shown in all periods - particularly the 1st.

The improvement over guidance (Table 7) was considerably less for the summer than for the spring. In fact, more high temperature forecasts were made worse than were improved in the 3rd and 4th periods, and in the 5th, the two were about even. Over half of the forecasts for the 5th period showed zero improvement. A zero improvement is achieved when either the forecast value is identical to the guidance value, or when the forecaster's error is equal in magnitude, but opposite in sign, to the guidance forecast. In the overwhelming majority of the cases, the zero improvement was due to the FP4 and MOS values being the same. Thus, the forecasters showed a great reluctance to change the guidance for the 5th period maximum. In fact, the sum for all the periods shows that about 40 percent of all the forecasts showed zero improvement. Again, a large majority of these were due to the FP4 values being identical to the guidance.

Fairly good improvement was made on the minimum forecasts for the first three periods, but the 4th and 5th periods showed negative improvements. Again, a sizeable percentage of the MOS forecasts were untouched.

c. Fall

The forecasters' errors for autumn (Table 8) again show overforecasting of the highs, especially in the 1st and 5th periods. Even more obvious is the underforecasting of the lows. In nearly all of the periods, twice as many minimum forecasts were on the low side as were too high.

Guidance (Table 9) gave opposite results for the highs, with a good deal of underforecasting being shown in the 1st through 3rd periods. It is less pronounced in the 4th period, and finally reversed in the 5th. On the other hand, the lows forecast by MOS again show a very strong bias toward the cold side.

A comparison of the FP4's and MOS (Table 10) shows that the forecasters equaled or did worse than guidance in forecasting the the autumn highs. Better results were shown in the lows, especially in the first period.

d. Winter

Again, the forecasters showed a strong tendency to overforecast the highs and underforecast the lows (Table 11).

This result is somewhat surprising in view of the large negative temperature anomaly shown for the winter season (-.4 degrees F in December, -7.8 degrees in January and -7.4 degrees in February). One would think that the extreme cold would have led a person to forecast both highs and lows that were too high (since the forecaster would lean toward climatology and look for any possible warm-up), or highs and lows that were too low (since the forecaster would finally begin to "think cold" due to the persistence of the pattern). In actuality, what seems to have happened is that the forecasters adjusted quite nicely to the persistent cold pattern, and thus showe the same tendencies that they showed during periods of "normal" weather.

Guidance for the winter season (Table 12) again showed a bias toward overforecasting the highs, with the inexplicable exception of the 3rd period. Among the four seasons, the bias was strongest in the winter, in spite of the 3rd period's results. Once again, the lows were greatly underforecast.

It should be noted here that the final guidance temperatures in December, January and February were contaminated by an error in the 7 layer PE-based trajectory model. The amount of deterioration caused by this error is not known (4).

The forecasters showed some improvement over the guidance maximum temperature forecasts (Table 13) in the 1st, 4th and 5th periods, but the FP4's were worse than MOS in the 2nd and 3rd. The greatest improvement was in the 4th period. The lows were improved upon in all periods, with the best being the 1st and 3rd.

5. Periodic Errors of the Guidance Forecasts

In the course of analyzing the year's worth of data, it was noticed that guidance would at times drastically underforecast or overforecast both the highs and lows for periods as long as a week and a half. Table 14 gives an example from late spring of 1978. Out of 100 forecasts for the dates given above the line, only 15 percent showed positive errors. 85 percent of the errors were negative or zero, with 16 percent showing errors of -10 degrees F or more. Over half had errors of -5 degrees F or more. The great majority of the positive errors were only one or two degrees.

The departures from normal of the daily average temperatures (right-hand column) were generally +5 degrees F or greater during the period that MOS showed the negative errors. As the temperature anomalies became smaller and occasionally negative, the guidance forecasts began to break out of their pattern of forecasts that were too cold, and the errors became more mixed (bottom portion of Table 14).

A few other instances similar to the one above were also found, although none persisted as long as this case from late May. Most featured negative errors (i.e. forecasts that were too cold), but some showed positive errors. Table 15 shows one such case. This occurred during a very cold period in early January of 1979 and lasted about nine days. 85 percent of the forecasts for the dates shown above the line were too high and 31 percent had errors of +10 degrees F or more. The daily average temperature anomalies were often -20 degrees or more during the period. Once again the guidance errors became less consistent as the departures from normal became closer to zero.

The two examples given would lead one to suspect that long periods of MOS forecasts that are too high or too low are associated with periods of temperatures well above or below normal. Indeed, most of the cases that were examined (not presented here) showed this to be true. Not all of them did, however.

Table 16 shows a 6-day period in mid-October of 1978 where a little over 80 percent of the guidance forecasts were too low. But the departures from normal of the temperature during that period were negative for the first three days, and did not show large positive departures until toward the end of the six-day period.

In view of the above discussion, a forecaster can make large improvements upon guidance if he realizes that he is in one of the periods where MOS is running consistently too high or too low. The trick, of course, is for the forecaster to know when guidance has entered one of these periods, and when the period is coming to an end. No hard and fast rules could be derived from the limited sample examined. Tentatively, it can be said that guidance has probably gone into one of these periods whenever a stretch of unseasonably warm or cold weather has set in and the MOS forecasts have demonstrated a strong tendency toward over- or underforecasting most of the temperatures for a couple of days. The period is likely to come to an end when the departures from normal come back toward zero or change sign or the guidance forecast errors become mixed in sign again. In most cases, the transition was gradual and quite similar to that shown in the lower portions of Tables 14 and 15.

Long periods of MOS forecasts that were consistently too high or too low appeared rather infrequently during the one year examined here. Shorter periods occurred much more frequently, but would have been difficult for a forecaster to detect until they had already run their course. Sometimes several short periods occurred within a time span of a month or so, and at times, a long period would be followed by several short periods. The short periods showed the same positive or negative error tendencies as the long ones preceding them.

6. SUMMARY AND COMMENTS

Analysis of both machine and man-made temperature forecasts for one year at Milwaukee has shown the following:

1. The forecasters had a distinct tendency to overforecast the high temperatures. This was true for nearly every forecast period for every season. The tendency was most pronounced in the summer and winter and least pronounced in the spring.

2. The forecasters also put out minimum temperature forecasts that were too low about 20 percent more often than forecasts that were too high. This was most pronounced in the fall and winter.

3. For the year as a whole, guidance tended to overforecast the highs, but not as consistently as the forecasters did. In the winter season, guidance max temperature forecasts were too high nearly 30 percent more often than they were too low. Yet in the autumn, the highs were underforecast in 4 out of 5 periods. Spring and summer showed mixed results, but overall the highs were overforecast in those two seasons.

4. Guidance underforecast the lows in all periods during all four seasons. The worst were fall and winter, when minimum forecasts that were too low occurred about 40 percent more often than those that were too high.

5. The forecasters showed their overall best improvement over the guidance maximum temperature forecasts in the spring, but did worse than guidance in the fall. The FP4's were consistently and considerably better