VERIFICATION OF MRF-BASED STATISTICAL GUIDANCE FOR MOLINE, ILLINOIS

James D. Hladik and Ray A. Wolf
NOAA/National Weather Service Forecast Office
Davenport (Quad Cities), Iowa

1. Introduction

Medium range statistical guidance providing forecasts of several meteorological variables through 192 hours was implemented by the National Weather Service in December, 1992 (Jensenius et al. 1993, 1995). Initially, a perfect prog approach was used to develop the statistical equations, but by the mid 1990s, this approach was improved upon by use of the MOS technique (Erickson and Carroll 1999). The statistical guidance is based on output from the National Center for Environmental Prediction (NCEP) Medium Range Forecast (MRF) 0000 UTC model cycle.

The MRF was upgraded in April 2002, and is now known as the Global Spectral Model (GFS) within the Global Forecast System. In parallel testing, this upgrade had minimal impact on MOS output (see www.nws.noaa.gov/mdl/synop/gfs/mrfmos.html). Thus, findings from this study based on MRF output should still be valid for GSM output despite changes in the numerical modeling system.

Since the implementation of MRF statistical guidance (hereafter FMR), NWS forecasts with leads time of seven days have become available to the general public. Since the FMR is one of the inputs into the medium range forecast process, an understanding of the skill level of these forecasts is important to using them properly to convey the most accurate forecast with the appropriate level of confidence and detail.

This paper verifies FMR temperature and 24-hour probability of precipitation forecasts (POP) through 192 hours for Moline, Illinois. Section 2 discusses the methodology of the assessment. Sections 3 and 4 detail the findings for temperature and POP forecasts, respectively. Section 5 summarizes the conclusions of the study, including considerations for the forecast process.

2. Methodology

Maximum and minimum temperatures, precipitation data, and FMR output for Moline, Illinois (MLI) were collected for the year 1999. Statistics were generated for each daily forecast period during the entire year, then collated into seasonal components (spring - March, April, May; summer - June, July, August; fall - September, October, November; and winter - December, January, February).

The study includes an assessment of maximum forecast temperatures for Day 1 and the maximum and minimum forecast temperatures for days 2 through 8. Daily precipitation forecasts (POPs) were evaluated for days 2 through 8 and in two ways; first, assuming trace observations are zero (*trace=0*), as used in the development of the forecast equations; and second, assuming trace events as measurable precipitation (*trace=measurable*). This was done for comparison purposes. In addition, climatological values of temperature and POPs, which are included with the FMR product, were also assessed for comparison to forecast values.

Statistics calculated for temperature forecasts include mean absolute error (MAE) (Eq. 1.0), bias (Eq. 1.1), and standard deviation (Wilks 1995). Statistics calculated for POP forecasts include the Brier Score (BS) (Eq. 1.2) and Skill Score using climatology as a reference (Eq. 1.3) (Wilks 1995).

Where F equals a forecast value and O equals the observed value,

$$MAE = |F - O| \tag{1.0}$$

$$\text{Bias} = F - O \tag{1.1}$$

$$\text{Brier score} = 1/n \sum (F - O)^2 \tag{1.2}$$

$$\text{Skill score} = 1 - (BS/BS_{ref}) \tag{1.3}$$

where BS_{ref} is the climatological Brier score.

The MLI ASOS is representative of the surrounding area and has no unusual local effects. ASOS does have known problems measuring frozen precipitation. These errors have been manually quality controlled and corrected in real time. Thus, the findings derived from the data set should be representative of the area.

3. Temperatures

a. Mean Absolute Error (MAE)

Looking at the data by seasons, spring temperature forecasts (Fig. 1) showed a gradual increase in MAE from day 1 through 8. Days 1 through 4 had reasonable forecast errors between about 2 and 4 degrees. The average error increased to between 5 and 6 degrees for days 5 through 8. The largest MAE's occurred where expected, in the later periods, with forecast maximum temperatures for days 6 through 8 having the highest errors. Interestingly, the climate maximum temperature had the greatest error (over 7 degrees), which is not unexpected since this period averaged warmer than normal (Fig. 2).

The summer period (Fig. 1) was similar in trend to spring, but with lower MAEs. An average error of less than 4 degrees extended through day 5, after which MAEs climbed to around 5 degrees by days 7 and 8. Performance was relatively good as would be expected in a forecast season with fewer extremes and air mass changes.

The fall forecast period (Fig. 1), as a transitional season with changeable weather, had much higher errors than summer, especially after day 3. Days 1 through 3 had an MAE of around 4 degrees, but starting with day 3, FMR forecast maximum temperatures showed large increases in error with day 7 maximums off over 8 degrees. Forecast maximum temperature MAEs for days 5 through 8 averaged 7.5 to 8 degrees. The climatological maximum temperature MAE was the highest overall for the fall forecast period, over 10 degrees. Forecast minimum temperatures had a relatively smaller error, but still rose to 6 degrees for days 7 and 8.

FMR MAEs were poorest during the winter season (Fig. 1) as a whole for days 1 through 8, accounting for both forecast maximums and minimums. An MAE around 4 degrees for days 1 and 2 maximum temperatures grew into an average error of close to 8 degrees by days 7 and 8, while MAEs for minimum temperatures averaged between 6 and 8 degrees for days 2 through 8. Again climatology had a large MAE of over 10 degrees. This may be in part due to a milder than normal winter season, especially in February (Fig. 2). Large forecast errors, especially in later periods, could result in some part to the output being overly weighted toward climatology.

Averaging the four seasons into annual values (Fig. 1), the MAE for maximum forecast temperature for day 1, and the max/min for days 2 and 3 showed a fairly good score with an average error 3 to 4 degrees. However at the end of the period, days 6 through 8 showed MAEs of 5 to 7 degrees. FMR forecast temperatures had an average 5 degree MAE in the heart of the medium range forecast period for days 4 and 5. The annual MAEs were biased by high errors from days 4 through 8 in the fall and winter seasons.

b. Standard Deviation

The standard deviation of FMR temperature forecast error (Fig. 3) showed the same general trend as MAEs. Summer forecasts had with the least variability with an average 3 to 5 degree spread from Day 1 through 8. In other

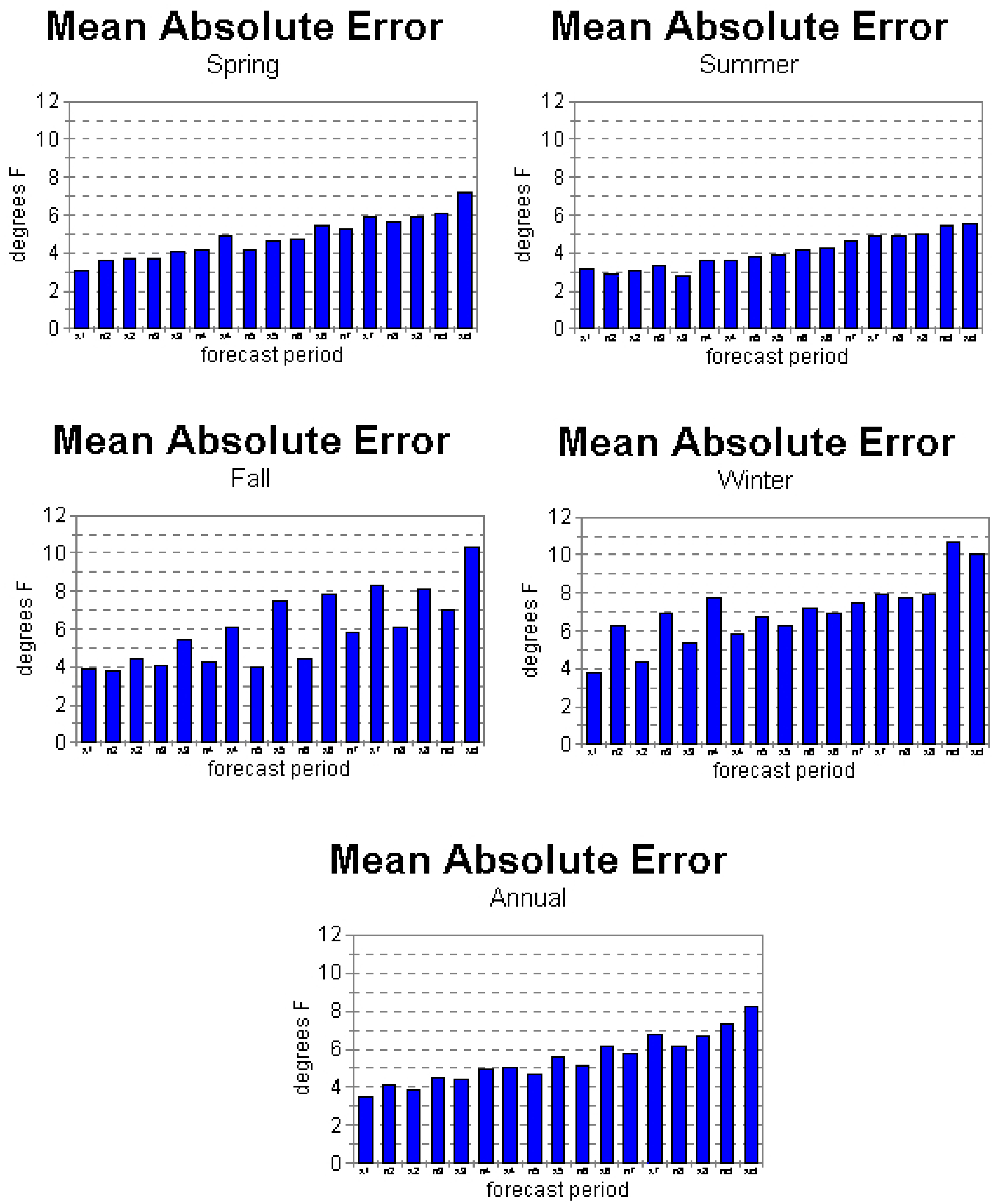

Fig. 1. Maximum and minimum temperature mean absolute error (°F) for days 1 through 8 and climatology. From upper left, spring and summer, fall and winter, annual.

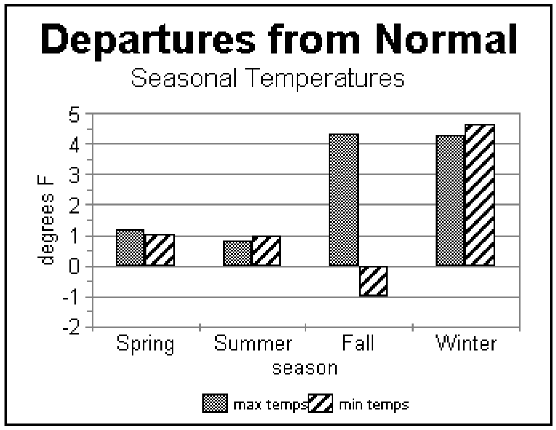

Fig. 2. Seasonal temperature departure from normal (°F). Solid = maximum temperature and striped = minimum temperature.

words, two-thirds of the summer forecast errors were 5 degrees or less and one-third were greater than 5 degrees. The spreads were greater during the fall and winter seasons, with both seasons showing increasing spread for days 1 through 8. Days 1 and 2 showed a standard deviation of just over 5 degrees, which climbed to over 9 degrees by day 8. Again, high standard deviations in fall and winter skewed the annual values with days 1 and 2 averaging around 5 degrees which climbed to over 8 degrees by Day 8. Not surprisingly, the spread of forecast error was higher in seasons with more transitional weather, when the potential for large misses in forecast temperatures is highest.

c. Bias

Looking at the seasonal breakdown, spring (Fig. 4) showed a rather consistent small, positive bias from days 2 through 8. This indicated that the FMR was on average 0.5 to 2 degree too warm for both forecast maximum and minimum temperatures in March, April, and May, with a near 2 degree warm bias for days 3 to 5. Even the climatological minimum and maximum forecasts showed at least a 0.5 degree warm bias, the only season in 1999 where the climatological forecasts for the period showed a positive bias. This is interesting since the 1999 spring season at Moline was slightly above normal temperature-wise overall, and yet the FMR still had a modest warm bias.

The summer season bias (Fig. 4) from forecast days 1 through 5 was generally 1 to 2 degrees negative (cool) for both highs and lows, with days 3 through 5 showing only a slight negative bias on the maximums, and almost a 1 degree warm bias on the minimums. In general, temperatures showed little consistent bias in this period. For days 6 through 8, both forecast maximums and minimums had a general 1 to 2 degree warm bias. This may be attributable to the fact that in general, the MRF systematically does not bring cold fronts far enough south across the United States late in the forecast cycle during the summer months. The summer climate forecasts showed a general 1 degree negative bias. Compared to all seasons, FMR was had the lowest bias in summer for days 1 through 8, which would be expected for a relatively unchanging season where temperatures averaged near or slightly above normal for the three month period (Fig. 2).

The fall season averaged warmer than normal at Moline for maximum temperatures but below normal for minimum temperatures (Fig. 2). This may have been one factor contributing to the rather poor performance by the FMR this forecast season, especially on maximum temperatures after day 3 (Fig. 4). FMR forecast maximum temperatures for days 4 through 8 averaged 4 to 5 degree cooler than observed as FMR consistently underforecast the maximum temperatures. Interestingly, the FMR had only a 1 to 2 degree warm bias on the minimum temperatures from day 2 through 8, with the magnitude of the bias decreasing after day 4. With a mild late fall, FMR guidance did not respond accordingly, trending its forecast maximum temperatures toward climatology. This was apparent as the climate maximum had the largest negative bias of all, tallying over 6 degrees too cool for the fall season. However, the climatological minimum showed just about a degree of negative bias, even though minimum temperatures

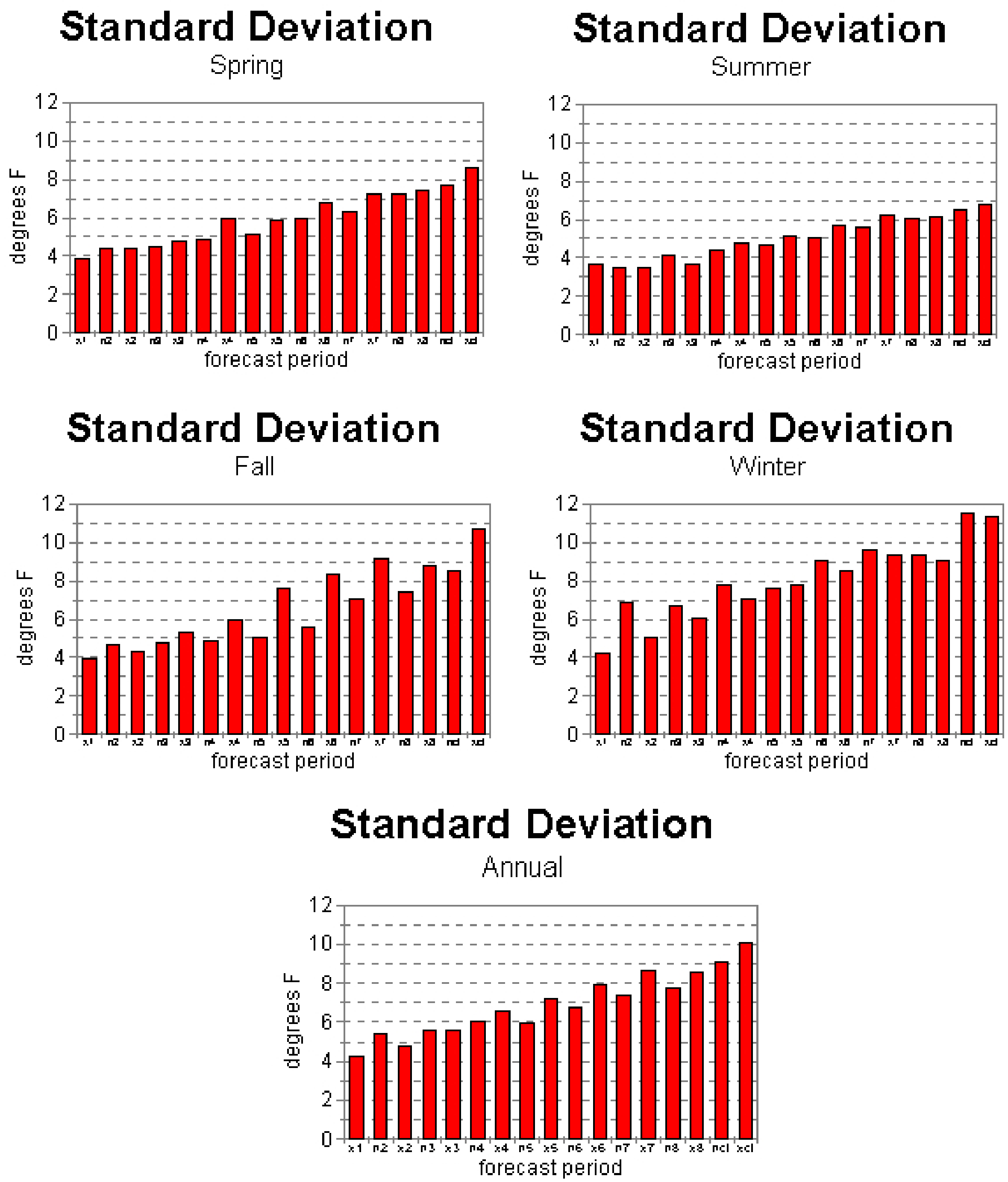

Fig. 3. Standard deviation of temperature forecast error (°F) for days 1 through 8 and climatology. From upper left, spring and summer, fall and winter, annual.

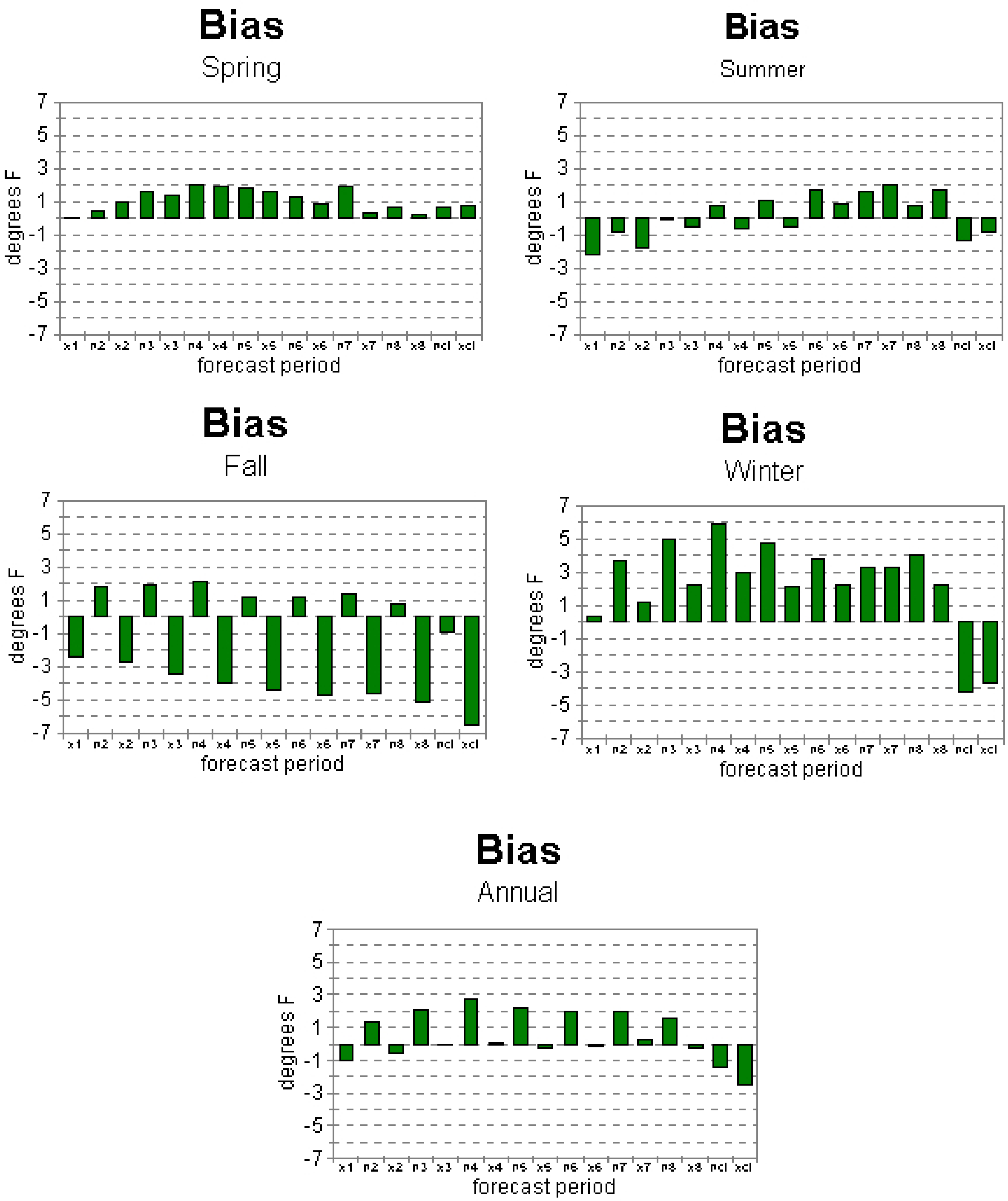

Fig. 4. Forecast temperature bias (°F) for days 1 through 8 and climatology. From upper left, spring and summer, fall and winter, annual.

averaged below normal.

For the winter months (Fig. 4), the FMR temperature forecasts showed a strong warm bias from days 2 through 8. Forecast minimum temperatures were especially prone to a positive (warm) bias error, with forecast minimums for day 4 averaging almost 6 degrees too warm. The forecast minimum temperatures averaged 3.5 to 4.5 degrees too warm through the eight day forecast period, while the forecast maximums averaged between 2 and 3 degrees too warm. The winter season showed the largest one-way bias of any season. This is particularly interesting since the winter months (especially February) were above normal (Fig. 2). Only January was close to its normal monthly average temperature. However, the winter season climatological temperatures showed close to a 4 degree cold bias. But yet even with a milder than normal season, it seems that the FMR over-compensated and had a warm forecast bias. Even in the later parts of the forecast period when the model should trend toward the climatology, FMR still showed a 3 degree warm bias. This suggests that during an abnormally warm or cool season, FMR guidance will not necessarily trend opposite the climate anomaly.

Looking at the FMR forecast year as a whole (Fig. 4), the forecast maximum temperatures averaged close to a zero bias from days 1 through 8. This is due to the large negative maximum temperature bias of the fall season being balanced by positive biases of the spring and winter seasons. Forecast minimum temperatures through the eight day forecast period reflected a positive bias. The FMR averaged almost a 2 degree warm bias for minimum temperatures from day 2 through 8 through the year. The 1 to 2 degree negative bias of the climatological maximums and minimums illustrates the warmer climate regime in place during most of 1999.

4. Probability of Precipitation

a. Brier Score

Overall, FMR guidance performed generally well for precipitation forecasting in a year when precipitation averaged below normal in each season (Fig. 5). Brier scores range from 0 to 1 with 0 being a perfect score. Most seasonal scores for FMR forecasts were 0.2 or less. The highest Brier scores for both *trace=measurable* and *trace=0* data sets (Figs. 6 and 7 respectively) occurred during the winter season followed by the summer, while the lowest (best) scores occurred during the very dry fall. Scores trended gradually higher as the lead time progressed from day 2 to 8. Annual Brier score were similar in trend and magnitude to seasonal scores (Figs. 6 and 7).

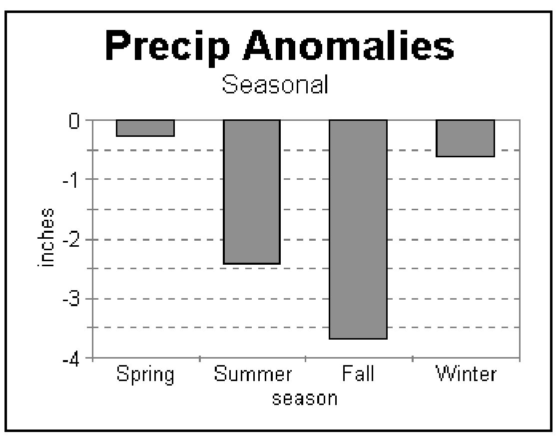

Fig. 5. Seasonal precipitation departure from normal (inches).

b. Skill Scores

For seasonal skill scores with *trace=measurable* events (Fig. 8), day 2 FMR showed a near 70% improvement over climate in spring which tailed off to only around a 5% improvement by days 7 and 8. For *trace=0* events (Fig. 9),

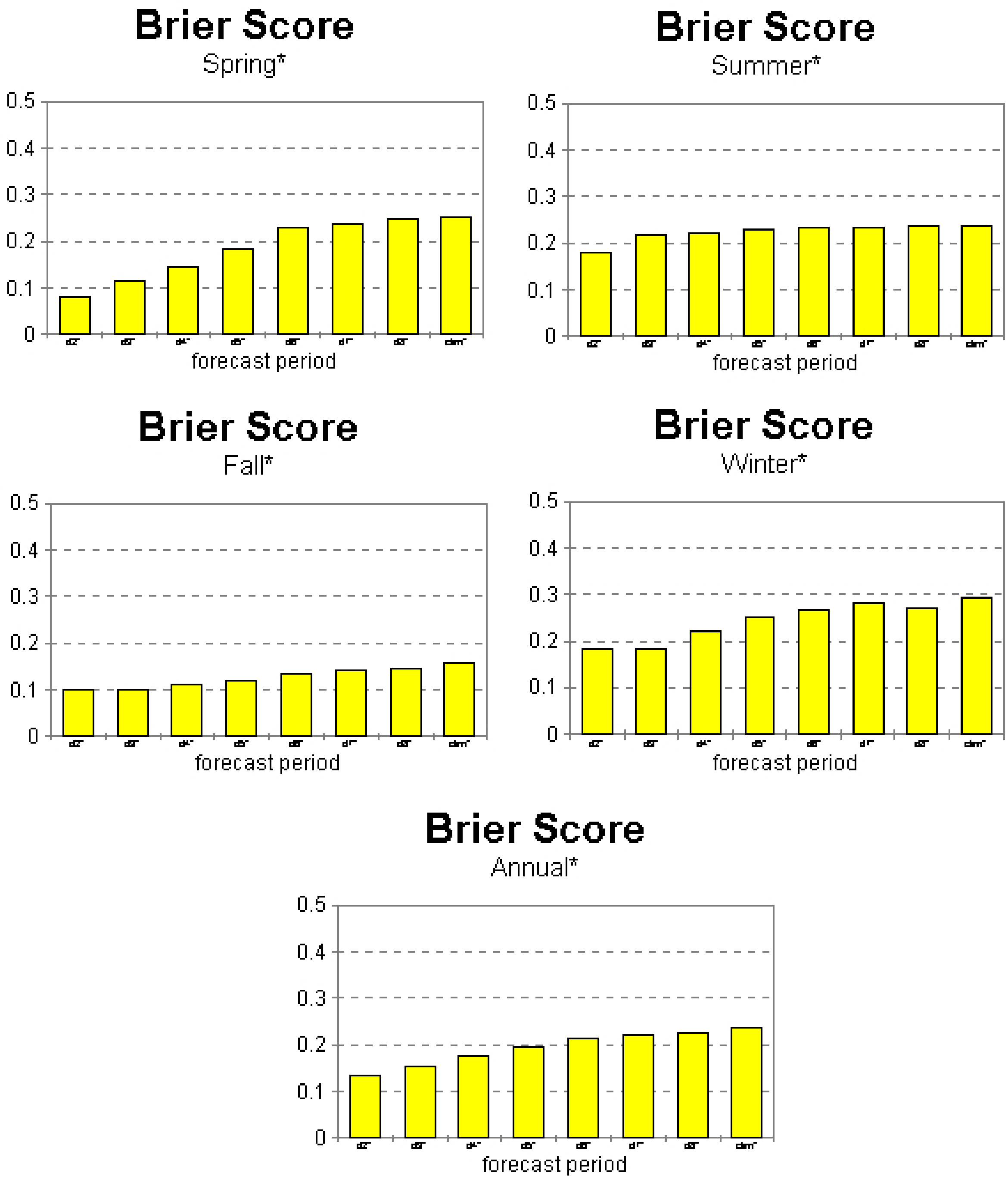

Fig. 6. Brier Score for *trace=measurable* events.

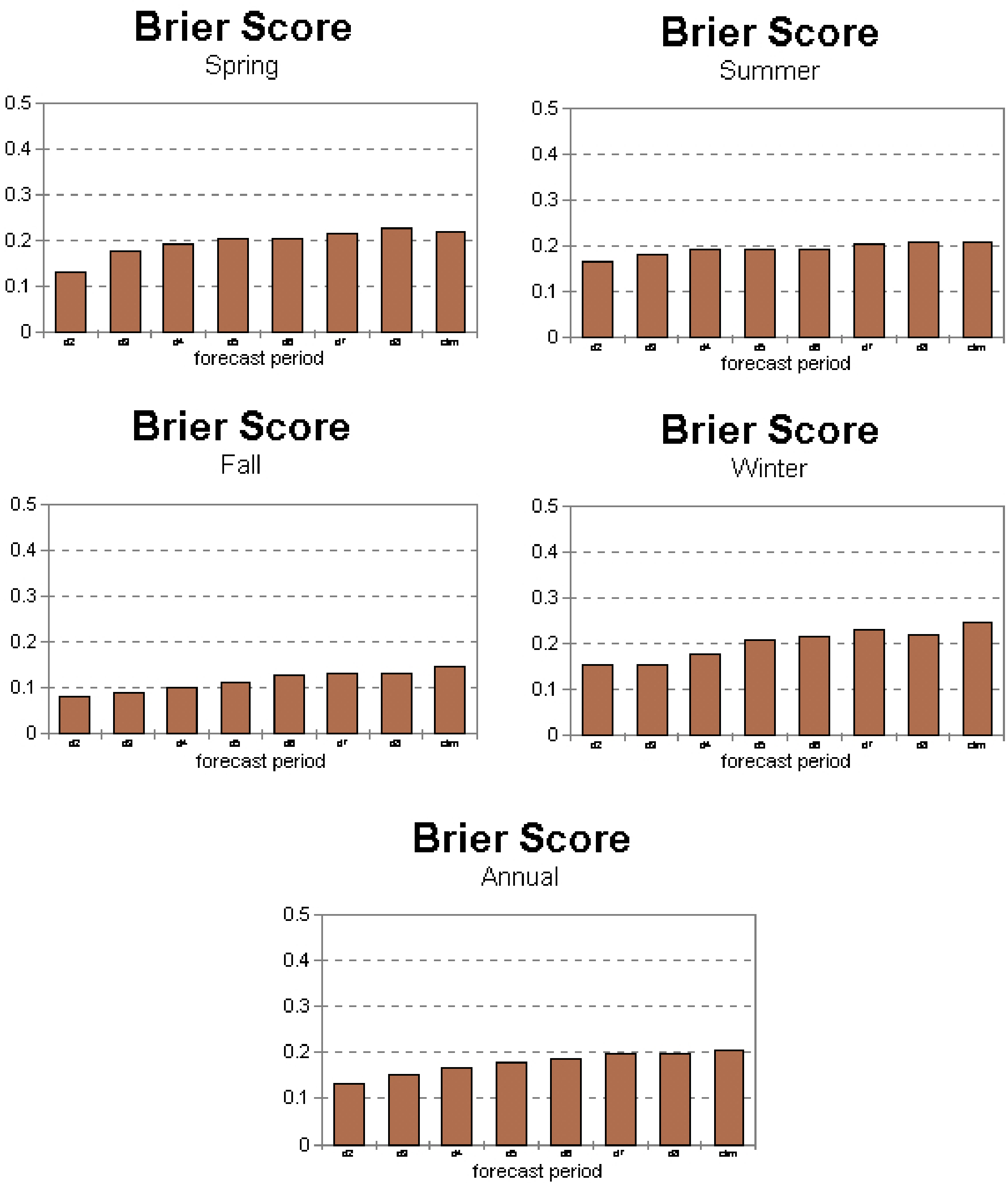

Fig. 7. Brier Score for *trace=0* events.

the results were less impressive at shorter lead times. FMR had a 35% improvement over climate on day 2, but then quickly trended down to only a 5% or less improvement by days 7 and 8. Days 2 through 4 showed a relatively large improvement over climate precipitation percentages for events greater than a trace. The POP equations were developed assuming trace events as zeroes. Also, unlike the temperature equations which were developed for a single station, the POP equations were developed for a regional set of locations (Jensenius et al. 1993).

Summer forecasts had lower skill scores over climatology, especially for *trace=0* events (Fig. 9). Day 2 forecasts had a 25% improvement over climatology, but quickly tailed off to 5% or less for day 4 and beyond. For *trace=measurable* events, FMR performed somewhat better for days 2 through 4 (Fig. 8), but overall, the summer season is when FMR had the least skill with respect to climatology and in comparison to other seasons.

Both fall and winter fell in between the spring and summer results and were similar to each other (Figs. 8 and 9). FMR scores showed a general 30-40% improvement over climatology for days 2 through 4 for both *trace=0* and *trace=measurable* events before decreasing to 10% or less from day 6 on. For *trace=measurable* events, both seasons mirrored each other with Skill Scores ranging close to a 35% improvement for day 2 which diminished to 5% or less improvement by days 7 and 8.

Annually (Figs. 8 and 9), FMR forecasts showed the same general trend in skill as in the seasonal breakdowns. Again, looking at both *trace=0* and *trace=measurable* events, FMR showed no negative skill. It also showed a decent of improvement of 20-40% in the forecast period for days 2 through 4, but then decreased to showing only little improvement over climate for days 7 and 8. Overall the *trace=measurable* data set had more forecast skill than the *trace=0* data set.

5. Summary

Several findings in this study are well known to experienced forecasters. By most measures, the skill of temperature and precipitation forecasts decreases with increasing lead time, showing marginal skill over climate by days 7 and 8. Findings from the MLI data were similar to the results discussed in Jensenius et al (1995) in this regard. Also, temperature forecasts in seasons of frequent air mass changes and/or climatic extreme tend to be less skillful and show a higher degree of spread in the forecast error (i.e., higher standard deviations) than forecasts during unchanging, near seasonal conditions.

Other findings are not as straightforward. Errors in forecast temperatures did not average to be of an opposite sign of the climatic anomaly (i.e., MOS forecasts were not too cool during unusually warm weather). In fact, forecast temperatures during unseasonably warm periods in 1999 still had a warm bias, not a cold bias as might be expected from the MOS approach which relies heavily on climatology. Moreover, the low Brier Scores for precipitation may be due to the relatively dry year in 1999. For example, forecasts with zero or very low POPs would consistently beat climatology on dry days, as was frequently the case in 1999. Interestingly, temperature forecasts had the lowest MAEs during the summer when precipitation forecasts showed the least skill. Precipitation forecasts showed the highest skill during the winter when temperature forecasts were least skillful (along with fall). FMR precipitation forecasts, in general, verified better when assuming a trace observation to be measurable as opposed to setting a trace equal to zero, as was done in the process of equation development. Most of this improvement occurred during the spring season with slightly below normal precipitation.

This study was confined to only one year of data. To increase the value of statistical forecasts and consequently improve medium range forecasts, this type of verification needs to be conducted on a continual basis with feedback routinely provided to forecasters. Moreover, a detailed look at occasions when large forecast misses occur should be conducted. And these occurrences should be tied into an assessment of synoptic and large scale patterns to help understand how model performance and thus statistical guidance skill varies with different flow regimes.

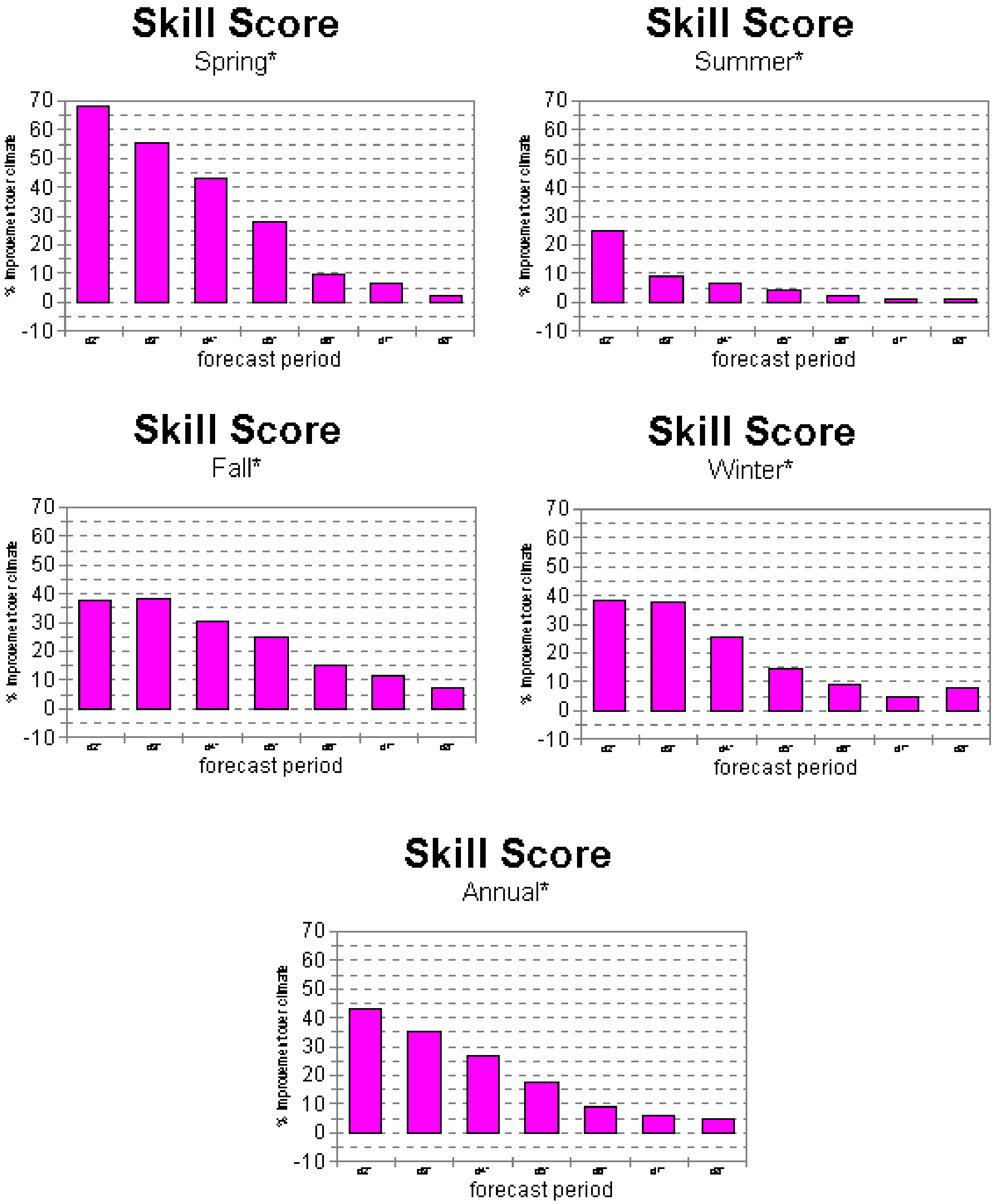

Fig. 8. Skill Score for *trace=measurable* events.

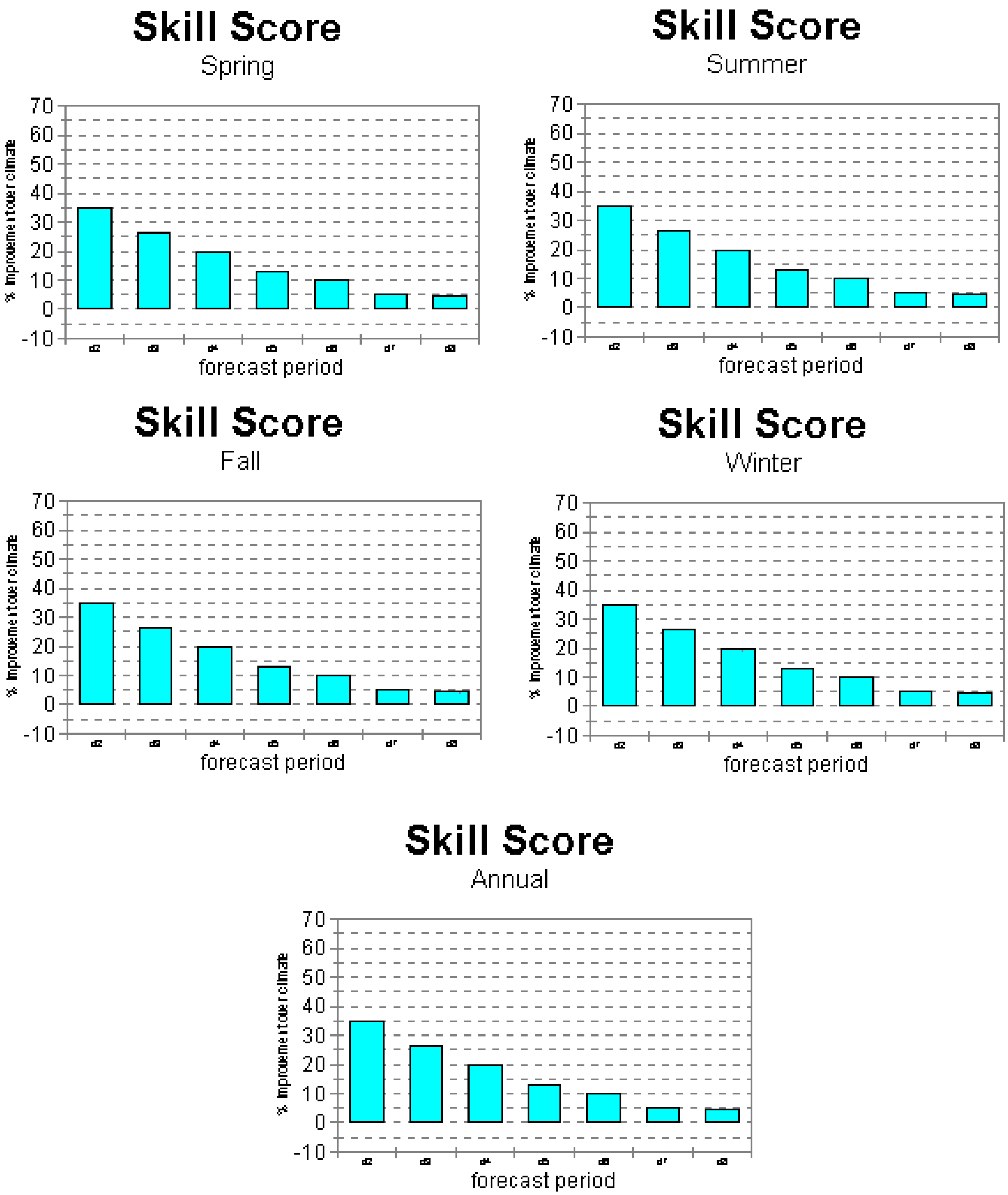

Fig. 9. Skill Score for *trace=0* events.

References

Erickson, M.C. and K.L. Carroll, 1999: Updated MRF-based MOS guidance: Another step in the evolution of objective medium-range forecasts. Preprints, *17th Conf. Weather Analysis and Forecasting*, Denver, CO, Amer. Meteor. Soc., 190-195.

Jensenius, Jr., J.S., J.P. Dallavalle, and S.A. Gilbert, 1993: The MRF-based statistical guidance message. NWS Technical Procedures Bulletin No. 411, NOAA, U.S. Dept. of Commerce, 14 pp.

Jensenius, Jr., J.S., K.K. Hughes, and J.B. Settelmaier, 1995: The National Weather Service's AVN- and MRF-based statistical weather forecast systems. Preprints, *14th Conf. Weather Analysis and Forecasting*, Dallas, TX, Amer. Meteor. Soc., 163-170.

Wilks, D.S. 1995: *Statistical Methods in the Atmospheric Sciences*, Academic Press, 467 pp.