Experiments in Temperature and Precipitation Forecasting for Illinois

JOHN R. GYAKUM

Department of Atmospheric Sciences, University of Illinois, Urbana, IL 61801

(Manuscript received 13 January 1986, in final form 4 April 1986)

ABSTRACT

Six years of daily temperature and precipitation forecasting are studied for Urbana, Illinois. Minimum temperature forecast skills, measured against a climatological control, are 57%, 48%, 34% and 20% for the respective forecast ranges of one, two, three, and four days. Maximum temperature skills are comparable. Precipitation probability skills of 29%, 19%, 6% and −2% are found for the same respective forecast ranges. However, our skill in predicting precipitation amount, given that a measurable quantity occurs, is only 17% at the first day range and negligible thereafter. An examination of objective National Weather Service (NWS) forecasts shows this guidance to be slightly less skillful than our consensus in forecasting temperature and precipitation. Some temporal improvement is found in both the consensus and guidance temperature forecasts, but none can be found in the more difficult problem of forecasting precipitation.

Significant warm and dry biases are frequently found in both our consensus and NWS guidance forecasts, especially during the summer season. These biases may be associated with the organized convective character of the precipitation in Illinois. Forecasts often miss these key events and, therefore, will often predict excessively warm maximum temperatures.

Finally, the results show that our consensus skill is comparable to the state of the art. Student or faculty individuals usually lose to our consensus, as does the NWS objective forecast guidance. This establishment of the consensus forecast as typically being superior to an individual forecast has been reported by investigators in eastern United States cities.

1. Introduction

This paper discusses results of a series of temperature and precipitation forecast experiments initiated in the spring 1980 semester at the University of Illinois Department of Atmospheric Sciences. Although recent studies have suggested long-term improvements in large-scale circulation prognoses at 500 mb (e.g., Bengtsson, 1985), surface temperature and precipitation forecast skill, as measured against a climatological control, has shown only limited improvement during the past 20 years (Sanders, 1979, 1986; Charba and Klein, 1980; Ramage, 1982; Bosart, 1983). However, recent studies by Glahn (1985) and Murphy and Sabin (1986) show considerably more temporal improvement in National Weather Service (NWS) forecasts nationwide. The purpose of this paper is to document temperature and precipitation skill for Urbana, Illinois, and possible trends in these skills. We will also compare and contrast these results with comparable results of other investigators.

2. Forecast details

The University of Illinois forecasts are made by faculty and students of the Department of Atmospheric Sciences. Daily (Monday through Friday) forecasts are typically made by five to ten graduate and undergraduate students and by one or two faculty members. A forecast deadline of 1700 local (central) time is observed. This corresponds to 2200 GMT when local daylight time is observed and to 2300 GMT when local standard time is used. Since 1980, forecasts of minimum and maximum temperature and precipitation probability have been made for each of four 24-h periods, beginning at 0600 GMT the following day. A climatological control forecast is established for these forecast categories each day on the basis of daily data taken from the Morrow Plots (Urbana) observing station since 1889. Forecasts of precipitation amounts for each of the following four calendar days have been made since the summer of 1983. The quantitative precipitation climatology is the median daily amount of precipitation occurring in a given month during the period 1951–80. The precipitation probability climatologies are established by mean monthly data, while the temperature controls are based upon smoothed daily data. The forecast results are obtained for each of the spring (15 January through 15 May), summer (16 May through 15 August), and fall (16 August through 20 December) semesters.

3. Consensus skill results

A consensus is found each day by averaging the individual forecasts of temperature, precipitation probability, and precipitation amount. Of course, National

---

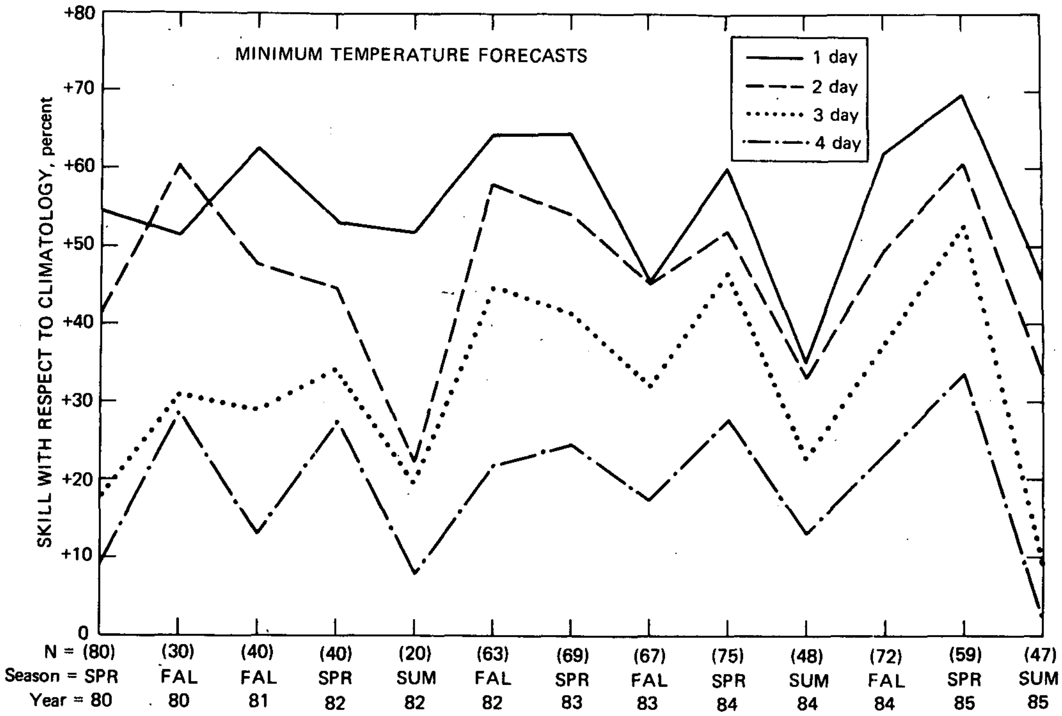


FIG. 1. Minimum temperature forecast consensus skill for ranges of 18–42 h (solid), 42–66 h (dashed), 66–90 h (dotted) and 90–114 h (dotted-dashed) after the 1200 GMT guidance time. Parentheses surround the number of consensus forecasts made for each semester.

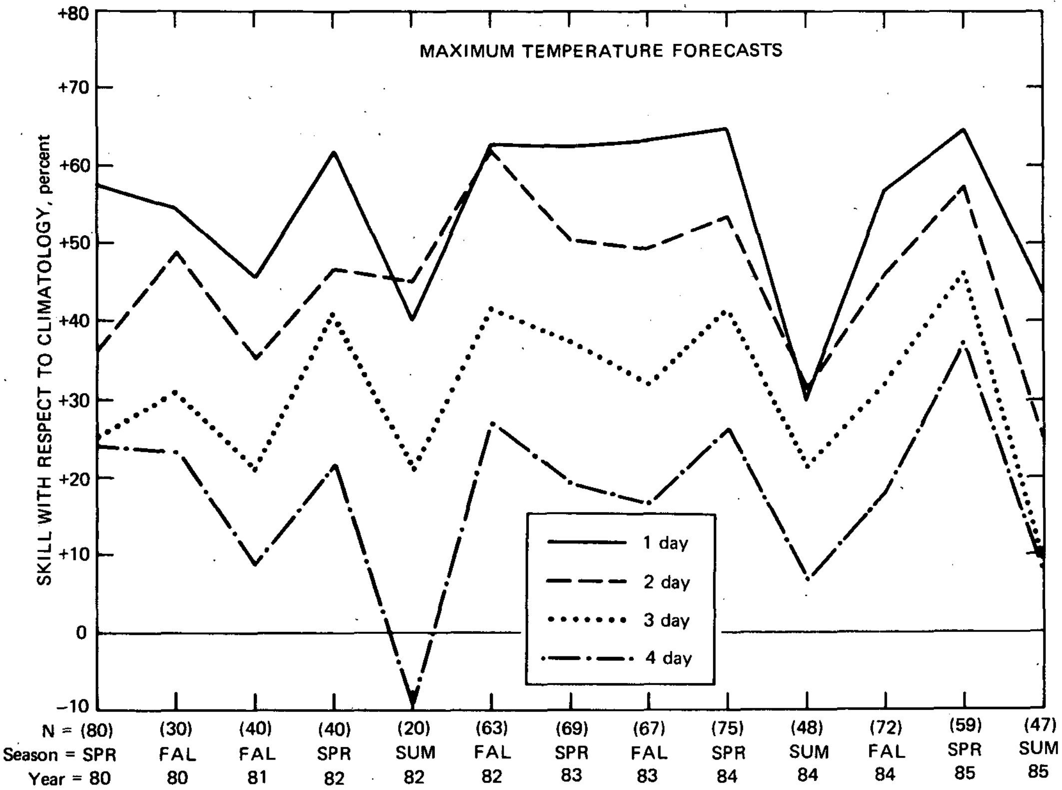


FIG. 2. As in Fig. 1 except for consensus maximum temperature forecasts.

Weather Service, persistence, and climatological forecasts are excluded from the consensus calculations. The skill of a consensus forecast (SK) with respect to climatology is defined as

$$\mathrm{SK} = \frac{\sum E_{CL} - \sum E_{CO}}{\sum E_{CL}} \times 100 \qquad (1)$$

where the summations denote the number of forecasts in a particular semester, and E_{CL} and E_{CO} are, respectively, the summed absolute value of climatology and consensus forecast errors. For example, the forecast error points, E_F, in the temperature category are defined as

$$E_F = |T_F - T_O| \qquad (2)$$

where T_F is the forecast, and T_O is the observed temperature.

Figures 1 and 2 show, respectively, the consensus skill against climatology in forecasting minimum and maximum temperatures. Temperature forecast results have been available continuously since the fall 1981 semester, except for the summer 1983 term, which was devoted to the introduction of the quantitative precipitation contest. The results clearly indicate a large semester-to-semester variability with a suggestion of lower scores during our limited summer term sample. Overall minimum temperature skills of 57%, 48%, 34% and 20% and maximum temperature skills of 57%, 46%, 32% and 20% are found for the respective forecast ranges of one, two, three and four days. The temperature results show that greater than 80% of the first day's skill is maintained into the second day, about 70% from the second to third days, and 60% from the third to fourth days. The minimum temperature skill levels are similar to those reported by Bosart (1983) and Sanders (1979), except that skills reported here at ranges of two through four days are as much as ten percentage points higher. Since the forecast deadline is 2300 GMT during the cold season and 2200 GMT during the warm season, we make full use of 1200 GMT guidance not used by Bosart and Sanders, as their forecasts are submitted by 1800 GMT. This guidance includes the National Meteorological Center's (NMC) three to five day subjective temperature guidance, along with the global spectral model (Sela, 1980) output extending to a 60-h range. We have few published results with which to compare our maximum temperature skills. However, Bosart's (1983) more limited sample showing 36-, 60- and 84-h forecasts as, respectively 49%, 34% and 19% are as many as 13 percentage points lower than results reported here. Rea-

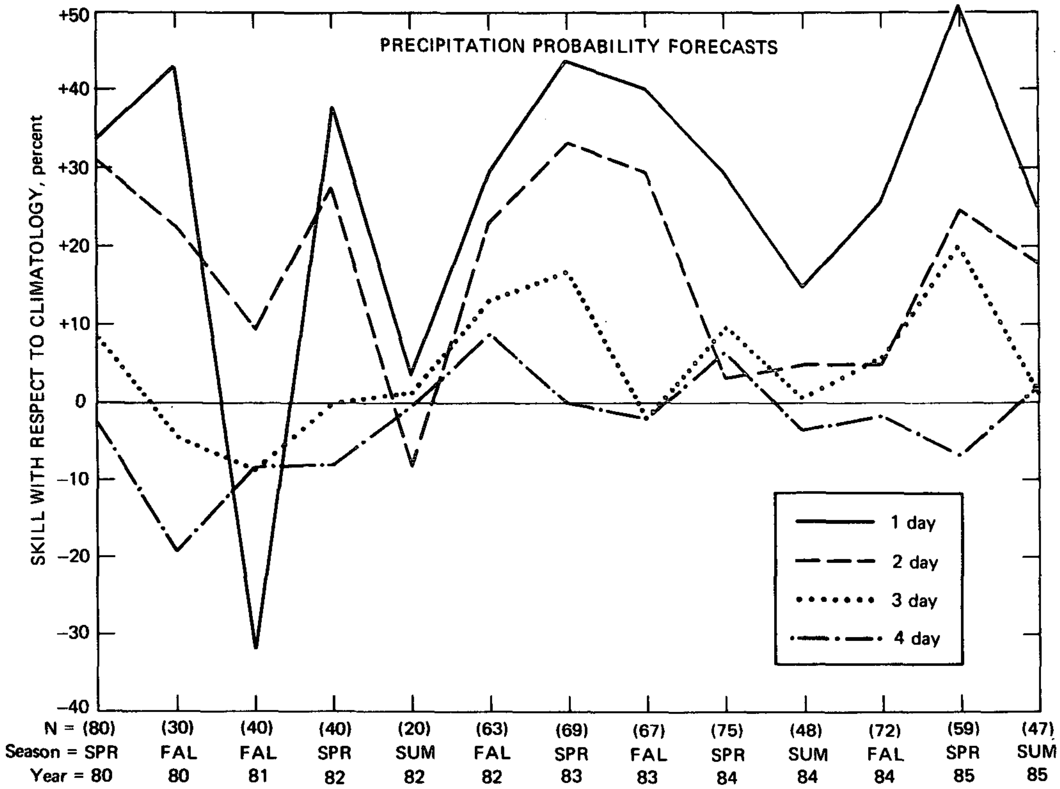


FIG. 3. As in Fig. 1 except for consensus precipitation probability forecasts.

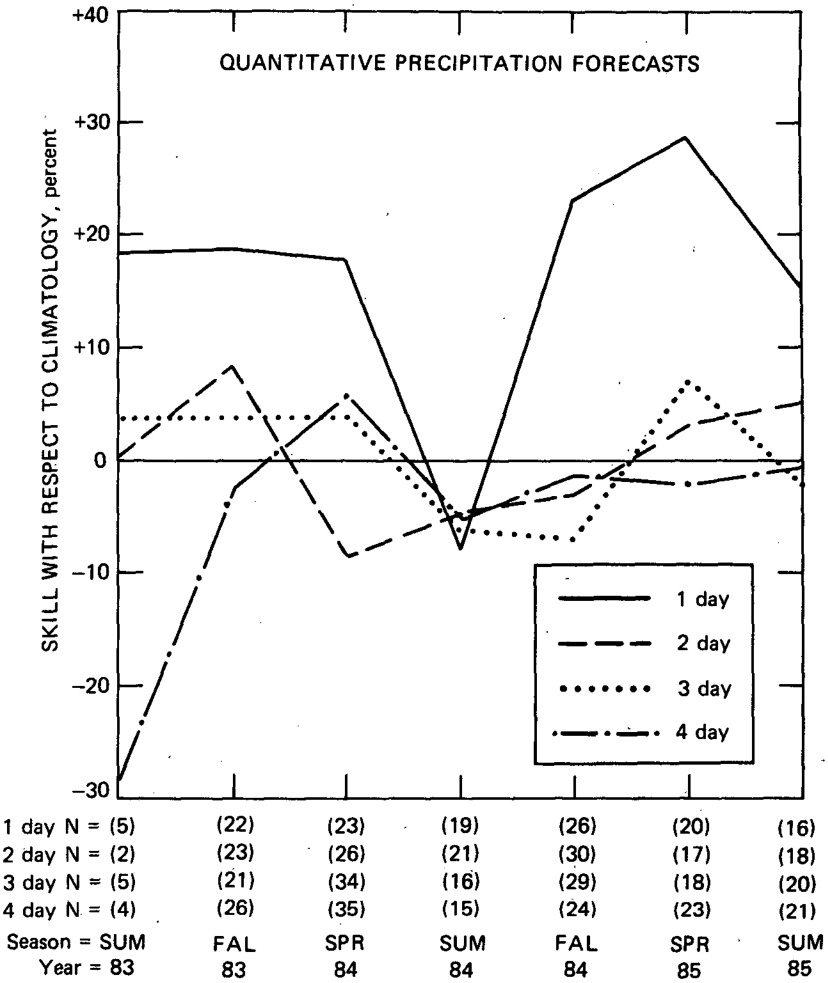


FIG. 4. As in Fig. 1 except for consensus quantitative precipitation forecast skill.

sons for these differences are unclear, although our later forecast deadline is possibly a factor in these differing results.

The precipitation probability contest involves the forecasting of probabilities of six classes of precipitation including none-trace, less than 1 mm, 1-3 mm, 3-7 mm, 7-15 mm and greater than 15 mm. Each of the five measurable precipitation classes occurs about 7% of the time. To find the skill of a forecast with respect to this climatological control, error points for a given forecast, E_F, are computed according to the ranked probability score (Epstein, 1969; Murphy, 1971)

$$E_F = \sum_{i=1}^{6} [P_F(X_i) - P_O(X_i)]^2 \tag{3}$$

where $P_F(X_i)$ and $P_O(X_i)$ are, respectively, the forecast and observed cumulative probability distributions, as functions of the precipitation category, X_i:

$$P_F(X_i) = \sum_{k=1}^{i} p_F(X_k) \tag{4}$$

$$P_O(X_i) = \sum_{k=1}^{i} p_O(X_k). \tag{5}$$

The consensus precipitation probability skill, defined in Eqs. (1) and (3), is shown in Fig. 3. The results show average skills of 29%, 19%, 6% and −2% for each successive 24-h period. However, our statistics do not show precipitation probability skills significantly different from zero beyond the second forecast day, according to the Student's t test (Panofsky and Brier, 1968). Substantially greater loss of skill with each succeeding forecast period is found in precipitation forecasting than in temperature forecasting. A direct comparison with those results of Sanders and Bosart is not possible since our first forecast period begins 18 h after the 1200 GMT guidance and 12 h after the comparable Albany

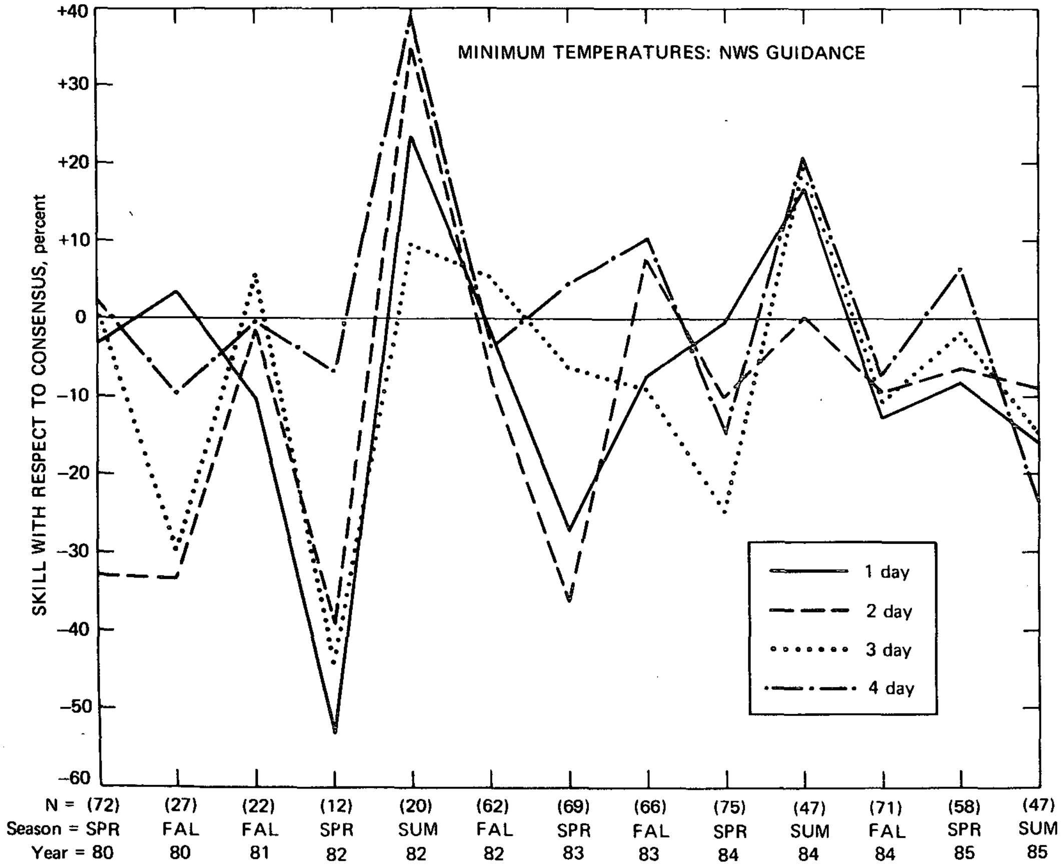


FIG. 5. As in Fig. 1 except for the guidance minimum temperature forecast skill with respect to consensus.

and Boston forecasts begin. In spite of this difficulty, it appears that our skill is somewhat less than those reported for Albany (Bosart, 1983) and is similar to those reported by Sanders (1979) for Boston. Reasons for our reduced precipitation skill are not clear, although the relatively high percentage of convective precipitation events through much of the year may play a role. This issue will be discussed in a later section.

Forecasts of precipitation amount, given that a measurable quantity occurs, were first introduced in the summer of 1983. The error in the precipitation amount is the absolute value of the difference between the forecast and observed values. The errors in the climatology and consensus forecasts are applied to Eq. (1). The results, seen in Fig. 4, average to 17%, −1%, 0% and 1% for each successive 24-h forecast period. Our skill at predicting how much precipitation will occur, given that a measurable amount falls, is substantially less than what is found for temperature or probability forecasting. In fact, our quantitative precipitation forecast skill ceases to exist beyond an 18–42 h forecast range. Clearly, this represents the most challenging of all of the forecasts, a point also reported by Sanders (1979), Charba and Klein (1980), and Bosart (1983).

4. Performance of objective guidance against consensus

As a check on the veracity of our consensus results, we routinely enter objective NWS guidance products in the forecasting contest. The guidance temperature forecasts were based upon the Model Output Statistics (MOS) guidance described by Klein and Lewis (1970) for the 18–42 and 42–66 h forecasts and upon the subjective three- and four-day temperature anomaly charts routinely received over the facsimile circuit.

The MOS guidance is based upon the current day's 1200 GMT output. The MOS 0000–1200 GMT minimum temperature and 1200–0000 GMT maximum temperature forecasts are applied to our 0600–0600 GMT extreme temperature forecasts. Although this practice is not strictly correct, our calendar day extremes typically do verify within the MOS time frame. Additionally, the MOS temperature forecast for Urbana is entered as the arithmetic mean of the Peoria (Illinois) and Indianapolis (Indiana) station forecasts, since Urbana is not an MOS station. Urbana is approximately equidistant (~160 km) along a line from each station, and this mean value corresponds closely to the interpolated values from the MOS forecast fac-

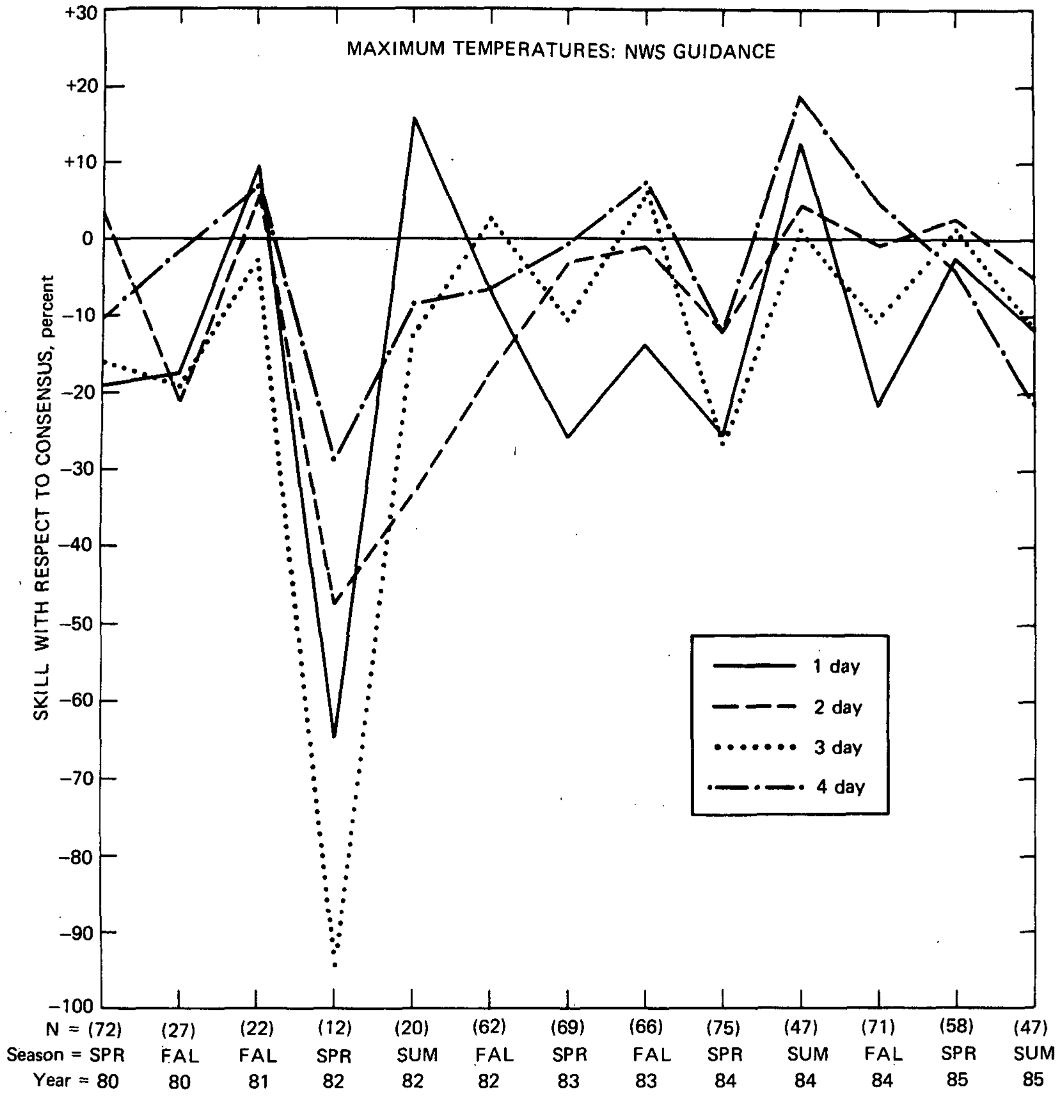


FIG. 6. As in Fig. 1 except for the guidance maximum temperature forecast skill with respect to consensus.

simile analysis. Quantitative precipitation guidance is based upon the limited-area fine-mesh model (LFM; Gerrity, 1977) precipitation forecasts. The Urbana precipitation forecast is entered as the arithmetic average of the Chicago (Illinois), Indianapolis, and Saint Louis (Missouri) station forecasts routinely transmitted over the teletype circuit. All of these products are available to the forecaster during the forecast preparation.

The performance of the minimum temperature guidance against our consensus (Fig. 5) suggests a large semester-to-semester variability and gives little indication of systematic guidance skill over our consensus. This same conclusion holds for the maximum temperature guidance (Fig. 6). In fact, the overall guidance skill with respect to consensus for days one through four is −7%, −12%, −6% and 0%, respectively. The corresponding guidance maximum temperature scores are −14%, −5%, −10% and −3%.

The more limited sample of conditional quantitative precipitation forecasts (QPF) is shown in Fig. 7. The overall score of −9% with respect to the Illinois consensus suggests that our forecasters used the guidance and generally improved upon it. This is symptomatic of all the guidance discussed in this section.

5. Performance of faculty against consensus

To assess the value of individual faculty forecasts against the student-dominated consensus, we have averaged all faculty scores of minimum and maximum temperatures and of precipitation probabilities for each forecast range. The results of this exercise (Fig. 8) show

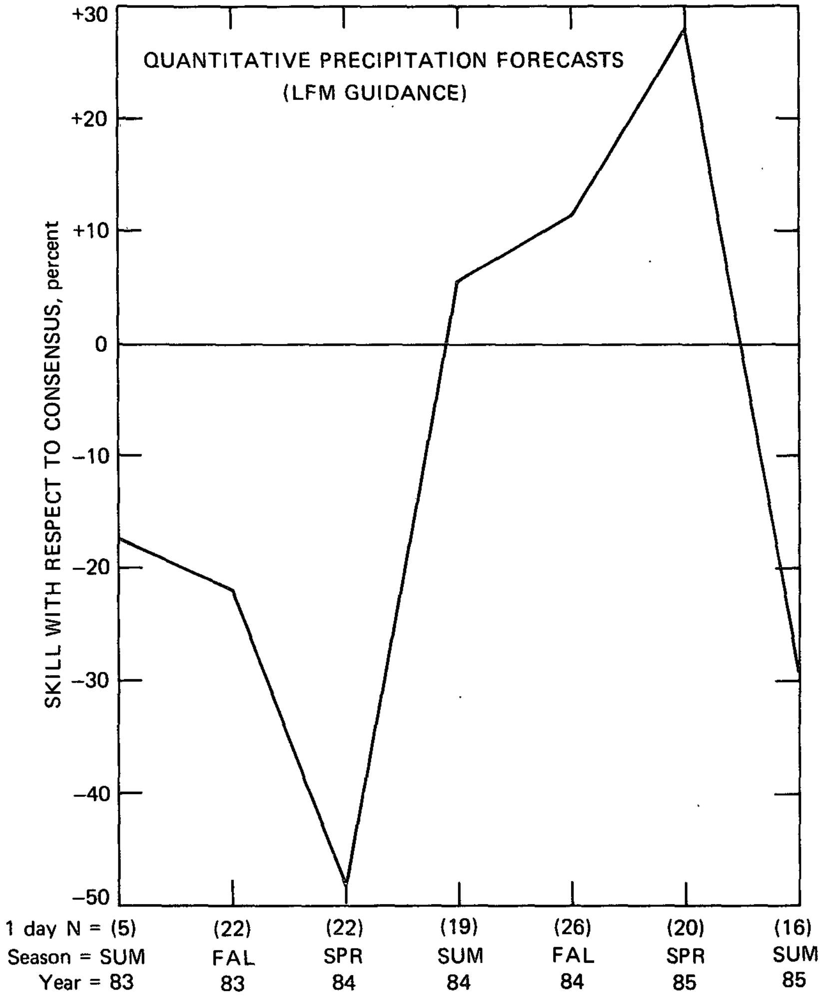


FIG. 7. The LFM guidance quantitative precipitation forecast skill for 18–42 h range with respect to consensus.

a slight loss against the consensus (averages of −6%, −8%, −6% and −4%) for each of the respective forecast ranges from days one through four. There is a weak suggestion of improved performance with increasing forecast range, consistent with the results of Bosart (1983). The participants included Professors John Anderson, Stan Kidder, John Lewis, John Walsh and the author. However, in no semester did more than two faculty members forecast more than 20% of the time (the criteria for inclusion in these statistics). Aside from the prominently poor performance in the fall of 1980, the faculty scores are typically only slightly below consensus.

Figure 8 also indicates that it is rare for any individual (student or faculty) to outperform consensus. Only in the spring of 1980 did more than one person beat consensus. This result is entirely consistent with the results of both Sanders (1979) and Bosart (1983), who have shown that the vast majority of individual forecasters lose to consensus.

6. Trends in forecast skill

We have tabulated the time series of ten academic semesters (excluding summer) of forecasting skill beginning with the spring of 1980. Summer terms were not included in these statistics because of inherently lower skill scores and the fact that all summers were not used during this six-year period. The linear regressions (Table 1) apparently show a temporal improvement in all categories except one. However, in order to assess whether these regression lines are significant, we have applied a Student's t test (Walpole and Myers, 1972) to the null hypothesis that the sample regressions are part of a population with zero temporal trend (slope). As shown in Table 1, the only regression with

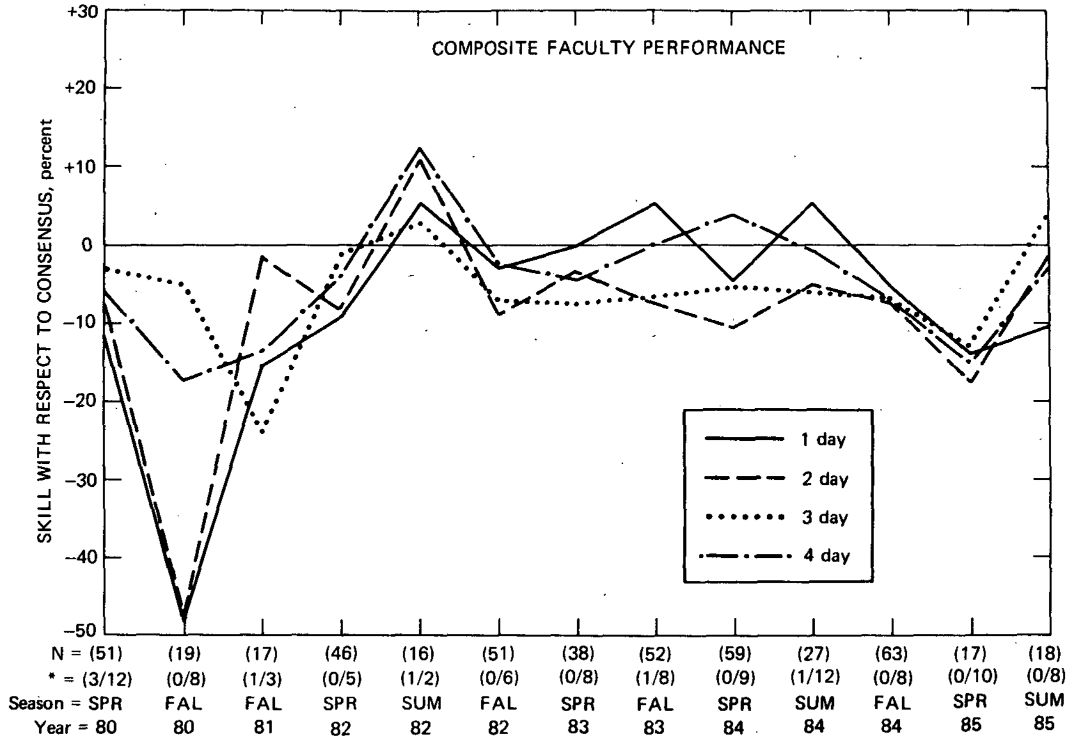


FIG. 8. As in Fig. 1 except for average faculty skill scores with respect to consensus for minimum and maximum temperature and precipitation probabilities. The number (N), shown for each season, indicates the number of forecasts used for each of the four forecast ranges. The left-hand numbers in parentheses (*) show how many student or faculty forecasters scored higher than consensus. The right-hand number shows total number of forecasters (students and faculty) with at least 20% of all possible forecasts.

a significantly different (at the 95% confidence level) slope from zero is the one computed for the 66–90 h minimum temperature forecasts. Weaker suggestions of improvement (90% significance) are found in some other temperature categories with the explained variance being higher than those found for the precipitation

TABLE 1. Linear regression formulas for consensus forecast skill against climatology (Y) as a function of time (X) for each of 12 forecast categories, where $T_{\min}$ is a minimum temperature forecast, $T_{\max}$ is a maximum temperature forecast, and PP is a precipitation probability forecast. The numerical subscript indicates the forecast range in days. Double asterisk indicates whether the trend (slope) falls outside of the 95% confidence interval of zero trend. Single asterisks are shown where the weaker 90% criteria are satisfied.

Fall/Spring Only (Spring 1980–Spring 1985)				
Linear regression (Y=)	Variance explained (%)	Mean overall forecasts	Projected skill, fall 2000 (X = 41)	Forecast category
$1.08X + 52.87$	19.8	59.18	97.25	$T_{\min 1}$
$0.73X + 47.41$	10.5	50.93	77.25	$T_{\min 2}$
$2.67X + 22.12$**	64.3	37.10	131.63	$T_{\min 3}$
$1.36X + 15.31$*	30.8	22.33	71.17	$T_{\min 4}$
$1.09X + 53.52$*	30.7	60.29	98.25	$T_{\max 1}$
$1.59X + 39.83$*	32.8	48.79	104.86	$T_{\max 2}$
$1.65X + 25.81$*	38.5	35.12	93.59	$T_{\max 3}$
$0.91X + 17.28$	13.4	22.66	54.63	$T_{\max 4}$
$2.18X + 18.29$	8.2	31.76	107.57	PP_1
$-1.20X + 27.54$	10.9	20.89	−21.66	PP_2
$1.60X - 3.00$	25.6	7.23	62.45	PP_3
$1.02X - 9.06$	14.9	−1.88	32.68	PP_4

TABLE 2. Linear regression formulas for NWS guidance skill against climatology (Y) as a function of time (X) for each of the eight forecast categories. $X = 1$ corresponds to Spring 1980. Double asterisk indicates whether the trend (slope) falls outside of the 95% confidence interval of zero trend. Single asterisk shows the same criterion being satisfied for the 90% confidence interval.

Fall/Spring Only (Spring 1980–Spring 1985)

Linear regression (Y=)	Variance explained (%)	Mean overall forecasts	Projected skill, fall 2000 (X = 41)	Forecast category
$0.86X + 50.62$	9.6	56.36	85.88	T_{min1}
$2.02X + 34.07$*	38.6	44.59	117.09	T_{min2}
$1.92X + 23.22$*	30.4	34.08	101.72	T_{min3}
$1.45X + 13.54$	24.9	21.57	72.95	T_{min4}
$1.19X + 46.21$	19.6	54.63	95.06	T_{max1}
$1.86X + 35.75$**	50.2	48.05	112.14	T_{max2}
$1.82X + 18.90$	17.5	30.60	93.67	T_{max3}
$0.69X + 17.33$	7.4	21.79	45.75	T_{max4}

probabilities. Our sample of ten academic semesters is small, and the trends in a few cases bring us to unrealistically high skills in excess of 100% by the fall of the year 2000.

To investigate the concurrent behavior of the NWS guidance skill, we have found the regression formulas (Table 2) for each of the guidance temperature forecasts. Although a comparison with Table 1 clearly shows the NWS guidance to be less skillful than our consensus, some temporal improvement is shown with these categories showing significance at the 90% confidence level. Although fewer forecast categories show a significant improvement than was found for our consensus, we cannot rule out the possibility of our consensus benefiting from improving guidance. Sanders (1986), using an 18-year sample of data at MIT, has also shown the objective guidance forecasts to have less skill than the corresponding consensus. However, Sanders also found the objective guidance to have gained substantially on the MIT consensus skill, especially in recent years.

The regression formulas from the composite faculty performance versus consensus (Fig. 8) are shown for twelve forecast categories in Table 3. Clearly, inconsequential significance can be attached to any trends in temperature or precipitation probability forecasting. The improvement in faculty performance against the student-dominated consensus is not evident using the Student's t test at the 90% confidence level previously described.

Thus, it appears that a constantly changing student population and its assured fresh supply of neophyte forecasters has no effect on degrading the quality of the consensus. Further, the relatively constant population of faculty has shown no significant improvement in its performance over this consensus.

TABLE 3. Linear regression formulas for faculty guidance skill against consensus (Y) as a function of time (X) for each of the 12 forecast categories. $X = 1$ corresponds to Spring 1980. No significance (at the 90% confidence level) can be found in any temporal trend.

Fall/Spring Only (Spring 1980–Spring 1985)

Linear regression (Y=)	Variance explained (%)	Mean overall forecasts	Projected skill, fall 2000 (X = 41)	Forecast category
$0.04X - 8.09$	0.0	−5.09	−6.44	T_{min1}
$-0.11X - 7.04$	0.4	−8.26	−11.64	T_{min2}
$0.31X - 9.26$	1.3	−7.42	3.63	T_{min3}
$0.70X - 14.80$	2.6	−7.39	13.84	T_{min4}
$1.44X - 14.80$	14.9	−7.11	44.18	T_{max1}
$-1.02X - 5.27$	16.2	−9.44	−47.03	T_{max2}
$-1.08X - 2.46$	7.6	−7.49	−46.77	T_{max3}
$0.47X - 7.90$	3.4	−3.32	11.49	T_{max4}
$5.56X - 46.53$	20.6	−7.63	181.31	PP_1
$5.01X - 44.30$	12.9	−9.53	161.44	PP_2
$0.84X - 11.78$	3.1	−3.01	22.95	PP_3
$0.39X - 5.32$	1.4	−2.02	10.73	PP_4

TABLE 4. Consensus biases.

	Minimum temperatures		Maximum temperatures		Quantitative precipitation	
	N	Mean error (°C)	N	Mean error (°C)	N	Mean error (mm)
Spring	1290	−0.13	1289	−0.70	196	−3.63
Summer	460	0.50	460	0.91	162	−6.35
Fall	1067	−0.05	1067	0.38	201	−4.48

7. Biases in forecasting

Calculations of mean temperature and QPF errors by both consensus and NWS guidance (shown in Tables 4 and 5, respectively) show qualitatively similar biases. Statistical analyses of these data for individual semesters reveal several semesters with a significant warm maximum temperature bias. The significance was determined through use of the Student's t test, and this criterion was applied at the 95% confidence level. This warm maximum temperature bias is especially pronounced during the summer semester. A systematic underprediction of precipitation is especially prominent during the summer semester.

This dry bias in the precipitation forecasts may be related to Urbana's precipitation climatology. Figure 9 illustrates this point by showing Urbana's total precipitation during the six-year forecast experiment. Measurable precipitation amounts associated with thunder heard are shaded. The number of days in which both thunder and measurable precipitation are recorded is also shown in Fig. 9. During the meteorological cold-season months of December, January and February, convection accounts for only about 10% of the precipitation total and occurs on less than 5% of the measurable precipitation days. After February, the amount of precipitation associated with thunder rises dramatically through the summer. During the warm-season months about 90% of all precipitation is convective. Even the transition months of March–May, and September–November are associated with substantial convective contributions to the total precipitation. In fact, over all months in the period, 60% of all precipitation was recorded on days of thunder.

Given this substantial convection in all months except the three cold-season months and the LFM guidance often missing these convective events or substantially underestimating them (Table 5), the tendency of the forecaster is to err on the side of a conservative precipitation forecast. Thus, our dry forecast bias is likely related to the prominence of the convective precipitation.

To address the question of whether the warm maximum temperature bias is related to precipitation underprediction, we examined forecast statistics for the two summers in which concurrent temperature and QPF experiments were conducted. Consensus maximum temperature errors found on days in which measurable precipitation was underestimated in the forecast were compared with maximum temperature errors on all other days. As can be seen in Table 6, both samples show a warm bias, and each sample shows a mean temperature error significantly different from zero (using the Student's t test and a 99.9% confidence limit). However, a comparison between the two samples also shows that maximum temperature errors were significantly more warm-biased on days in which measurable precipitation was underforecast than on other days. This conclusion is based upon the use of the Student's t test and is valid at the 99.9% significance level.

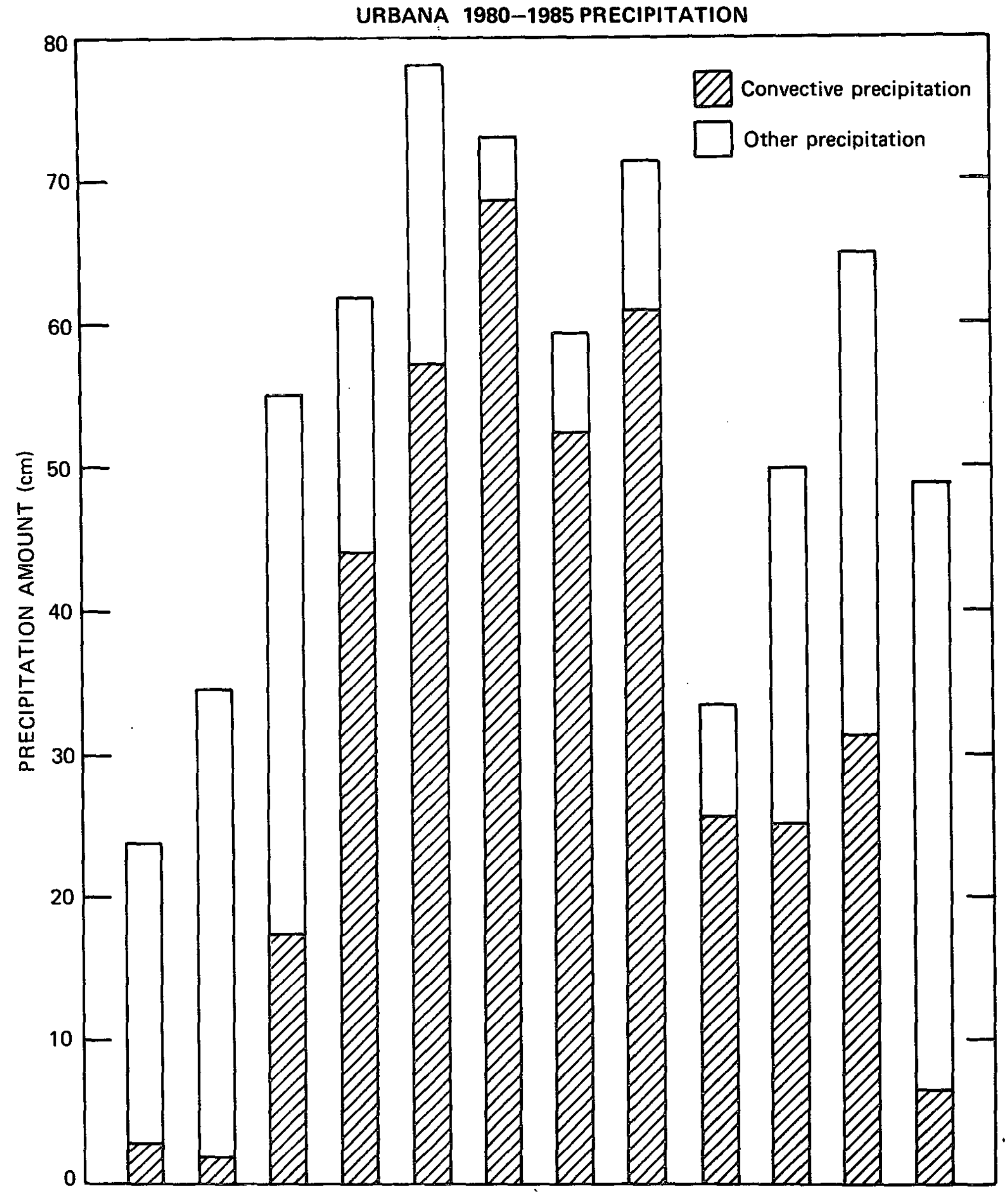
Thus, these data show that days in which measurable precipitation is underestimated are also typically days in which the maximum temperature is substantially overestimated. One possible explanation for this relationship is that warm-season precipitation events in Illinois, while usually convective (Fig. 9), are also often organized on scales substantially larger than the cumulus (such as mesoscale convective complexes, described by Maddox, 1980). This poorly predicted phenomenon often has cloudiness persisting over an area for up to 24 hours. This cloud cover can cause a substantial reduction in maximum temperature.

8. Concluding discussion

The results from six years of daily forecasting at Urbana, Illinois, show qualitatively similar results to those found by Sanders (1979) and Bosart (1983) for the respective East Coast cities of Boston and Albany. Although our minimum temperature forecast skill is slightly higher than these investigators have found, this may be associated with our later forecast deadline and the accompanying additional NWS guidance. Few published results exist with which to compare our maximum temperature skills. The precipitation probability and quantitative precipitation skills, lower than temperature forecasting skills, may be adversely affected by the convective nature of much of Urbana's precipitation.

TABLE 5. NWS guidance biases.

	Minimum temperatures		Maximum temperatures		Quantitative precipitation	
	N	Mean error (°C)	N	Mean error (°C)	N	Mean error (mm)
Spring	1133	−0.62	1131	0.15	42	0.72
Summer	452	0.30	452	0.70	40	−10.21
Fall	960	−0.05	459	0.44	48	−2.39



	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	Total
% of amount	12.1	5.1	31.6	71.2	73.3	93.9	88.3	85.4	76.3	50.2	48.3	13.3	
Days of T	2	2	13	33	34	38	38	42	18	11	17	3	251
All precip. days	51	59	73	77	65	61	53	67	50	65	67	70	758

FIG. 9. Urbana (Morrow Plots) monthly precipitation summary for the six year period 1980–85. Monthly total and convective precipitation (shaded) amounts are shown. The percentages of the monthly precipitation total found on days with thunder are shown directly below the graph. The number of days in which both thunder (T) and measurable precipitation are recorded is shown, along with the number of days in which measurable precipitation was recorded.

Our consensus is virtually unbeatable by any one individual (student or faculty) or by the standard NWS guidance. However, there is a suggestion of some temporal improvement in our consensus skill in temperature forecasting, along with a concurrent improvement in the NWS temperature guidance; we cannot rule out the possibility that the former is driven by the latter. However, no significant trend in precipitation probability forecast skill can be established. Our skill in predicting how much precipitation will fall, given

TABLE 6. Consensus maximum temperature error (forecast minus observed) statistics for 1984 and 1985 summer semesters for cases when forecasts underestimated measurable precipitation and for all other cases.

	N	Mean (°C)	Standard deviation (°C)
Precipitation underestimated	80	+2.07	3.10
Other cases	300	+0.80	2.16

that a measurable amount occurs, is small for up to a day in advance and negligible for ranges beyond one day.

Our rather high (and perhaps improving) temperature skills contrast sharply with our rather low (and not improving) precipitation skills. This leads to the philosophical question: What constitutes a good forecast? It is the author's belief that high forecast success based upon a sample dominated by quiescent weather conditions are poor indicators of "predicting the weather." The more interesting and less frequent meteorological events, such as heavy precipitation, often dominate the public's perception of the weather and attract much of the meteorologist's research time. Thus, it appears that these relatively rare events should be emphasized in a forecasting effort. Research into these areas will be reported upon in future papers.

Acknowledgments. This research has been supported, in part, by the National Science Foundation under Grant ATM-8305096. The author would like to thank Dr. Lance Bosart and anonymous reviewers for their helpful comments on this work. The data collection assistance of Gerry Bell, Mike Ferguson, and Karen Garrelts is gratefully acknowledged. John Brother drafted the figures, and Karen Garrelts and Norene McGhiey typed the manuscript. Special thanks is also extended to all those faculty, students, and friends who have been willing to apply their meteorological skills to the difficult task of forecasting.

REFERENCES

Bengtsson, L., 1985: Medium-range forecasting—The experience of ECMWF. *Bull. Amer. Meteor. Soc.*, **66**, 1133–1146.

Bosart, L. F., 1983: An update on trends in skill of daily forecasts of temperature and precipitation at the State University of New York at Albany. *Bull. Amer. Meteor. Soc.*, **64**, 346–354.

Charba, J. P., and W. H. Klein, 1980: Skill in precipitation forecasting in the National Weather Service. *Bull. Amer. Meteor. Soc.*, **61**, 1546–1555.

Epstein, E. S., 1969: A scoring system for probability forecasts. *J. Appl. Meteor.*, **8**, 985–987.

Gerrity, J. G., 1977: The LFM model 1976: A documentation. NOAA Tech. Memo. NWS-NMC-60, 68 pp. [NTIS PB-279-419.]

Glahn, H. R., 1985: Yes, precipitation forecasts have improved. *Bull. Amer. Meteor. Soc.*, **66**, 820–830.

Klein, W. H., and F. Lewis, 1970: Computer forecasts of maximum and minimum temperatures. *J. Appl. Meteor.*, **9**, 350–359.

Maddox, R. A., 1980: Mesoscale convective complexes. *Bull. Amer. Meteor. Soc.*, **61**, 1374–1387.

Murphy, A. H., 1971: A note on the ranked probability score. *J. Appl. Meteor.*, **10**, 155–156.

——, and T. E. Sabin, 1986: Trends in the quality of National Weather Service forecasts. *Wea. Forecasting*, **1**, 42–55.

Panofsky, H. A., and G. W. Brier, 1968: *Some Applications of Statistics to Meteorology.* The Pennsylvania State University, 224 pp.

Ramage, C. S., 1982: Have precipitation forecasts improved? *Bull. Amer. Meteor. Soc.*, **63**, 739–743.

Sanders, F., 1979: Trends in skill of daily forecasts of temperature and precipitation, 1966–1978. *Bull. Amer. Meteor. Soc.*, **60**, 763–769.

——, 1986: Trends in skill of Boston forecasts made at MIT, 1966–84. *Bull. Amer. Meteor. Soc.*, **67**, 170–176.

Sela, J. G., 1980: Spectral modeling at the National Meteorological Center. *Mon. Wea. Rev.*, **108**, 1279–1292.

Walpole, R. E., and R. H. Myers, 1972: *Probability and Statistics for Engineers and Scientists.* Macmillan, 506 pp.