Trends in Skill of Boston Forecasts Made at MIT, 1966–84

Frederick Sanders

9 Flint St., Marblehead, MA 01945

Abstract

Eighteen years of daily forecasts of temperature and precipitation at Boston, representing a consensus of MIT participants, are evaluated. The current state of the art in these forecasts is a skill of roughly 50% with respect to a climatological control for the first day ahead. About 50% of residual skill remains at the end of each subsequent 24 h. Temperature forecasts are somewhat more skillful than precipitation forecasts. Conditional forecasts of precipitation amount, verified only on days when precipitation occurs, are much less skillful.

Skills are increasing at the very slow rate of a few tenths of a percent per year, except for the first day and except for the conditional precipitation forecasts. First-day skill shows virtually zero trend, while the quantitative precipitation forecasts have deteriorated.

Objective-guidance forecasts obtained directly or indirectly from dynamical models have shown less skill. With time, however, they have gained substantially on the skill of subjective forecasts based in part upon them, and have reached approximate equality in a number of instances.

1. Introduction

Since the indication from the first six years of the Massachusetts Institute of Technology (MIT) forecasting contest that the state of the art of short-range-forecast skill was not advancing rapidly (Sanders, 1973), there has been considerable interest in this and related issues, in the real world (e.g., Pierce, 1976; Ramage, 1976, 1982; Zurndorfer *et al.,* 1979; Charba and Klein, 1980; Glahn, 1985; Gordon, 1985) as well as on the academic scene (Bosart, 1975, 1983; Gedzelman, 1978; Firestone, 1979; Scanlon and Anawalt, 1980; Baker, 1982). Six years have passed since the last report on the state of the MIT contest, and it is appropriate to report whether the modest trends found at that time (Sanders, 1979) have continued.

2. The MIT forecasts

Predictions are made Monday through Friday, on a mostly voluntary basis, for Boston for the first, second, third, and fourth 24-h periods following the normal submission time of 1800 GMT. Contestants range from freshmen to faculty, with widely varying amounts of experience. As shown in the sample forecast form (Fig. 1), the forecasts are for the 24-h maximum and minimum temperatures in °F (T_{MAX} and T_{MIN}), a probability distribution over six ranges of precipitation amount in percent (*PP*), and for an amount category in 20ths of inches (*P*). The T_{MAX} forecasts refer to 24-h periods 12 h earlier than those to which the others pertain, and the *P* forecast is discarded except on those occasions when a measurable amount falls (or, since 1982, when any amount falls).

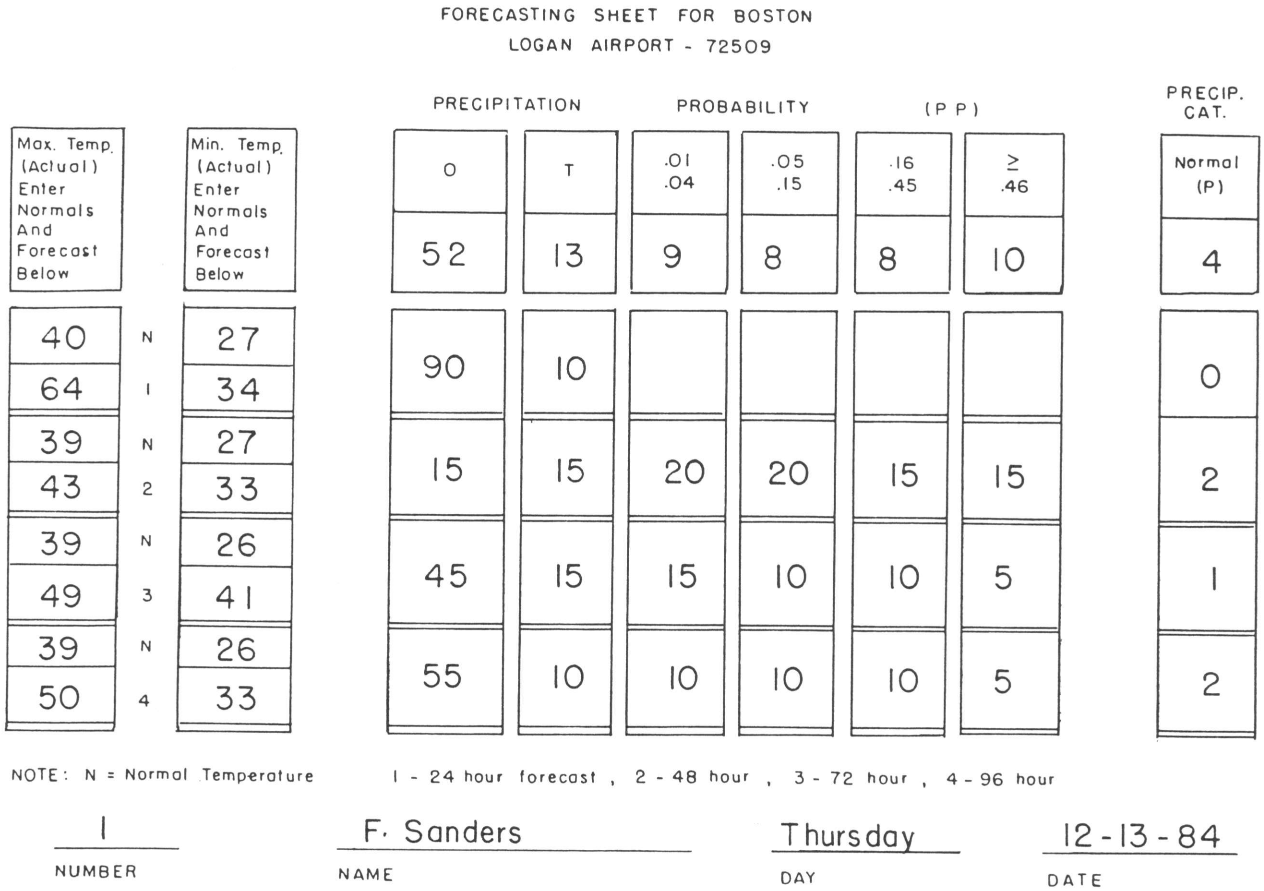
The scoring rules are absolute error for T_{MIN}, T_{MAX}, and *P*, and the ranked probability score (Epstein, 1969; Murphy, 1971) for *PP*. A consensus, formed on any day when two or more forecasts are submitted, will receive major attention here. Error points are accumulated during each session of the contest, which begins on each academic registration day in the Fall term (F), Spring term (Sp), and Summer term (Su).

A full range of graphical and numerical products, received from The National Weather Service, (NWS) are available for free and uninhibited consultation. Individuals are ranked according to their standing relative to consensus (almost always inferior), calculated for those days when they contributed to it; climatological MOS (model output statistics) (Klein and Hammons, 1975), LFM (limited-area fine mesh) (Newell and Deaven, 1981) and NWS local (FP4) forecasts are verified for comparison where appropriate. Other details are given by Sanders (1973, 1979).

3. Levels of Skill

Skill is defined as the reduction in error points of the forecasts in question relative to the score of the control forecast, understood unless otherwise stated to be a forecast of the climatological mean, divided by the control score and expressed as a percent. The MIT results, averaged over the 18 years from fall 1966 through summer 1984 for T_{MIN} and *PP*, over the period from fall 1974 through summer 1982 for *P* (the change in acceptance criterion without a change in the climatological forecast then put this control at a decided disadvantage), and from spring 1979 through summer 1984 for T_{MAX}, appear in Fig. 2. The skill for each 24-h period is attributed to the ending time of that period.

Temperature forecasts show more skill than precipitation forecasts, in terms of the scores computed. This is particularly so for the conditional *P* contest, in which the undeniable skill in distinguishing no-rain days from others is not reflected. The decay of skill with range is approximately exponential out to four days for temperature, with a loss of about 40% of the residual skill each day from the first-day value of about 55%. With precipitation, the exponential loss at the faster rate of about 50% of residual brings the modest first-day skills of about 45% for the *PP* and 20% for the *P* forecasts to dubiously useful levels by the third day, whereafter virtually all skill is lost. Bosart (1983) has found



FORECASTING SHEET FOR BOSTON
LOGAN AIRPORT - 72509

Max. Temp. (Actual) Enter Normals And Forecast Below		Min. Temp. (Actual) Enter Normals And Forecast Below	PRECIPITATION PROBABILITY (P P) 0	T	.01 .04	.05 .15	.16 .45	≥ .46	PRECIP. CAT. Normal (P)
			52	13	9	8	8	10	4
40	N	27							
64	1	34	90	10					0
39	N	27							
43	2	33	15	15	20	20	15	15	2
39	N	26							
49	3	41	45	15	15	10	10	5	1
39	N	26							
50	4	33	55	10	10	10	10	5	2

NOTE: N = Normal Temperature   1 - 24 hour forecast, 2 - 48 hour, 3 - 72 hour, 4 - 96 hour

1	F. Sanders	Thursday	12-13-84
NUMBER	NAME	DAY	DATE

FIG. 1. Sample forecast prepared on 13 December 1984. Climatological forecasts (°F for temperature, percent for precipitation probability, 20ths of inches for precipitation category).

somewhat greater skill levels for fall and spring in a parallel contest at the State University of New York at Albany (SUNYA), due mainly to the T_{MAX} forecasts in which uncertainty concerning the maritime factor at Boston limits MIT skill. Even discounting these forecasts, the SUNYA results show about 2% greater skill; Bosart's reasons for this convey more generosity than conviction. The small difference suggests that our results are representative for locations in the northeastern United States.

National Weather Service forecasts deal generally with 12-h periods and are thus more demanding. With some allowance for this effect, skill levels indicated in NWS forecasts by Zurndorfer *et al.* (1979) and by Glahn (1985) seem comparable to ours. In New Zealand, Gordon's (1985) results show a more-rapid decay with range for temperature and generally lower values for probability of precipitation, with both skills nearly vanishing by the end of the second day, illustrating the difficulty of forecasting in a maritime climate with relatively sparse data coverage.

4. Trends with time

Trends of skill over the available periods of time are displayed in Figs. 3–6. Because of prominent seasonal variations, results are shown separately for fall, spring, and summer, except for T_{MAX} in which the five-year period of record precluded such a stratification.

With seasonally stratified results merged, and the *P* forecasts excluded, the mean regression trends of skill are upward: +0.01% per year for the first day, and +0.54%, +0.31%, and +0.10% for succeeding days. Thus the loss of skill for the first day, deplored by Sanders (1979) in the last report, has been halted, but just barely, and at the cost of slightly more modest estimates of the rate of gain for subsequent days.

Bosart (1983), for the SUNYA equivalent of our T_{MIN} and *PP* forecasts, shows averaged *loss* of skill at the annual rate of eight tenths of a percent (from the regression coefficients in his figures 1 and 2) per year. Ramage (1982), in a study of NWS 12-h forecasts of precipitation probability up to 36-h range, found evidence of increasing skill only in the eastern and central regions in winter and there at an annual rate of only one or two tenths of a percent. In a study designed to see whether Ramage's pessimism would withstand reexamination of the data and extension in time, Glahn (1985) considered these NWS forecasts from 1967 to 1982. He found increasing skill year-round for all regions and all periods, at a rate for the northeastern region of eight tenths of a percent per year in the cooler season and seven tenths in the warmer.

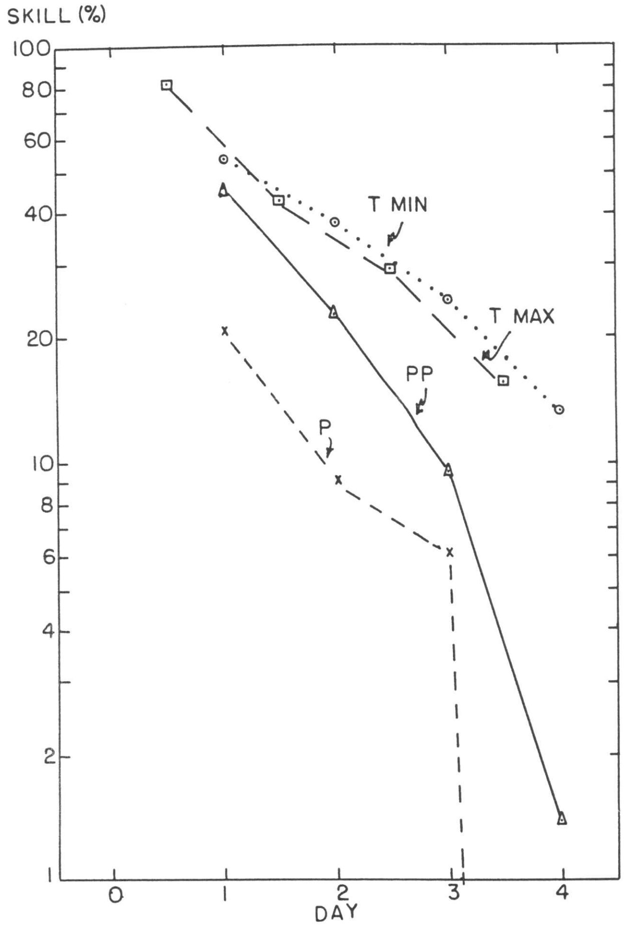


FIG. 2. 18-year average forecast skill of consensus with respect to climatology, as a function of range.

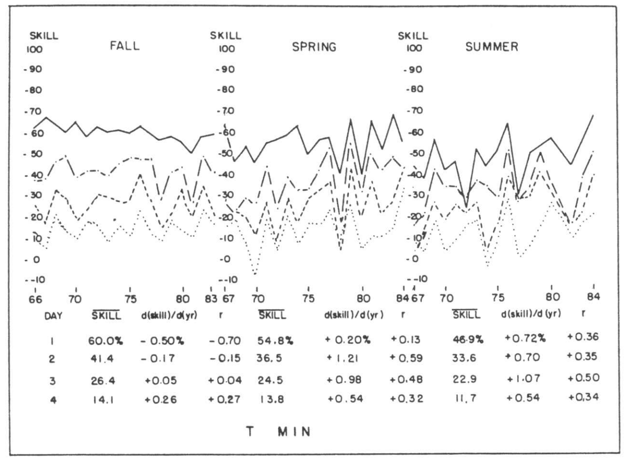


FIG. 3. Times series of consensus skill over 18 years in minimum temperature forecasts, by season, for days 1 (solid), 2 (dot-dashed), 3 (dashed) and 4 (dotted).

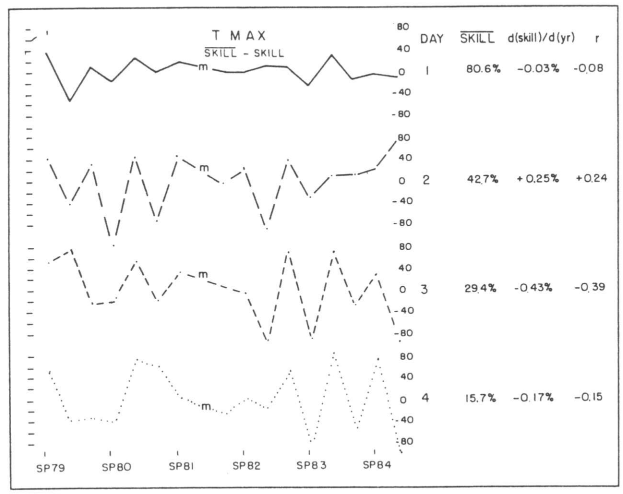


FIG. 4. Time series of deviation of consensus skill in maximum temperature forecasts from five-year or six-year seasonal average.

The trends for each of our curves in Figs. 3–5 cannot be individually relied upon, because the explained variances are so small. Nevertheless, however they are combined or viewed, the preponderance of regression coefficients is positive. So our results fall within the range found by others, but rather closer to Glahn's results than to Ramage's. What seems clear is that we are a long way from perfection and will likely take a long time to get there, even with these common variables.

The conditional highly resolved quantitative *P* forecasts at MIT and SUNYA, evidently more difficult than precipitation forecasts undertaken elsewhere, create both opportunity and frustration. At MIT, trends were mostly upward from 1974 to 1978 (c.f. Sanders [1979], his table 3), but since then they have suffered mightily, being distinctly downward from 1974 to 1982 (Fig. 6) except in fall. (The rise since then is spurious, as indicated earlier.) At SUNYA, on the other hand, the trends were up from 1976 to 1982 (c.f. Bosart [1980], his figure 3) At both places, however, the result for a given semester typically depends on the outcome of a few big events, which can be either great triumphs or crushing defeats. As a result the variation from one semester or year to the next is large and erratic, and computed trends are far from robust.

5. Consensus versus others

a. The author

The group that constitutes consensus has consisted mainly of students, many of whom have been freshmen whose participation is their first (and occasionally last) exposure to meteorology. Were they individually much less skillful than the author, with 40 years of forecast experience, mostly in New England, then we might expect that the latter would show skill relative to consensus. Table 1 shows that this is not the case. Overall, the author loses to consensus by about 2%, with no clear distinction among the types of forecast. Beginning students tend to lose individually by substantially larger amounts, but their consensus is entirely respectable. We have not developed the data to confirm or refute Gedzelman's (1978) finding that an individual plateau of forecasting skill is reached after about 20 forecasts, but we believe that our experience is similar. There is little consistent trend in the author's skill relative to consensus, with the mean close to zero and about half the regression coefficients

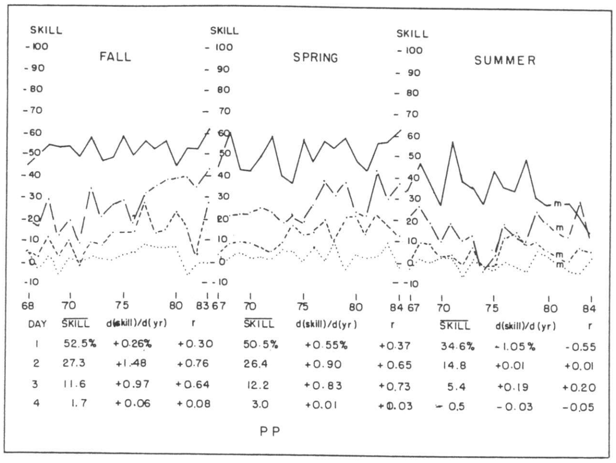


FIG. 5. As in Fig. 3, but for precipitation-probability forecasts.

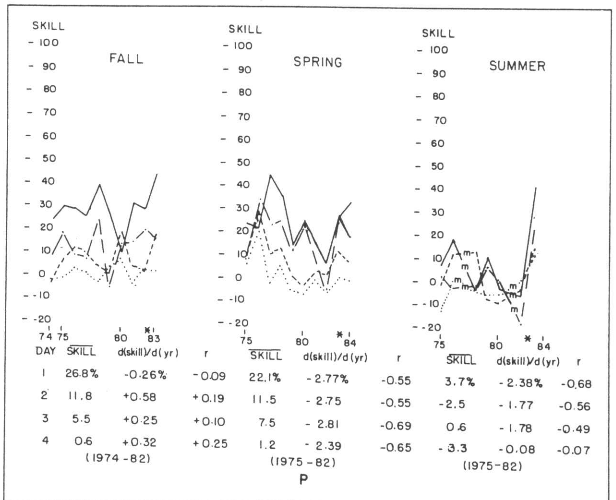


FIG. 6. Time series of deviation of consensus skill in precipitation category forecasts from nine-year or eight-year seasonal average. Asterisk indicates exclusion from calculation of mean skill and trend.

of either sign. A similar analysis by Bosart (1983) shows similarly small trends, but his skill averages slightly above consensus. Evidently the students are less-skillful forecasters at SUNYA than at MIT. At either place an individual, whether experienced or tyro, rarely beats consensus on the average.

b. NWS FP4 forecasts

The forecasts for Boston formulated by the local NWS office shortly after our initial time of 1800 GMT, and disseminated as part of the FP4 message, have been entered in our T_{MIN} and T_{MAX} contests as appropriate since the 1978/79 academic year. Their skill relative to consensus is shown in Table 2. That these forecasts for 12-h intervals lose to consensus, by a substantial margin in the case of T_{MIN}, is partly attributable to a disadvantage placed on them by verifying them from 24-h extremes. Possibly the main reason is our freedom to concentrate on a single station, whereas the NWS office has a forecast responsibility for several. It does not appear in any case that our results can be discounted as representing less than the state of the art.

TABLE 1. Mean and trend of skill of author forecasts with respect to consensus, by season for T_{MIN} and PP, merged for T_{MAX} and P. The overbar denotes the skill, averaged over the period; $d[\ ]/d(yr)$ its annual rate of change (percent); and r is the correlation coefficient between skill and time. Fall of 1967, spring of 1968, and spring of 1981 have been excluded.

Season	Years	Day	$\overline{[\ ]}$ (%)	$d[\ ]/d(yr)$ (%)	r
		T_{MIN}			
Fall	1966–83	1	+0.2	−0.09	−0.05
		2	−3.5	+0.78	+0.53
		3	−4.2	+0.01	+0.01
		4	−3.5	−0.33	−0.35
Spring	1967–84	1	−5.3	−0.04	−0.02
		2	−8.1	−0.13	−0.07
		3	−5.6	−0.89	−0.54
		4	−4.8	−0.11	−0.09
Summer	1967–84	1	+1.3	−0.73	−0.37
		2	−1.0	−0.30	−0.14
		3	+0.8	+0.34	+0.22
		4	−1.3	−0.09	−0.05
		PP			
Fall	1966–83	1	−2.4	−0.39	−0.18
		2	+0.4	+0.04	+0.02
		3	+1.7	+0.51	+0.37
		4	−1.2	−0.25	−0.27
Spring	1967–84	1	+1.2	+0.36	+0.20
		2	−5.2	−0.50	−0.33
		3	−4.5	−0.76	−0.38
		4	−0.9	−0.08	−0.14
Summer	1967–84	1	−1.0	+0.72	+0.28
		2	+2.1	+0.61	+0.33
		3	−2.7	−0.12	−0.01
		4	−1.1	−0.19	−0.14
		T_{MAX}			
Spring 1979– Summer 1984		1	−12.0	−1.10	−0.23
		2	−2.3	+1.20	+0.38
		3	+1.6	+1.22	+0.66
		4	+1.4	+0.97	+0.51
		P			
Fall 1974– Summer 1984		1	−5.3	+0.05	+0.02
		2	−1.6	+0.07	+0.04
		3	−1.0	+0.25	+0.18
		4	−1.5	+0.39	+0.45

TABLE 2. Mean and trend of skill of FP4 temperature forecasts with respect to consensus. Variables as in Table 1.

Day	$\overline{[\ ]}$ (%)	$d[\ ]/d(yr)$ (%)	r
	T_{MIN} (F78–Su84)		
1	−14.8	−0.59	−0.24
2	−11.6	−0.26	−0.12
3	−12.0	−0.33	−0.16
	T_{MAX} (Sp79–Su84)		
2	−1.6	+0.21	+0.17
3	+0.0	−0.26	−0.11

c. MOS temperature forecasts

Forecasts of maximum and minimum temperature, prepared by the MOS technique from the current 1200-GMT LFM model run, are available to the forecasters and are

TABLE 3. Mean and trend of skill of MOS temperature with respect to consensus, by season for T_{MIN} and merged for T_{MAX}. Variables as in Table 1. Year of equality, by linear trend, indicated as yr[0]. In far right column, asterisk denotes 1978 forecast; ± is sign of $d[\ ]/d(\mathrm{yr})$ 1972–78.

Season	Years	Day	$\overline{[\ \ ]}$ (%)	$d[\ ]/d(yr)$ (%)	r	yr[0]
			T_{MIN}			
Fall	1972–83	1	−28.4	+0.59	+0.17	2026 −1963*
		2	−13.8	+0.46	+0.17	2009 +1992*
		3	−10.1	+3.84	+0.85	1980 +1987*
		4	−3.1	+0.99	+0.35	1981 −1997*
Spring	1973–84	1	−26.2	+2.49	+0.52	1989 +1990*
		2	−10.6	−0.01	−0.00	1253 +2017*
		3	−18.0	+0.95	+0.32	1997 −1900*
		4	−7.2	+1.17	+0.39	1985 −1959*
Summer	1973–84	1	−17.4	+1.35	+0.26	1991 +2009*
		2	−6.7	+0.21	+0.05	2010 +1978*
		3	−11.2	+0.45	+0.09	2002 −1973*
		4	−8.1	+0.85	+0.28	1988 −2012*
			T_{MAX}			
Spring 1979–Summer 1984		2	−0.6	+0.25	+0.14	1982
		3	+4.0	+0.92	+0.43	1980
		4	−4.8	−1.20	−0.47	1976

verified for T_{MIN} on days 1 and 2 and for T_{MAX} on days 2 and 3. The temperatures from the five-day forecast received a few hours after our forecast deadline are used for T_{MIN} on days 3 and 4 and for T_{MAX} on day 4. These are denoted, collectively albeit somewhat inaccurately, as the MOS forecasts. Their skill relative to consensus is documented in Table 3, for the periods indicated. A similar comparison was made earlier (Sanders, 1979) but results were not then stratified by season. They now have been and are included in Table 3.

The level of skill has been less than that of consensus in all contests (except T_{MAX} day 2), averaged over the periods of comparison. Similar results are shown by Bosart (1983). The trends, however, even more now than at the earlier time, show that the objective-guidance forecasts are gaining on us and have overtaken us, or will do so in the near future in a number of contests. For reasons that are not entirely clear, we seem to do poorly on T_{MAX} forecasting, with respect not only to SUNYA, but also to the climatological control (Figs. 2 and 4) and to the MOS guidance (Table 3).

d. LFM quantitative precipitation forecasts

The amount of precipitation indicated at Boston in the alphanumeric message based on the 1200-GMT LFM run has been used for day 1 of the *P* contest since fall 1974. This message also provides a forecast for the first 18 h of day 2, which since spring 1982 we have completed by arbitrarily adding the last 6 h of precipitation forecast at LaGuardia, some 300 km to the southwest whence most storms come. This guidance material is available to the forecasters. The level and trend of skill of these forecasts relative to consensus is indicated in Table 4. As in the case of the MOS temperature forecasts, the skill is less but the trend is upward (except in the fragile summer contest); the exclusion of a single year (dominated by a single event in which a preposterous "bull's eye" of heavy rain happened to be predicted directly over Boston but mercifully failed to occur), changed the estimated trend dramatically. These trends no doubt reflect in part the beneficial effect of the removal during the period of study (Newell and Deaven, 1981) of the large positive bias identified by Bosart (1980) in the LFM forecasts. Note, however, that the short-series day 2 forecasts do not enjoy this effect, yet show a pronounced upward trend, for what little it is worth. The results in Table 4 indicate that the day of overtaking is upon us, contrary to the indications given by Bosart (1983).

6. Concluding Comments

From an 18-year experiment we have established that:

1) The current state of the art in these daily forecasts shows a loss of skill with respect to a climatological control of roughly 50% per 24 h of increasing range. This result is representative of locations in the northeastern United States.

TABLE 4. Mean and trend of skill of LFM precipitation forecasts (*P*) with respect to consensus, by season for day 1, merged for day 2.

Season	Years	[ ] (%)	*d*[ ]/*d*(*yr*) (%)	*r*	yr(0)
		Day 1			
Fall	1974–83	−31.3	+3.90	+0.43	1987
Spring	1975–84	−30.6	+5.21	+0.74	1985
Summer	1975–84	−31.7	−3.87	−0.30	1971
Excluding Summer 1983		−18.3	+2.03	+0.62	1988
		Day 2			
Spring 1982–Summer 1984		−8.2	+3.26	+0.35	Sp 84*

*Season (0).

2) This skill is increasing at the very slow rate of a few tenths of a percent per year. This result is confirmed in NWS forecasts by studies conducted by Zurndorfer *et al.* (1979) and by *Glahn* (1985). Conditional forecasts of precipitation amount, besides being less skillful, fail to show this improvement.
3) Objective-guidance forecasts obtained directly or indirectly from dynamical prediction models show a skill that has gained substantially on the skill of subjective forecasts based upon them in part, and that is now near equality in a number of instances. This result is also consistent with findings of Zurndorfer *et al.* (1979).

What should the response of the forecasting community to this state of affairs be? Perhaps we should set the beating of guidance as a specific goal, thinking to assuage our wounded pride as well as to provide slightly better forecasts. But is this an effective use of human skill? Perhaps we should restructure the forecasting enterprise. Retail-specialist forecasters, weeding out the occasional nonsense that emerges from the computers, might concentrate on more effective ways of conveying the routine information to those who wish to hear it. Warning-specialist forecasters might concentrate on the few extreme events. We assume without much evidence, aside from that suggested by Firestone (1979), that experienced forecasters have a relative advantage here.

How skillful *should* forecasts be, and what rate of advance would be satisfying? There is great public interest in the subject of forecast skill, but these questions are never asked. Rather, there seems to be an unexamined assumption that "if we can put a man on the moon we ought to be able to forecast tomorrow's weather." There is a comparable attitude in medicine toward the curing of the common cold. With respect to the stock market, on the other hand, there seems to be little dismay at the limited skill in prediction a day ahead (although as with weather forecasts there is great interest in hearing about it). Why is more expected of us? Could it be, as Ramage (1982) has so succinctly put it, that "the great technological developments of the duodecennium have apparently belied the expectations of their proponents"?

Finally, what is the prognosis for the MIT forecasting contest? As Fig. 7 shows, there has been a slow decrease

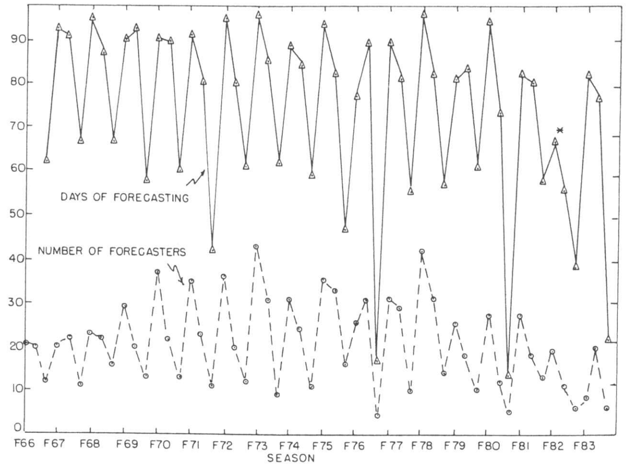


FIG. 7. Time series of number of days per academic term on which a consensus forecast was available, and number of contestants making more than one forecast in a given academic term.

over the last decade, both in the number of days over a given semester on which a consensus is assembled and the number of forecasters contributing to it. With the retirement of the author from MIT in the summer of 1984, we must anticipate a continuation of this trend. Perhaps it is just as well. At its inception this contest represented something new and untried. Surely there are more new things to try.

Acknowledgments. This enterprise would not have existed without the willing and eager participation of these forecasters: M. Abkowitz, R. Adams, E. Ahcor, F. Aicaradi, D. Aldrich, S. Allen, S. Alford, R. Anawalt, J. Anderson, G. Andrews, J. Ardini, J. Avallon, D. Baker, J. Barnard, C. Barker, D. Becker, L. Bender, A. Berger, R. Berger, L. Berman, V. Beyer, C. Biber, C. Bisel, M. Blank, P. Bloniarz, M. Blosser, H. Bluestein, L. Bodony, H. Boehme, A. Bommer, L. Bosart, G. Bosch, J. Boyd, J. Bradstreet, L. Branscome, B. Brenner, C. Bretherton, D. Brooks, J. Brown, G. Brownstein, D. Brush, R. Bryant, D. Bukvic, W. Burleson, R. Burpee, J. Burrows, C. Buscemi, J. Bushell, J. Caesar, B. Callahan, M. Cane, J. Carlson, T. Carlson, K. Carson, S. Cascione, A. Cassel, M. Cassity, S. Causbie, V. Chang, J. Chung, L. Chiu, D. Clark, P. Clay, T. Coate, R. Cohen, S. Cohen, F. Colby, B. Colman, R. Colten, A. Corwin, R. Cowan, E. Crawley, G. Cuscino, W. Daentl, S. Danberg, P. Dave, M. Davis, G. Deblonde, M. Dershowitz, G. Dewitt, R. Diamond, C. Dieterich, R. Dill, P. Dirks, J. Distefano, C. Dobbs, D. Dobos, D. Dockery, G. Dolan, R. Dole, E. Dorchak, G. Domm, M. Douglas, D. Duffey, D. Durran, J. Duszlak, J. Dweck, W. Ebisuzaki, R. Edson, H. Eglowstein, J. Elliot, K. Emanuel, B. Emery, K. Eng, R. Errico, D. Essa, F. Faller, M. Fantini, S. Ferguson, F. Ferris, D. Ferruzza, P. Fiekowsky, M. Finn, J. Fix, T. Fletcher, C. Fodor, J. Franklin, R. Franzosa, L. Freeman, J. Fullmer, I. Fung, E. Funk, J. Furhman, J. Gaertner, R. Gahl, M. Gamer, D. Gardner, S. Garner, A. Garrett, N. Gelband, S. Geotis, D. Gerhardt, M. Gilman, T. Glickperson, J. Glaudemans, M. Goldsmith, S. Golson, N. D. Gordon, N. B. Gordon, S. Gore, D. Gottfried, M. Graboe, C. Grant, J. Griggs, B. Greenberg, S. Greene, P. Guido, N. Guivens, P. Goldberg, W. Gutowsky, D. Gutzler, J. Gyakum, S. Gyoeri, R. Halstead, S. Hamilton, P. Hancock, R. Hankin, M. Hannig, B. Harrington, K. Haug. T. Hayes, G. Heffner, W. Henderson, S. Henderson, R. Hendl, G. Herman, R. Heuman, R. Hodgin, R. Hoffman, J. Holmes, R. Horn, T. Houghtby, R. Houze, M. Howard, G. Huffman, J. Hsiung, B. Ison, D. Jaffe, L. Jankauski, M. Jarvin, W. Johnson, G. Johnson, B. Jones, R. Jung, P. Justesen, J. Kaste, D. Katz, P. Kaufman, F.

Kenney, L. Keshishian, E. Kessler, J. Kester, J. Kidson, M. Keiper, W. Klein, T. Klein, S. Dur Ko, K. Kolodzy, J. Kowalski, J. Krakauer, J. Kristl, J. Kroll, KTPU, R. Kuklinski, T. Kush, P. Kuzmenko, A. Kyle, J. Lacorte, J. Lague, R. Lamson, T. Lavery, J. Lay, C. Leary, G. Lemke, M. Lepper, E. Lesser, J. Lester, L. Lesniewski, C. Lin, E. Lindstrom, J. Linn, P. Lui, N. Lukitsh, M. Lutz, D. Macintyre, W. Madden, J. Maguire, J. Maher, M. Majewski, M. Maniero, F. Marks, L. Marx, R. Martin, R. Mastrangelo, J. Matarese, T. Matejka, J. McCombie, J. McHenry, B. McKeller, B. McKee, K. McKinley, J. Meneely, G. Mejail, G. Melvin, J. Merritt, M. Middlebrook, P. Michael, D. A. Miller, D. Miller, S. Molica, J. Moses, S. Mudrick, B. Mullan, B. Myers, J. Myers, Nprime, T. Nehrkorn, P. Neilley, J. Neilsen, E. Nelson, D. Newman, G. Norman, R. Norton, C. Notis, P. Novick, S. O'Neil, T. Ozemko, R. Paine, K. Patch, D. Paton, L. Penn, L. Perrin, G. Perry, J. Pinkowish, A. Polansky, J. Pollack, D. Porter, J. Powley, J. Plotkin, A. Presser, W. Pritchard, T. Ransohoff, C. Rappaport, W. Raub, K. Redman, K. Redmond, D. Reiff, D. Reinhold, J. Richardson, P. Riegel, G. Riley, E. Rivas, J. Roads, P. Robbins, J. Robinson, A. Robock, S. Rodgers, P. Roebber, R. Rosen, H. Rosenblum, M. Rosen, J. Rosenfeld, L. Rosenstein, M. Rosenstein, M. Rosenthal, C. Rowell, K. Ruggles, M. Russell, D. Russell, S. Ryan, P. Sabella, H. Safren, J. Salley, S. Salamone, E. Sambuco, W. Sauer, K. Saunders, C. Schmandt, J. Schmidt, D. Schultz, D. Schwartzkopf, J. Scire, S. Scolnik, F. Scarabino, R. Selkirk, M. Seltzer, T. Semple, D. Shapiro, T. Shea, J. Sheldon, J. Shreter, L. Shulman, R. Shyderoff, W. Skamarock, W. Slade, C. Smith, J. Snyder, J. Soltes, H. Song, S. Soroka, S. Spiegel, S. Spura, R. Stall, R. Stallard, R. Steinweg, E. Steinhauer, P. Stellar, W. Stern, D. Stevens, D. Stewart, J. Stokes, T. Strat, S. Strauss, G. Street, K. Struhl, G. Swinton, M. Tanaka, B. Tavrow, L. Taylor, P. Tayntor, G. Ternullo, R. Thompson, J. Tillman, S. Tracton, K. Trenberth, A. Turner, P. Van Kleek, T. Vasicek, P. T. Venman, S. Vick, D. Vincent, D. Vogel, M. Wakim, H. Walker, B. Walker, J. Walsh, K. Walsh, F. Walter, P. Warner, M. Wasilewski, B. Weare, S. Webb, H. Weed, J. West, M. Whelan, S. Whitaker, P. Whitman, R. Whitney, R. Wilcox, D. Willett, H. Willett, J. Willett, F. Williams, P. Willissen, K. Wiltsee, M. Wojcik, W. Wolfe, D. Wolfsberg, P. Wolnik, R. Wong, D. Woodruff, P. Woods, K. Wooton, W. Wright, J. Wurman, P. Young, E. Yu, F. Zbar, S. Zebiak. The author is grateful to each of them.

Thanks are due to Isabelle Kole for the preparation of this manuscript and the accompanying figures. This investigation was supported by the National Science Foundation, most recently under grant ATM-8407142.

References

Baker, D. C., 1982: Synoptic-scale and mesoscale contributions to objective operational maximum-minimum temperature forecast errors. *Mon. Wea. Rev.*, **110**, 163–169.

Bosart, L. F., 1975: SUNYA experimental results in forecasting daily temperature and precipitation. *Mon. Wea. Rev.*, **103**, 1013–1020.

——, 1980: Evaluation of LFM-2 quantitative precipitation forecasts. *Mon. Wea. Rev.*, **108**, 1087–1099.

——, 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 of ranked catagories. *J. Appl. Meteor.*, **8**, 985–987.

Firestone, J. K., 1979: Comment on "Forecasting skill of beginners." *Bull. Amer. Meteor. Soc.*, **60**, 1206–1207.

Gedzelman, S. D., 1978: Forecasting skill of beginners. *Bull. Amer. Meteor. Soc.*, **59**, 1305–1309.

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

Gordon, N. D., 1985: The Kelburn forecast contest: Results from the first two years. New Zealand Meteorological Service Sci. Rep. **11**, 21 pp. [ISSN 0112-2398.]

Klein, W. H., and G. A. Hammons, 1975: Maximum-minimum temperature forecasts based on model output statistics. *Mon. Wea. Rev.*, **103**, 796–806.

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

Newell. J. E., and D. G. Deaven, 1981: The LFM-II model-1980. NOAA Tech. Memo. NWS NMC 66, 20 pp. [Available from Environmental Science Information Center (D822) NOAA, 6009 Executive Blvd., Rockville, MD. 20852.]

Pierce, C. H., 1976: Are weather forecasts improving? *Weatherwise*, **29**, 136–137.

Ramage, C. S., 1976: Prognosis for weather forecasting. *Bull. Amer. Meteor. Soc.*, **57**, 4–10.

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

Sanders, F., 1973: Skill in forecasting daily temperature and precipitation: Some experimental results. *Bull. Amer. Meteor. Soc.*, **54**, 1171–1179.

——, 1979: Trends in skill of daily forecasts of temperature and precipitation, 1966–78. *Bull. Amer. Meteor. Soc.*, **60**, 762–769.

Scanlon, R. J., and R. A. Anawalt, 1980: A survey of university weather forecasting games. *Bull. Amer. Meteor. Soc.*, **61**, 197–201.

Zurndorfer, E. A., J. R. Bochieri, G. M. Carter, J. P. Dallavalle, D. B. Gilhousen, K. F. Hebenstreit, and D. J. Vercelli, 1979: Trends in comparative verification scores for guidance and local aviation-public weather forecasts. *Mon. Wea. Rev.*, **107**, 799–811. ●