The Complex Relationship between Forecast Skill and Forecast Value: A Real-World Analysis

PAUL J. ROEBBER

Department of Geosciences, University of Wisconsin—Milwaukee, Milwaukee, Wisconsin

LANCE F. BOSART

Department of Atmospheric Science, State University of New York at Albany, Albany, New York

(Manuscript received 31 August 1995, in final form 5 June 1996)

ABSTRACT

For routine forecasts of temperature and precipitation, the relative skill advantage of human forecasters with respect to the numerical–statistical guidance is small (and diminishing). Since the relationship between forecast skill and the value of those forecasts is complex, the authors have examined their value across a range of real-world user contexts. It is found that although in most cases the meteorological information possessed considerable value to the users, human intervention in making those forecasts (as measured by National Weather Service forecasts) has generally led to minimal gains in value beyond that which is obtainable through direct use of numerical–statistical guidance. An important exception is the use of meteorological information by gas utilities during peak wintertime periods; in those circumstances, the value of human intervention was considerable. The presence of information in the National Weather Service forecasts independent of that contained in the numerical–statistical guidance was also established. Despite this, application of the additional information through a combined National Weather Service/guidance forecast provided only a small gain in value in most cases. In the most successful forecast context (the gas utility), the combined approach led to a loss of value relative to the unaltered National Weather Service forecasts.

However, recent trends toward increased skill in probability of precipitation forecasts have led to some gains in the relative value of the National Weather Service forecasts, concurrent with a shift toward smaller optimal cost–loss ratio distributions, findings that are significant with respect to practical business considerations. Furthermore, all of the applications studied showed the potential for considerable further growth in forecast value with continued increases in forecast skill. The relevance of our findings to the future of public and private meteorological forecasting is briefly discussed.

1. Introduction

There exists ample evidence substantiating the claim that synoptic-scale forecasts have shown significant increases in skill over the past three decades (e.g., Landis 1994; Kalnay et al. 1990; Shuman 1989). These improvements, reported largely in terms of measures of skill in forecasting height fields and associated winds at 500 and 250 hPa, also appear to be reflected in part in surface-based forecasts of temperatures and probability of precipitation, hereafter referred to as POP (Landis 1994). However, these latter skill improvements appear to have slackened considerably since the mid-1980s (see also section 3), an aspect that is particularly evident in the slow growth of skill in forecasting warm season POP (Fig. 2 of Landis 1994) and thunderstorm activity (Bosart and Landin 1994). Roebber and Bosart (1996) provided evidence that POP forecast skill at the 24–36-h range is now close to saturation (with respect to current model capabilities), whereas gains at shorter range (12–24 h) are likely being obtained through the application of specific knowledge concerning local controls on precipitation. Accordingly, the relative skill advantage of human forecasters with respect to the model output statistics (MOS) is small (and diminishing) and now largely reflects the ability of human forecasters to recognize those instances in which the MOS approach, which seeks to minimize forecast error by correcting for the systematic bias in the numerical model forecasts taken as a whole, does not account for bias that is specific to certain synoptic situations. Thus, further gains in POP forecast skill at 12-h and longer timescales will likely require significant improvements in operational numerical model forecasts on the mesoscale, such that forecasters will be able to recognize and account for smaller-scale flow features relevant to the occurrence of precipitation. Such gains in the numerical models

Corresponding author address: Dr. Paul J. Roebber, Department of Geosciences, University of Wisconsin—Milwaukee, Lapham Hall 352, P.O. Box 413, Milwaukee, WI 53201.
E-mail: roebber@csd.uwm.edu

would also be applicable to the MOS approach, and thus it is not clear a priori that such forecast gains would translate into continued human advantage with respect to the guidance. However, history suggests that such a continued advantage is likely. Olson et al. (1995) indicate that quantitative precipitation forecasters at the National Centers for Environmental Prediction (formerly known as the National Meteorological Center) have maintained their relative advantage in threat score with respect to numerical–statistical guidance throughout three decades of numerical model advances.

The issue of forecast value is related to the slow growth of skill in recent years and the diminishment in the skill gap between human forecasters and MOS (Sanders 1986; Roebber and Bosart 1996). Given small skill differences, one may ask whether *routine* National Weather Service (NWS) POP and temperature forecasts have any value above MOS. It is important to note that in this paper we shall address only routine NWS forecasts and do not seek to address the issue of quantitative precipitation forecasts (QPF) or public safety, two key forecast areas that hold considerable potential for economic impact. Further comments on this aspect are provided in the concluding discussion (section 4).

Studies of forecast value must consider explicitly the user context, which in its simplest form is expressed in terms of a cost–loss analysis. Murphy (1977) has shown from a theoretical basis that at the cost–loss extremes no-skill forecasts such as climatology and weather-independent decisions (always protect and never protect against adverse weather) approach the value of perfect forecasts. Since value is a function of decisions that are changed by the forecast, rather than necessarily the extent of the change, the relationship between forecast skill and value is, in general, complex. Readers interested in a concise overview of this issue should consult Murphy (1994) and the references listed therein. Accordingly, in section 2 we will discuss the techniques used for estimating forecast skill and forecast value for a range of real-world user contexts, along with a description of the forecast and verification data-set used for the subsequent analyses. Section 3 presents the details of the specific user contexts and presents the results of the analyses of skill and value for each of those cases. Finally, we will discuss the implications of the results in section 4.

2. Procedure

a. Analysis of skill

A skill score (SS), defined with respect to the mean square error (MSE) of a reference forecast that is the mean of the variable being forecast (climatology given a sufficiently long time series and a stationary climate, denoted here as MSE_c), can be written as

$$\mathrm{SS} = 1 - \frac{\mathrm{MSE}_f}{\mathrm{MSE}_c}, \qquad (1)$$

where the MSE_f is

$$\mathrm{MSE}_f = \frac{1}{N}\sum_{i=1}^{N} (f_i - O_i)^2 \qquad (2)$$

and f_i and O_i are the ith forecast and observation, respectively. Correspondingly, the MSE_c is defined by the substitution of climatological values for the forecasts, f_i, in (2), such that

$$\mathrm{MSE}_c = \frac{1}{N}\sum_{i=1}^{N} (\bar{O} - O_i)^2, \qquad (3)$$

where the (mean) climatological forecast is defined by

$$\bar{O} = \frac{1}{N}\sum_{i=1}^{N} O_i. \qquad (4)$$

The skill score defined by (1) is 1.0 for perfect forecasts, 0.0 for forecasts that are only as accurate as the reference (climatology), and negative for forecasts less accurate than the reference. Murphy (1988) discusses a further decomposition of this skill score, which allows for an examination of some of the properties of the joint distribution of forecasts and events (Murphy and Winkler 1987). However, in this study, our interest is to examine the relationship between forecast skill and value rather than the individual elements (such as bias) composing the skill score; accordingly, we shall report skill scores defined by (1) without recourse to further decomposition.

b. Analysis of value

For the purposes of examining the value of the forecasts, we shall restrict our analysis to decisions that are optimal from the point of view of expected return. In other words, we shall consider the prescriptive approach (how the forecasts should be used) rather than the descriptive approach (how they are used); thus, our analysis will present the maximum expected return for decisions based on the forecasts and the reference (climatology). It should be noted that nonoptimal decision making (a chronic feature of the real world) would in general lead to lower economic returns for both the forecasts and the reference. Consequently, the value of the forecast, defined with respect to the baseline measure, can sometimes be greater than is the case for that associated with optimal decision making, even though the expected return of the latter is greater. Thus, one should consider the values as reported in this paper to represent potential value, rather than actual value, which may vary in either direction as a result of nonoptimal constraints on the decision making process. Further discussion of these issues can be found in Katz and Murphy (1996).

	Adverse Weather	No Adverse Weather
Do Not Protect	a	b
Protect	c	d

FIG. 1. The structure of the basic payoff table with outcomes *a*, *b*, *c*, and *d*.

The most general description of value for a two-event, two-decision model takes the form of a payoff table, as shown in Fig. 1. Two events (adverse or nonadverse weather, for example, frost in a citrus farming application) and two decisions (protect or do not protect against the possibility of adverse weather, for example, spraying crops with water to protect against freezing) are provided, allowing for the possibility of four outcomes, each with an associated payoff. An optimal decision rule can then be constructed, which then takes the form of

$$p\,(\text{event}) \geqslant \frac{(b-d)}{(b-d)+(c-a)} \rightarrow \text{take protection}, \qquad (5)$$

where p = the forecast probability of the event, a = net payoff for adverse weather and no protection, b = net payoff for nonadverse weather and no protection, c = net payoff for adverse weather and protection, and d = net payoff for nonadverse weather and protection. Note that a net payoff in (5), which is an expense, is negative, such as might typically be the case for a, c, and d. In this rule, the expected return is optimized if the user protects against adverse weather when the forecast probability of such an event (p) exceeds the ratio defined by (5). For the special circumstances of $c = d$ = the cost of protection and a = loss (where the average net profit for each occasion, exclusive of the cost of protection or the losses suffered, is subtracted out such that $b = 0$), (5) reduces to the familiar cost–loss ratio [C/L or d/a in Eq. (5)]. A series of forecasts then produces a set of optimal decisions, based upon the assigned values of a, b, c, and d, and associated payoffs. The total payoff for the series of forecasts can then be computed according to

$$\text{total payoff} = \sum_{i=1}^{N} G_i, \qquad (6)$$

where N is the number of forecasts and G_i is the payoff associated with the ith decision, which takes on one of the values from a to d. Specific examples of these types of evaluations are presented in sections 3a and 3b. The determination of the value of that set of forecasts is then made by subtracting from (6) the total payoff when decisions are based upon forecasts produced by the standard of reference (e.g., climatology or MOS).

In some cases, it is more appropriate to express the value of a set of forecasts in terms of utility rather than in purely economic terms. For example, in many northern regions of the United States, municipal parks departments set up outdoor ice skating rinks in winter. This activity is effectively supported by local taxes and, where admission is not charged, it is generally not possible to assign dollar values to the four outcomes (a, b, c, and d). Here, a utility function can be defined, expressing the relative desirability of the outcomes. A specific example of this type of scenario will be discussed in section 3c.

Frequently, the above approaches, defining so-called static decision making, are not justified. More complex scenarios involving sequential decisions, in which actions taken and events occurring at a given point affect options and the consequences of those options at a later stage are often more appropriate; a specific example of such is discussed in section 3d.

In some circumstances, one wishes to evaluate cost savings rather than direct economic gains. Thus, costs associated with forecast errors are computed using a variety of forecasts (e.g., NWS, MOS, climatology, etc.); the value of such forecasts is defined by the cost savings, relative to the basic standard. An example of the cost savings approach is discussed in section 3e. It should be noted that the ''cost savings'' and ''payoff gains'' frameworks are fundamentally equivalent, since costs are simply negative payoffs for risk-neutral decision makers.

c. Forecast and verification data

The initial impetus for this study was provided by questions arising from the analysis of the skill of POP and temperature forecasts from the forecast contest at the State University of New York at Albany (SUNYA), as reported by Roebber and Bosart (1996). In documenting the small differences in skill between various constituent forecast groups, stratified according to education and experience, we became interested in whether these skill differences were meaningful in a more practical context such as forecast value. However, because of the constraints of the academic year, specific forecast periods are missing from the SUNYA data; for example, the contest is not held during the summer months (May–August), nor during the winter recess (mid-December–mid-January).

Consequently, we have turned to a dataset compiled by the NWS for the purposes of forecast verification, covering the period April 1966 through March 1994 at approximately 100 locations across the continental United States (Carter and Polger 1986; Landis 1994). For this study, we have extracted the following NWS and MOS forecasts (issued from the 1200 UTC forecast cycle) along with the accompanying observations for Albany, New York (ALB), for the period January 1970 through March 1994: (a) 12–24-h POP, (b) 12–24-h minimum temperature, (c) 24–36-h POP, and (d) 24–36-h maximum temperature. In addition, the NWS and MOS 12–24-h POP forecasts issued from the 0000 UTC cycle were extracted in order to focus on those people whose work activities the following day might be affected by precipitation and who might use this as the last forecast that they would hear before the workday began. Monthly climatological values of both POP and temperature for the associated 12-h periods were derived directly from the observations for the complete 29-yr period (1966–94). These values were then used as the basic reference for computing skill in (1) and also to define a baseline for the value calculations.

One limitation to our application of the forecast value analysis to forecasts and users specific to Albany is the necessary reduction in the generality of our results. However, given our experience in forecasting for other sites, we suspect that the basic results of this study are germane to the more general forecast problem and that many of our conclusions would hold across a wide variety of sites. A second question is whether our results would hold for a more general cross section of users than was possible to study in this paper. In section 4, we have made a preliminary attempt at testing the sensitivity of our results to arbitrary user cost–loss distributions.

Another limitation of the dataset is the application of a rain/no-rain methodology to situations in which the amount of precipitation may also be a determining factor. Since the NWS does not routinely make a public forecast of QPF, we are precluded from incorporating this aspect into our analysis. However, we suspect that QPF may indeed have considerable utility for many users and, as such, may be an interesting component in the debate concerning the future of forecasting. One application of QPF forecasts for specific users, probabilistic quantitative precipitation forecasts for river basins for use in NWS River Forecast Centers, has been discussed by Krzysztofowicz et al. (1993).

3. Results

Figure 2 shows the trend of annual NWS forecast skill score [defined by Eq. (1)] for 24–36-h POP and

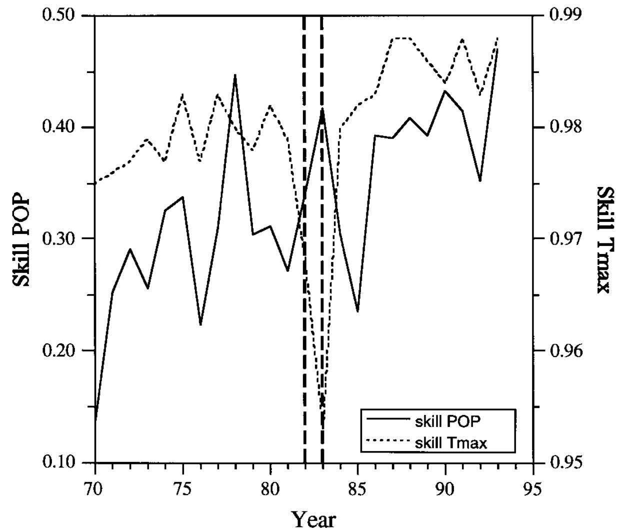


FIG. 2. The time series of NWS forecast skill for the period 1970–93 at Albany, New York, for 24–36-h POP (solid) and maximum temperature (dashed). The vertical dashed lines indicate the period of extensive data loss (January 1982–October 1983) cited in the text.

24–36-h maximum temperature at ALB for the period 1970–93. A major computer change in January 1982 caused an extensive loss of data until October 1983 (Carter and Polger 1986); consequently, skill scores during this period are not robust, and this time is indicated in the figure by the vertical dashed lines. Excluding 1982–83, the respective regression equations for the two forecast variables are skill (POP) = 0.239 + 0.008 (year-1970); skill (TMAX) = 0.976 + 0.000474 (year-1970), which yield an overall rate of improvement of about 0.8% per year for POP and 0.05% per year for maximum temperature.

It is obvious that skill in forecasting maximum temperature is well ahead of that for POP, a result consistent with 1) forecaster experience, 2) Brier score verification results from ALB and Boston (see Roebber and Bosart 1996; Sanders 1986; Bosart 1983) that have shown that POP skills are lower than temperature skills at comparable time periods and fall off more rapidly with time, 3) similar results showing that QPF skills are less than POP skills for the same time periods, and 4) spatial autocorrelation functions of temperature, height, and relative humidity/specific humidity from radiosonde data (Boyle 1981; Thiebaux 1975), which show that moisture drops off most rapidly with distance, indicating the inherently more mesoscale nature of moisture distributions.

Given the complex relationship between forecast skill and value and the substantial level of skill evident in temperature forecasts, one must wonder whether these slow advances have any practical utility. To investigate this question, we turn to the analysis of specific user contexts in the next several sections. Over the period 1970–94, a variety of numerical weather pre-

TABLE 1. Summary of model and guidance stratifications for the period 1970–94, for ALB Weather Service Forecast Office (WSFO) 24–36 h maximum temperature and POP forecasts.

Period	Parameter	Guidance	Model
Jan. 70–Jul. 73	Temp	Subjective/perfect prog.	Six-layer primitive equation (6L PE)
Aug. 73–Mar. 80	Temp	MOS	6L PE
Apr. 80–May 93	Temp	MOS	Limited fine mesh (LFM)
Jun. 93–Mar. 94	Temp	MOS	Nested grid model (NGM)
Jan. 70–Dec. 71	POP	Subjective	6L PE
Jan. 72–Mar. 80	POP	MOS	6L PE
Apr. 80–May 93	POP	MOS	LFM
Jun. 93–Mar. 94	POP	MOS	NGM

diction (NWP) models and associated guidance forecasts have been operationally available (Table 1). In the subsequent analyses, we shall refer generically to this guidance as MOS. The reader should also be aware that all dollar amounts are expressed in constant 1994 dollars, except where otherwise noted. Finally, the time series of skill and value for each of the applications are presented for the years 1970–93, 1994 is excluded as it represents only a 3-month dataset. However, summary information presented in Tables 3–5 includes the data from 1994.

a. *Concrete (flatwork) contractor: Norton Home Builders*

Telephone interviews with a local contractor (Norton Home Builders) allowed us to determine the manner in which meteorological information is used for flatwork concrete pouring (sidewalks, patios, basement floors, etc.). Specifically, the contractor must decide whether to risk pouring, given the possibility of rain, which can ruin the job at great expense and inconvenience. Through consultation with the contractor, it was determined that the payoff table approach described in section 2b was applicable to this situation. Specific dollar values were assigned to the analysis: the loss associated with rain ruining a concrete pouring job (which would entail the necessity of redoing the job after removing the ruined concrete), a, equals \$5500; the value of a successful pour, b, equals \$2000; the cost of protection (do not pour) against adverse weather (rain), c and d in this context, equals the cost of keeping the crew labor idle (about \$700). The decision rule, defined by Eq. (5), was then applied such that for forecast probabilities greater than or equal to 36%, the concrete was not poured. Some additional constraints were applied: the value analysis was run only for those days during the contracting season (April–October) on which daytime maximum temperatures exceeded 4.4°C (40°F).

The 12–24-h POP forecast issued from the 0000 UTC forecast cycle was used as the basis for this analysis. The lack of QPF forecasts renders the analysis somewhat simplistic, since the curing properties of the flatwork require relatively substantial precipitation (order 0.10 in. as opposed to the 0.01 in. used to verify measurable precipitation) for damage to ensue. However, since this is warm season work, convective rather than stratiform precipitation events would dominate, and local amounts might typically be relatively high for verifying episodes. We find further justification of our approach in the assurances of the contractor that the primary variable used to determine their pouring decisions was the NWS POP forecast.

Other weather-related effects could complicate the analysis. For example, a temperature dependence also exists for extreme high temperatures (~100°F or 37.8°C). Under such conditions, the quality of the concrete work deteriorates, yet the developing flaws are not evident until a couple of years later. Thus, the value of a joint forecast (temperature and precipitation) under such circumstances also requires some lagged component, since the poor work in real-world markets might affect the contractor negatively at some later date. However, we note that the maximum temperature at Albany exceeded 95°F (35°C) on only 20 days during the period 1970–94 and was as high as 99°F (37.2°C) on only 1 day in this period.

The results for the complete period from 1970–93 are shown in Table 2; the trend in annual value and skill of those forecasts is plotted in Fig. 3. Note that 1994 is not included since our data ends in March 1994. For this case, the meteorological information did not add value relative to climatology, which in this example was equivalent to the simple procedure of always

TABLE 2. Skill and value analysis of forecasts (1970–93) for concrete contractor. Negative values are enclosed in parentheses.

Forecaster	Skill score	Payoff per job	Value (above MOS)
Perfect forecast	1.000	\$1330.27	\$332.84
Climatology	0.000	\$1009.18	\$11.75
ALB WSFO	0.409	\$1007.94	\$10.51
MOS	0.382	\$997.43	\$0.00
NWS/MOS	0.421	\$1052.76	\$55.33
Always pour	—	\$1009.18	\$11.75
Never pour	—	(\$700.00)	(\$1697.43)

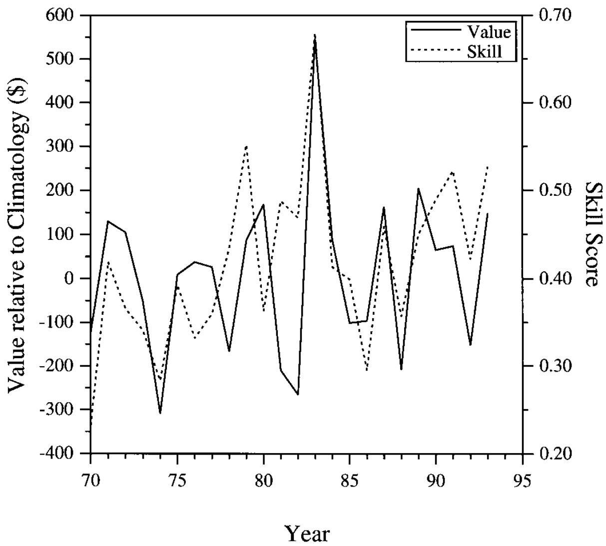


FIG. 3. The time series of the relative value of NWS forecasts (with respect to climatology, solid) and skill (dashed) for the period 1970–93 at Albany, New York, for the concrete flatwork application. The value is expressed as dollars per pouring job. Payoff table details are provided in the text.

pouring the concrete, regardless of the forecast. However, the value of perfect forecasts (relative to climatology) was $321.09, suggesting that improvements in forecast skill might allow such forecasts to improve their value. Unfortunately, Fig. 3 provides little comfort in that regard, with only very marginal gains in value in evidence over the past quarter century of operational forecasting. Worse still, an examination of the value of the NWS forecasts relative to MOS (Table 2) suggests that human intervention in this context has yielded little benefit to date.

In spite of these discouraging results, one should not necessarily conclude that the forecast enterprise is useless for this application. Interviews with the contractor revealed that he found the forecasts useful, because he is quite risk averse and preferred to avoid unsuccessful pouring jobs (even if this meant losses associated with missed opportunities). There is additional reason for hope: it is important to note that comparative performance and *independent* information content are two quite different concepts. Our analysis has shown that the NWS forecasts have limited value relative to MOS in this user context, but it may well be that the NWS forecasts possess significant independent information (which might represent itself in terms of value) from that contained in MOS. Clemen et al. (1995) have described the underlying conceptual and methodological frameworks for assessing independent information content. One means for assessing the presence of independent information is to examine whether combining two forecasts produces a consensus forecast of greater skill (or value) than the individual components (e.g., Vislocky and Fritsch 1995). In order to evaluate this aspect, we reran the analysis, using the combined forecast defined as the simple arithmetic average of the NWS forecast and the MOS forecast on each day. The resultant forecast (denoted as NWS/MOS in Table 2 and subsequent analyses) yielded a payoff per job of $1052.76, which represents a value with respect to climatology (MOS) of $43.58 ($55.33) per job. Thus, the presence of independent information in the NWS forecasts allows the possibility of obtaining additional value over that possible through direct application of either forecast alone.

b. Newspaper delivery: The Albany Times-Union

Most newspaper customers would like to avoid having their daily newspaper delivered in a soggy state; thus, precipitation poses a risk to newspaper delivery. Court (1982) attempted an analysis of this problem but found that the cost–loss analysis was complicated by the fact that a large percentage of customers would not complain and request a new delivery, but would simply seek to dry the existing paper (perhaps pausing only to curse the deliverer). In order to determine the value of forecasts in this context, we need controls for two unknown variables: the complaint rate, denoted by α, and the loss associated with precipitation ruining the paper and requiring a replacement (accounting for the costs associated with the extra delivery), denoted by β. Since the value of a successful delivery, with no protection against precipitation, b, equals $0.50 (the price of the newspaper), the net payoff for adverse weather and no protection, a, is $-\alpha\beta + \$0.50(1 - \alpha)$. The factors c and d in this context represent the payoff associated with protecting against adverse weather, which equals the value of a successful delivery ($0.50) minus the cost of wrapping it in plastic ($0.04, a cost obtained directly from the *Albany Times-Union*) or $0.46. A payoff table analysis [using the decision rule defined by Eq. (5) to determine whether or not to wrap the papers] was then run for all the days in the period from 1970–94 to assess the value of the forecasts for this user; the sensitivity of the results to the complaint rate (α) and replacement costs (β) was also tested. Again, the 12–24-h POP from the 0000 UTC forecast cycle was used to simulate the last information that would have been available to the newspaper carriers before delivery.

In order to determine whether the meteorological information adds value in this context, we have plotted (as a function of the two unknown variables, α and β) in Fig. 4 the difference in the daily payoff (based upon 102 717 daily deliveries, as reported directly by the *Albany Times-Union*) between the NWS forecasts and climatology. The relative value of the NWS forecasts is maximized along the bold-dashed line in Fig. 4. It is apparent that substantial value is added but only within a fairly narrow range of complaint rates for a fixed

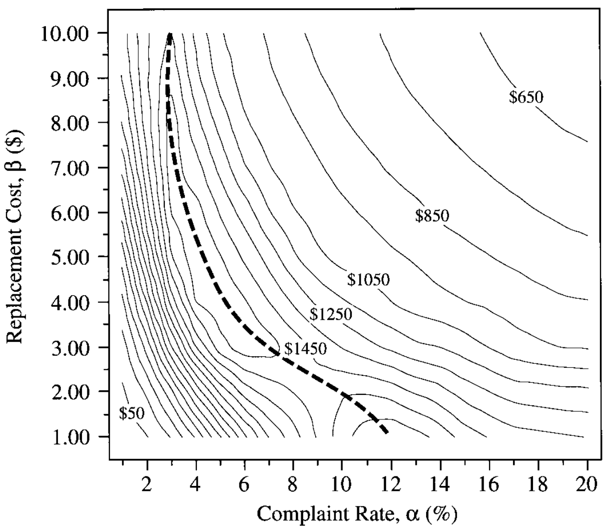


FIG. 4. The relative value of NWS forecasts (with respect to climatology) at Albany, New York, for the period January 1970 through March 1994 for the newspaper delivery application. The relative value is expressed as dollars per day, based upon 102 717 daily deliveries, as a function of complaint rate (α) and replacement cost (β). Payoff table details are provided in the text. The bold-dashed line denotes the axis of maximum relative value.

replacement cost. For moderate replacement costs (e.g., a 30-km round-trip at \$0.20 per km yields β = \$6.00), the forecasts quickly lose value with falling complaint rates (a conclusion that is circumstantially reinforced by the observation that wrapping newspapers is not a universal procedure). Gains in value with respect to the numerical–statistical guidance are small; for example, when the MOS payoffs are subtracted from the NWS payoffs, the resultant values along the dashed line of Fig. 4 are in the \$25–\$50 range, while peak values within a plausible range of α and β fail to exceed \$125. Nonetheless, with improvement in the forecasts, significant further gains in value are possible (e.g., the relative value of perfect forecasts with respect to NWS forecasts is \$1615.92 for α = 5% and β = \$6.00) and these gains are in evidence across a wide range of complaint rates and replacement costs (not shown). However, the trend in value and skill of those forecasts, plotted in Fig. 5 (with 1994 excluded due to partial data) for α = 5% and β = \$6.00, suggests that little gain in value has occurred in concert with the skill improvements in POP since the early 1980s (Fig. 2).

The question as to whether independent information in the NWS forecasts added value beyond MOS was also addressed. Some potential gain would appear possible, given that POP skill scores were 0.428, 0.404, and 0.444 for NWS, MOS, and NWS/MOS, respectively. In fact, the NWS/MOS forecast added \$50.71 of value for α = 5% and β = \$6.00, while peak values within a plausible range of α and β ranged upward to \$225.

c. Outdoor skating rinks: Town of Colonie, New York

In this example, the Colonie, New York, Parks and Recreation Department, through its general funds (ultimately supplied through local taxes filtered through the town budget for all recreational activities), sets up and maintains a series of outdoor skating rinks during the winter months. Because admission is not charged, it is difficult to analyze the forecast value using a payoff table approach. Instead, we turn to measures of relative utility by assigning numbers (on a scale of zero to one, with one representing highest utility) to the combinations of decisions with weather conditions. The decision is whether to flood the rinks with water, which under the appropriate conditions, provides a pleasant skating surface. The best weather for flooding is associated with temperatures below freezing and no precipitation. If the decision is to flood under these prime conditions, we have assigned a utility of 1.0. If the flood choice has been made with temperatures above freezing or precipitation occurring, we assign a utility of 0.0 (since this wastes labor and ruins the ice surface). More difficult is the assessment of the utility of the final two decisions and weather events. Although nonoptimal choices may occur, an important consideration to the operations of the town of Colonie is the ability to use such workers for other tasks within the department. For example, for the choice to not flood the ice under good conditions (temperatures below freezing and no

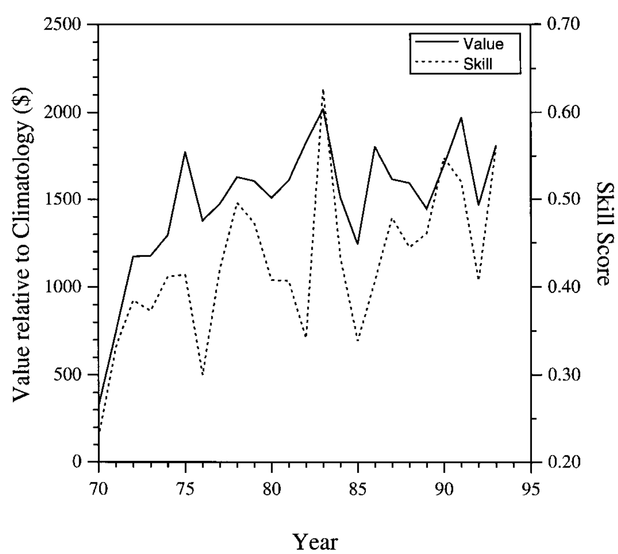


FIG. 5. The time series of the relative value of NWS forecasts (with respect to climatology, solid) and skill (dashed) for the period 1970–93 at Albany, New York, for the newspaper delivery application. The value is expressed as dollars per day, based upon 102 717 daily deliveries, with complaint rate α = 5% and replacement cost β = \$6.00.

TABLE 3. Skill and utility analysis of forecasts (1970–94) for maintenance of town outdoor ice skating rinks. The utility function is $a = 0.00$, $b = 1.00$, $c = 0.75$, and $d = 0.25$, as defined in Eq. (5).

Forecaster	POP skill	Temp skill	Utility
Perfect forecast	1.000	1.000	0.847
Climatology	0.000	0.000	0.525
ALB WSFO	0.397	0.774	0.735
MOS	0.374	0.742	0.723
NWS/MOS	0.408	0.782	0.735
Always flood	—	—	0.389
Never flood	—	—	0.56

precipitation), it is reasonable to set the utility to a nonzero value, since although an opportunity has been missed and the ice quality is degraded, the labor can still be applied elsewhere. If the choice is made to not flood the ice, and weather conditions are adverse, a utility less than unity is plausible, since although the proper choice has been made given the circumstances and the labor can be applied elsewhere, the most desirable outcome (a pleasant skating surface) has not been achieved. Initially, we will choose fixed utility values of 0.25 and 0.75 for these latter two choices, respectively. However, we shall also examine the sensitivity of our results to variable utility assignments. In keeping with seasonal considerations, we have restricted the analysis to the months December through February and because the labor assignments are made a day in advance, we have used the 24–36-h POP and maximum temperature forecasts issued from the 1200 UTC forecast cycle to analyze this case.

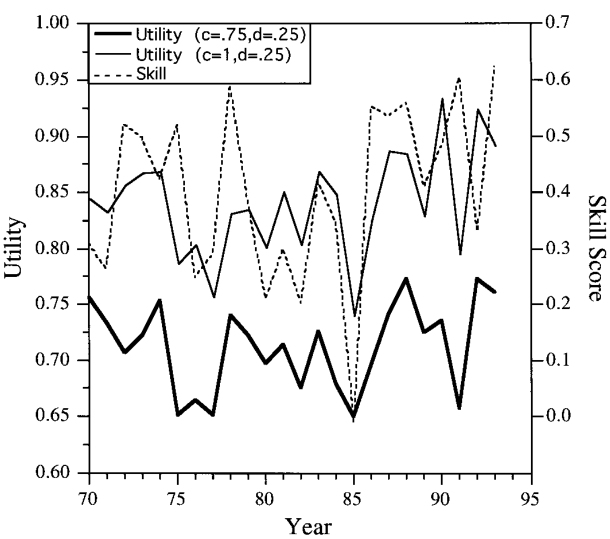


FIG. 6. The time series of NWS forecast utility (solid) and skill (dashed) for the period 1970–93 at Albany, New York, for the outdoor skating rink application. The utility is expressed on a scale of zero to one. Details are provided in the text.

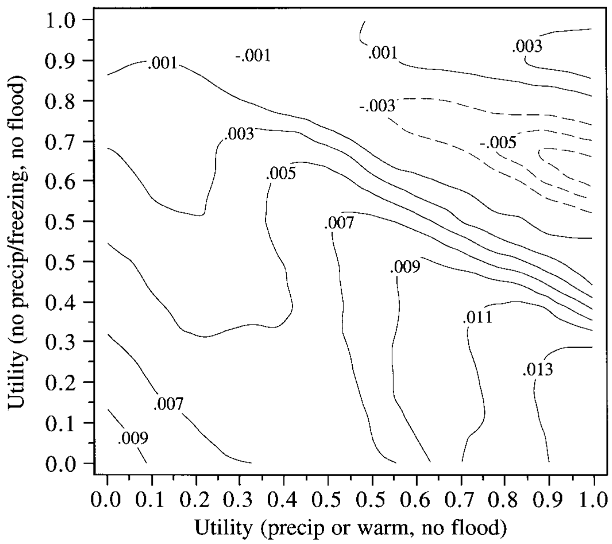


FIG. 7. The difference in utility between the forecasts from the NWS and MOS, for the period 1970–94, for the outdoor skating rink application, as a function of variable utility assignments for the poor condition/no-flood and good condition/no-flood situations. Details are provided in the text.

The results for the complete period from 1970–94 are shown in Table 3. Here, the meteorological information has considerable utility (as compared to the nonmeteorological strategies of always flood or never flood and the combined climatological probability of precipitation and subfreezing temperatures). In addition, there is room for growth in utility as forecast skill improves. However, the differences in utility between the NWS and MOS forecasts are fairly small, and no measurable gain is achieved by forming the combined NWS/MOS forecast. Figure 6, which shows the trend in the annual utility and skill of the NWS forecasts (with 1994 data excluded), demonstrates little growth in either variable for this user context during the entire period of record, a conclusion which does not appear to be very sensitive to reasonable modification of the utility assignments discussed above. As a final check on this statement, we have analyzed the difference in utility between the NWS and MOS forecasts across a complete range of utility assignments for the poor conditions/no flood and good conditions/no flood cases (Fig. 7). The results provide some encouragement in that the relative NWS forecast utility is maximized for reasonable ranges of these variables; however, the absolute differences are quite small, and there is little room for improvement by simply adjusting utility assignments.

d. Class AA baseball: The Albany–Colonie Yankees at Heritage Park

We have also examined the value of forecasts for Albany to a user with a more complicated decision structure than can be studied using a simple payoff table approach. Here, we are concerned with the value of forecasts to the local minor league baseball team (the Class AA Albany–Colonie Yankees at the time of the study). In this case, the grounds crew must decide whether to protect the infield against rainfall by unrolling the tarp. These choices are made for the overnight period, which makes the field available for practice the following day, and during the day, which preserves the field for the regularly scheduled evening game. We have restricted the analysis to the period of the Class AA season, roughly 10 April through 5 September and have used the 12–24-h POP (for the overnight period) and the 24–36-h POP (for the following day) from the 1200 UTC forecast cycle. These choices were driven by the scheduling needs of the Town of Colonie Parks and Recreation Department, which oversees the maintenance of the playing field.

A number of complications arise in performing such an analysis. In principle, a fully dynamic model of the decision process would need to incorporate the cumulative impact (on a seasonal timescale) of the number of rainouts on travel schedules and postseason commitments. In practice, it has been our observation that Class AA leagues take a more ad hoc approach to rescheduling missed games than in the more familiar example of Major League Baseball. Rescheduling is attempted on a priority basis for contending teams, subject to a myriad of other logistical considerations. Likely, the only means for accurately reconstructing such scheduling effects would be to research the complete sequence of games played in each of the studied seasons. Since such an analysis is beyond the scope of this study, we have adopted a sequential, decision tree approach for the 24-h period of each game sequence (overnight period and day of game). Thus, actions taken and events occurring at a given point in a sequence of decisions affect options, as well as the consequences of those options, at later stages in the overnight and following day periods *but are not considered to impact decisions at later dates in the season.* This simplification means that our projected losses due to rainouts are probably greater than would actually be experienced, since some of the missed games could eventually be replayed. We further assume that the decision makers are cognizant of the potential consequences of each of the options selected at decision points and choose the options that give the highest expected return (based upon the forecast POP). Decisions are then evaluated with respect to the weather conditions that actually occurred, and the payoffs are tabulated. The basic structure of the decision tree for this example is shown in Fig. 8.

An additional complication arises in this analysis since we must consider the value of practice. We know that it has some value, or the field would never be tarped overnight, since the overnight tarping preserves only the practice option, while tarping during the day

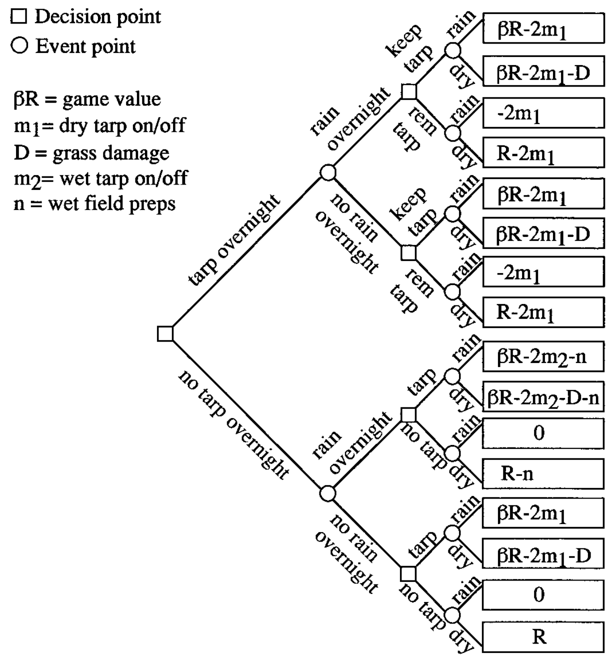


FIG. 8. The decision tree structure for the Class AA baseball application. Parameter values used in the analysis are $R = \$10,000.00$, $m_1 = \$35.00$, $m_2 = \$140.00$, $D = \$240.00$, and $n = \$70.00$.

preserves the evening game (at the expense of practice). To handle this complication, we suppose that if R represents the revenue attached to the game/practice combination, then αR is the practice value and βR is the game value, where $\alpha + \beta = 1$. We will examine two options: the first representing the case where the practice and game are equally valuable ($\alpha = \beta = 0.50$), and the second (likely more realistic) case where the game is much more valuable than the practice ($\alpha = 0.25$; $\beta = 0.75$).

The results of the analysis for the period from 1970–93 (with 1994 excluded since our data ends in March 1994, prior to the start of the Class AA season) are presented in Table 4. Again, although the meteorological information is quite valuable in both scenarios, in neither case does human intervention as represented by the NWS forecasts serve to add substantial value beyond that gained through direct use of the MOS forecasts. However, considerable further gains are possible with improving forecasts, and there is some evidence that the growth in skill since the early 1980s has led to corresponding increases in value over the same period (Fig. 9). Furthermore, the presence of independent information in the NWS forecasts can contribute to additional gains in value relative to the simple-minded application of the MOS forecast.

e. Gas utility

Suchman et al. (1979) have discussed the use of meteorological information by gas utilities. We have applied their approach, using values representative of a moderate-sized utility in the northeastern United States, to the Albany forecast data. Accordingly, the value results in this section should be interpreted as constant 1979 dollars.

The utilities seek to limit costs associated with additional demand during peak wintertime periods. These costs arise because the utility can only draw so much gas from the pipeline, and demand above this amount must be drawn from alternative and, generally more costly, sources (such as liquid natural gas, propane, or storage supplies). The controlling factors are the critical point, that is, the degree-day value above which alternative sources are needed and tolerance, the amount of error that can be accommodated without incurring cost. Errors that result in costs may arise in two situations: 1) the observed degree day exceeds the critical point, and the forecast error is in excess of the tolerance (in the case of an overforecast, excess gas is generated at expense while in the case of an underforecast, the alternative supplies are tapped); and 2) the observed degree day falls below the critical point, but the forecast exceeds the critical point and the forecast error is in excess of the tolerance (again resulting in the generation of excess gas).

For this analysis, we have assumed a critical point of 53 degree days (corresponding to a daily mean temperature of −11.1°C or 12°F), tolerance of 0.6°C (1°F), usage of 53.1×10^3 m^3 (1875×10^3 ft^3) per degree day, and a cost factor of \$123.60 $(1000\ m^3)^{-1}$ (\$3.50 $(1000\ ft^3)^{-1}$). Accordingly, we have limited the analysis to those days in which the critical point was exceeded (observations or forecasts), for the period 1970–94. The number of such days in any season ranged from 1 (1990) to 27 (1970), with a mean of 10.5 yr^{-1}. Clearly, the overall value of such forecasts is strongly dependent on the ''opportunity'' provided by exceedances of the critical point, with warm winters such as 1990 limiting the seasonal value of such forecasts. However, we note that no secular trend is apparent in the data and that January through March 1994 featured the second greatest number of exceedances in the entire period of record. The 12–24-h minimum temperature and 24–36-h maximum temperature forecasts issued from the 1200 UTC cycle were used for this analysis.

The overall results are summarized in Table 5. In this user context, it is clear that the meteorological information has tremendous value and, further, that human intervention has led to significant gains beyond MOS. As a result, in this application, the use of combined NWS and MOS forecasts degrades the value of the forecasts relative to NWS alone. Although MOS does forecast extreme values, experienced forecasters are well aware of the tendency for MOS to underforecast deviations from climatology during anomalous temperature regimes. Thus, in this user context, NWS forecasters are capable of producing forecasts of considerable value. We observe that progress in forecast skill could lead to further gains in the already substantial value of this information, although such gains will

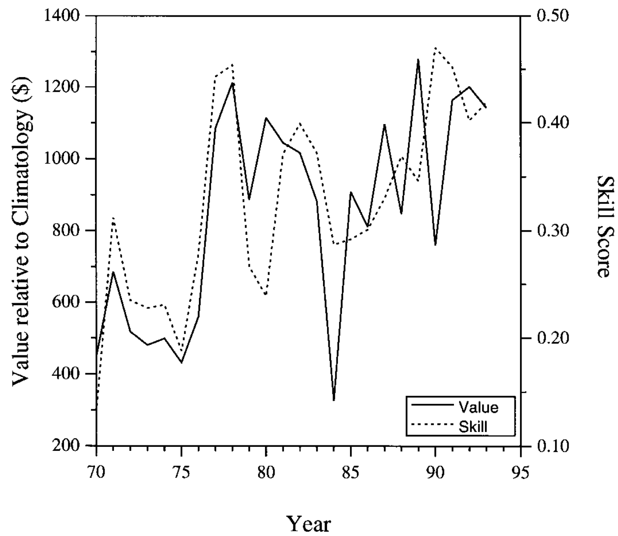


FIG. 9. The time series of the relative value of NWS forecasts (with respect to climatology, solid) and skill (dashed) for the period 1970–93 at Albany, New York, for the Class AA baseball application. The value is expressed as dollars per game, with $\beta = 0.75$. Details are provided in the text.

TABLE 4. Skill and value analysis of forecasts (1970–93) for class AA baseball club. Shown for each forecaster are the payoff (per game) and the value (relative to MOS) as a function of the relative importance of the game (β). Negative values are enclosed in parentheses.

		$\beta = .50$		$\beta = .75$	
Forecaster	Skill score	Payoff	Value	Payoff	Value
Perfect forecast	1.000	\$8635.28	\$974.20	\$9244.00	\$980.58
Climatology	0.000	\$7434.90	(\$226.18)	\$7409.34	(\$854.08)
ALB WSFO	0.328	\$7678.08	\$17.00	\$8258.92	(\$4.50)
MOS	0.311	\$7661.08	\$0.00	\$8263.42	\$0.00
NWS/MOS	0.344	\$7699.75	\$38.67	\$8296.94	\$33.52
Always tarp	—	\$4691.44	(\$2969.64)	\$7161.05	(\$1102.37)
Never tarp	—	\$7434.90	(\$226.18)	\$7434.90	(\$828.52)

TABLE 5. Skill and value analysis of forecasts (1970–94) for a moderate-sized northeastern U.S. gas utility. Shown are the value with respect to climatological forecasts and MOS. All dollar amounts are expressed in constant 1979 dollars and give the total cost (value) per heating season. Negative values are enclosed in parentheses.

Forecaster	Skill score	Cost	Value (Climatology)	Value (MOS)
Perfect forecast	1.000	$0.00	$1,181,241.51	$365,559.30
Climatology	0.000	$1,181,241.51	$0.00	($815,682.21)
ALB WSFO	0.904	$273,898.65	$907,342.86	$91,660.65
MOS	0.849	$365,559.30	$815,682.21	$0.00
NWS/MOS	0.895	$287,382.45	$893,859.06	$78,176.85

become increasingly difficult given the currently high levels of skill exhibited for the degree-day forecasts.

4. Concluding discussion

The complex relationship between skill and value has been demonstrated, both across forecast systems (e.g., NWS, MOS, climatology) within a particular user context, and for a particular forecast system within a range of years. For example, in the concrete pouring application, although climatology was the lowest ranked of the five forecasts in terms of skill, it outperformed two of the higher-ranked forecasts in terms of value. Such shifts were also apparent for the outdoor ice rink and the Class AA baseball applications. When the annual time series of the NWS forecast skill is correlated with the value (relative to climatology) of those same forecasts, linear correlations ranged from .566 (gas utility) to .750 (newspaper delivery). Thus, the ability of the skill score to predict (linearly) the value of a set of forecasts for a particular application is quite limited.

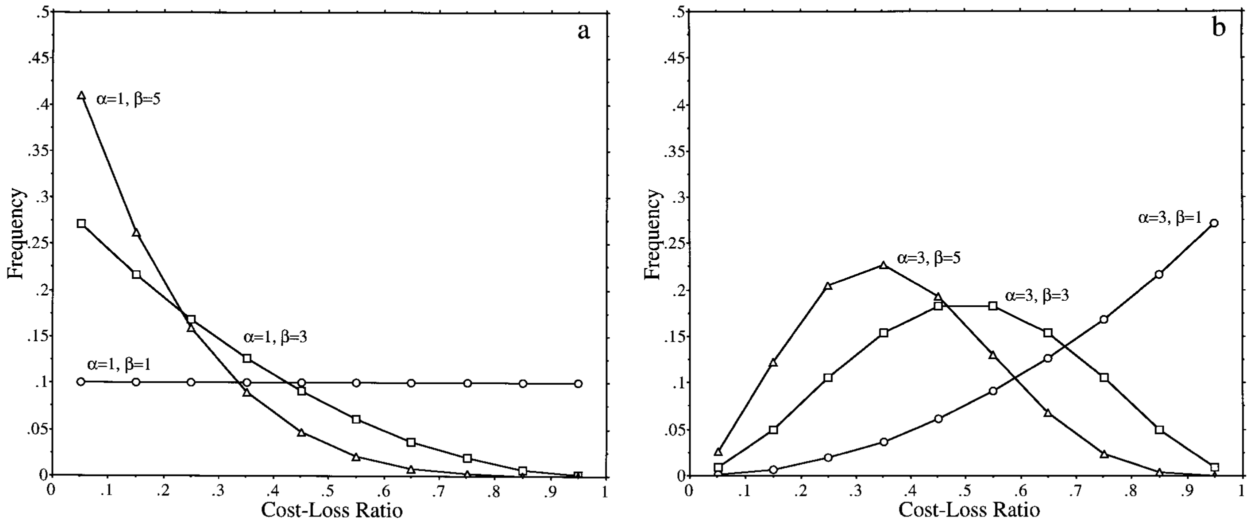
Accordingly, we have examined the value of the Albany NWS forecasts across a selected set of user contexts, in an effort to assess the meaning of the small (and diminishing) skill advantage that human forecasters currently enjoy over the numerical–statistical guidance represented by MOS. We have found that although the meteorological information in most cases possessed considerable value to the users, human intervention has generally led to minimal additional gains beyond that provided by MOS. An important exception was the use of meteorological information by gas utilities during peak wintertime periods; in those circumstances, the value of human intervention was considerable. The presence of information in the NWS forecasts independent of that contained in MOS was established. However, application of this additional information through a combined NWS/MOS forecast led to mixed results: in most cases, the use of this information provided a small gain in value, while in the most successful forecast context—the gas utility—this approach led to a loss of value (relative to the unaltered NWS forecasts).

One aspect of this issue that we have not addressed is the sensitivity of our results to the full cross section of users in the population. Our sampling of users has undoubtedly left out many important sectors, whose cost–loss factors differ significantly from those that we have studied. One approach to answering this question, on a preliminary level, was suggested to us by A. Murphy (1995, personal communication). This approach is based upon modeling the distribution of user cost–loss contexts and examining the value of forecasts as a function of the distributions so defined. Accordingly, we have followed Murphy (1969) and assumed that the cost–loss distribution of users takes the form of a beta distribution:

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1 - x)^{\beta-1}, \qquad (7)$$

where x is the cost–loss ratio ranging from 0 to 1 and Γ is the factorial function. Since the POP forecasts are reported in 10s of percent (0%, 10%, etc.), we have computed cost–loss user distributions based upon 10 equal-sized bins of width 0.1 (centered on the midpoint), ranging from 0.05 to 0.95. For example, the percentage of users in the economy with a cost–loss value of 0.05 is determined by the integral of (7) on the interval 0–0.1. The distribution parameters α and β were varied from 0.5 to 5 to represent a wide range of cost–loss distributions. Examples of the form of the beta distribution determined in this way as a function of the parameters α and β are shown in Fig. 10. Readers should note that the beta distribution exhibits a peak at $x = \alpha/(\alpha + \beta)$ and the variance is $\alpha\beta/(\alpha + \beta + 1)(\alpha + \beta)^2$; thus for $\alpha = \beta$, the distribution is symmetric (the limiting case of $\alpha = \beta = 1$ yields the uniform distribution), while for α greater (less) than β, the distribution is skewed to the left (right).

We then computed the value of the 24–36-h POP forecasts for the period 1970–94 for each of these idealized cost–loss distributions, with the results scaled such that perfect forecasts take the value of $1000 and climatological forecasts have zero value (Fig. 11). Figure 11a shows that over the period 1970–94 the NWS POP forecasts have added value across a wide spectrum of user contexts; this value is maximized in the region where α is small and β is large, representing the cases where the majority of user cost–loss ratios are small. However, the relative value of human intervention



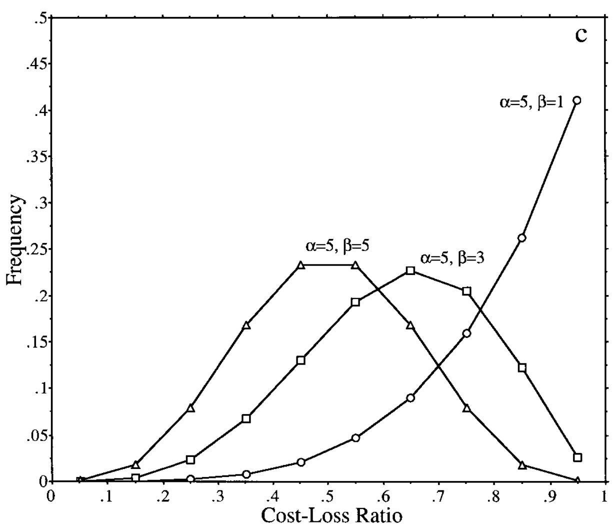


FIG. 10. Examples of hypothetical cost–loss distributions in the user community, modeled after a beta distribution. The parameters α and β are varied, with $\beta = 1$ denoted by O, $\beta = 3$ by □, and $\beta = 5$ by Δ, for (a) $\alpha = 1$, (b) $\alpha = 3$, and (c) $\alpha = 5$.

(with respect to MOS, defined as NWS value minus MOS value) is fairly small throughout the entire cost–loss distribution space (Fig. 11b), with maximum benefit (order $14) falling approximately along the line $\alpha = \beta + 1$. These distributions represent the cases where cost–loss ratios are centered somewhat above 0.5. We have also tried modeling the cost–loss distribution using the conceptually simpler Gaussian (normal) distribution, based upon a range of mean population cost–loss ratios with variable variance. Our results (not shown) were comparable to those obtained using the beta distribution, namely, that the relative value of the NWS forecasts (with respect to MOS) was maximized in a focused distribution region near mean cost–loss ratios of 0.5 with small variance.

We suspect that the generally modest improvements in value (despite measurable advantages in skill) that human intervention has demonstrated beyond MOS reflects the overall sensitivity of many users to precipitation, and the corresponding (relative) lack of skill in this forecast area. Skill improvements in precipitation forecasting have been demonstrably slow, and skill levels remain fairly low. In contrast, skill levels in forecasting temperatures are quite high, and thus decisions based upon this criteria can be finely tuned (e.g., in the case of the gas utility). This point is reinforced by inspection of the changes in 24–36-h POP forecast value during the past decade (1985–94; Figs. 11c and 11d). The increases in POP skill during this period have led to some increases in the overall value of the NWS forecasts, both with respect to climatology (Fig. 11c) and with respect to MOS (Fig. 11d). Furthermore, the maximum benefits during this period appear to fall along a line near $\alpha = \beta$, suggesting that there has been a shift in the optimal cost–loss ratio distribution toward lower values in recent years. This result has some practical

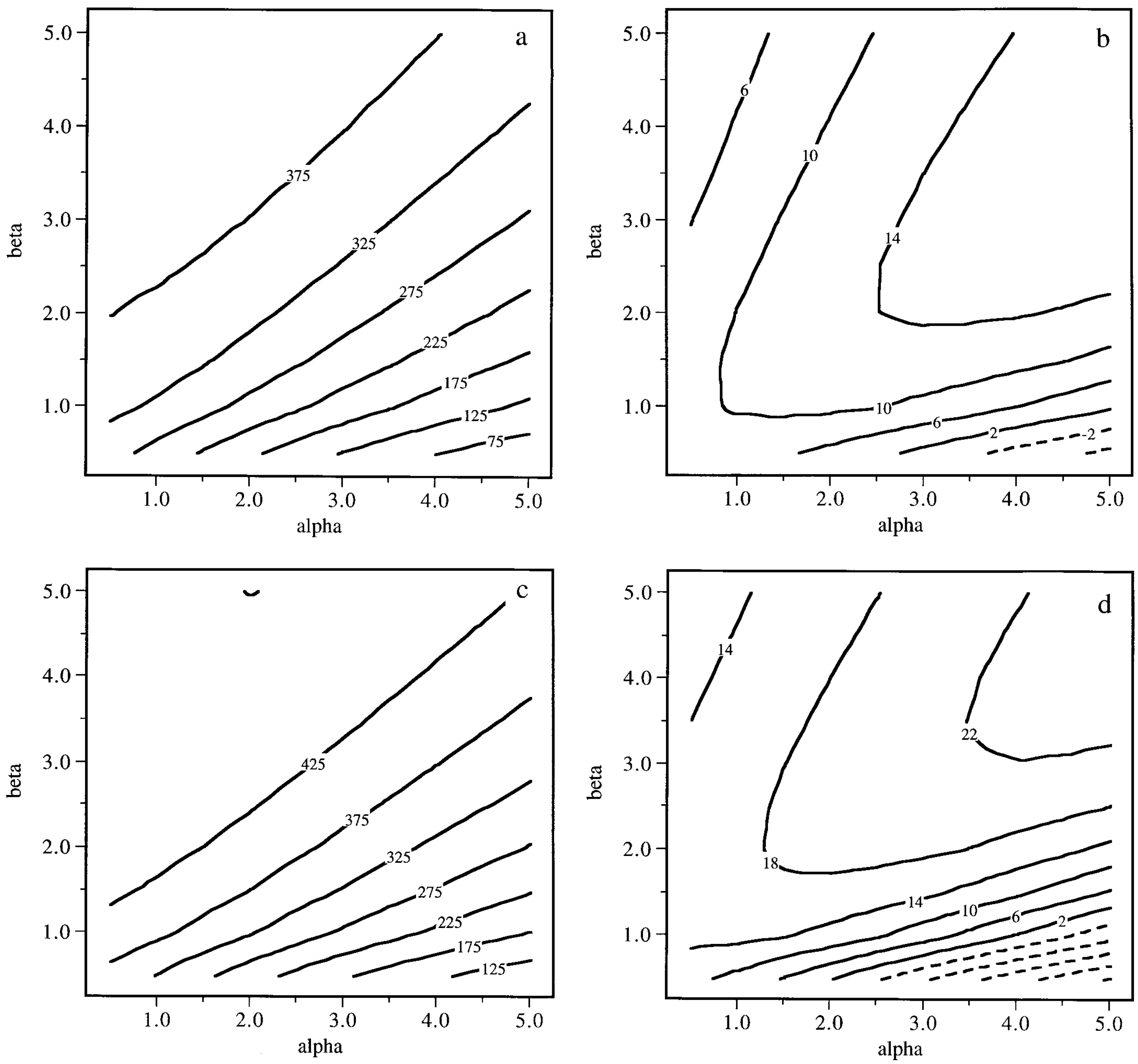


FIG. 11. The relative value of NWS forecasts with respect to climatology and to MOS at Albany, New York, as a function of user cost–loss distribution for 24–36-h POP. Value is scaled such that a perfect forecast is $1000.00 and a climatological forecast is $0.00. Shown are (a) value relative to climatology for the period January 1970–March 1994, (b) value relative to MOS for the period January 1970–March 1994, (c) value relative to climatology for the period January 1985–March 1994, and (d) value relative to MOS for the period January 1985–March 1994.

importance, since high cost–loss ratios represent a difficult operational environment from a business standpoint (the cost of protection is too high to maintain financial stability). As another check on these results, we have performed a similar analysis, using the 12–24-h POP issued from the 0000 UTC forecast cycle. Figure 12 shows the overall value of the NWS forecasts with respect to climatology (Fig. 12a) during the 1985–94 period and the value of the NWS forecasts with respect to MOS (Fig. 12b) for the same period. Except for an enhancement of the overall value of the forecasts (consistent with, although not guaranteed by, a concomitant increase in skill level), these results are in agreement with our findings for the 24–36-h POP forecasts. The presence of independent information in the NWS forecasts was found to be a source of some small additional value in several user contexts. As a check on these results, we show in Fig. 13 the overall value of the combined NWS–MOS forecast (relative to the MOS forecast alone) for a range of hypothetical

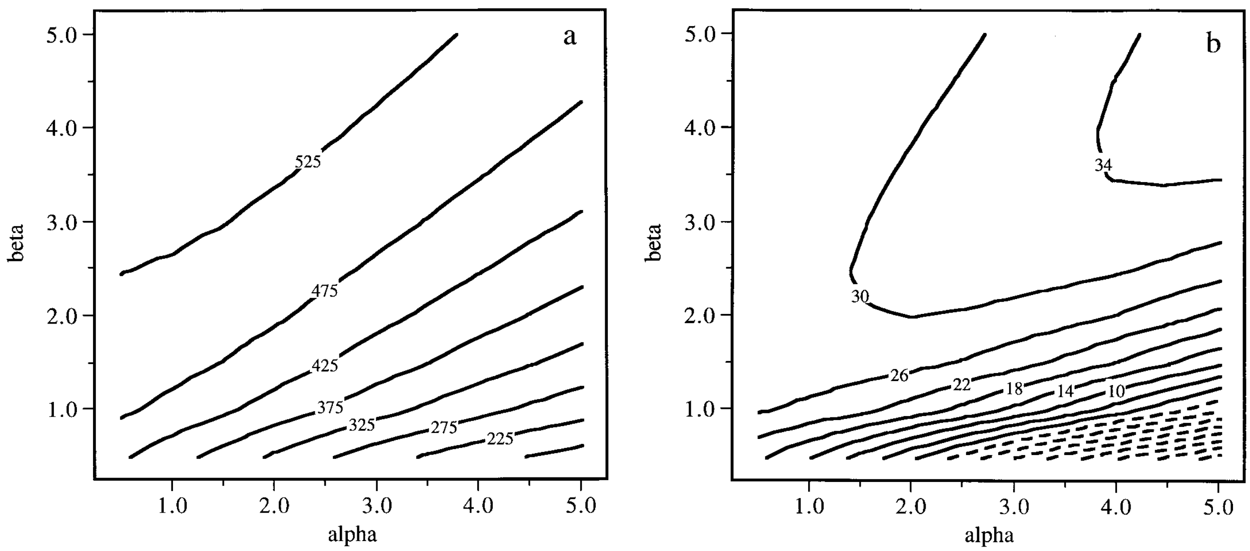


FIG. 12. As in Fig. 11 but for 12–24-h POP issued from the 0000 UTC forecast cycle. Shown are (a) value relative to climatology for the period January 1985–March 1994 and (b) value relative to MOS for the period January 1985–March 1994.

cost–loss distributions. The results for the 24–36-h forecasts (Fig. 13a) and the 12–24-h forecasts (Fig. 13b) are similar and provide mild encouragement: the independent information adds some value across a wide range of user distributions, with maximum effect occurring for small α and large β (corresponding to distributions with peak frequencies at small cost–loss ratios; Fig. 10a).

A component of value that we have not attempted to quantify here is the ability of the user to interact with the provider of those services. Such interactions often allow the user to better assess the degree of certainty in the forecasts and, even more basically, to draw out specific information that is pertinent to their concerns. One might also argue that such interaction would allow providers of forecast information to better tune their

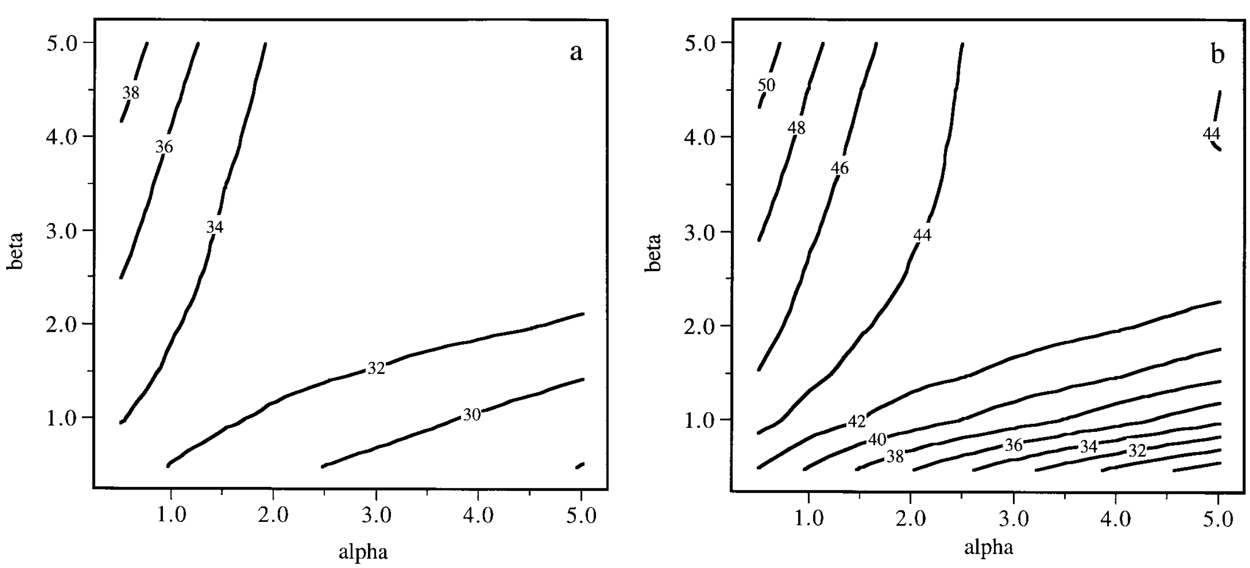


FIG. 13. The relative value of the combined NWS–MOS POP forecasts with respect to the MOS forecasts alone at Albany, New York, for the period January 1970–March 1994 as a function of user cost–loss distribution. Value is scaled such that a perfect forecast is \$1000.00 and a climatological forecast is \$0.00. Shown are (a) relative value for 24–36-h POP, issued from the 1200 UTC forecast cycle and (b) relative value for 12–24-h POP, issued from the 0000 UTC forecast cycle.

forecasts to the users' needs. H. Brooks (1995, personal communication) has pointed out that Fig. 11 hints at a possible increase in the role of the private sector in meteorological forecasting. Because it is difficult to add value for a wide range of users with different needs, it might be possible for private forecast services to achieve greater success than has been demonstrated by the NWS, by providing specific forecasts within narrow customer bases, and so maximizing the potential value of those forecasts. The value of such user–forecaster interactions is demonstrable (in part) by virtue of the fact that many users of weather information subscribe to private providers of forecast services, although all providers of meteorological information (public and private) rely on identical information, and likely possess comparable levels of skill.

Thus, while the absolute magnitude of the value added by human intervention has been and continues to be small, recent trends in improved forecasting of POP have resulted in encouraging trends in the value of those forecasts, which might bode well for the future (a point underlined by the fact that all of the specific users studied in this paper showed that improvements in forecast skill could lead to substantial increases in value). Whether humans will be able to exploit the same kind of pattern recognition techniques for POP and QPF that are currently employed for temperature (see Roebber and Bosart 1996) to the same advantage over MOS, remains to be seen.

Regardless of future potential, we wish to emphasize that our results should not lead to the conclusion that human intervention in weather forecasting currently adds little or no positive value to the economy as a whole. Although we have studied a range of users, we have focused narrowly on user contexts pertaining largely to the rain/no-rain scenario and have not addressed issues related directly to QPF or public safety, a fundamental task of the NWS forecast enterprise. Hudlow (1993) reported that 75% of presidentially declared natural disasters result from floods. A recent, dramatic example of such an event is the Great Flood of 1993 (U.S. Department of Commerce 1994), in which damages exceeded $15 billion and 48 fatalities were attributed to the disaster. Numerous additional examples were noted by Olson et al. (1995). It is likely that improvements in the skill of QPF would lead to comparable improvements for the river forecast system, which should yield substantial economic value throughout the user community. Evidence already exists that improvements in the forecasting of hurricanes and tornadoes has led to significant reductions in the fatalities resulting from these natural disasters (Sheets 1990; Ostby 1992). Clearly, these forecast tasks are of considerable value and represent a role that has not as yet been superseded by the independent performance of numerical models. We believe that our results serve to emphasize the conclusions of Roebber and Bosart (1996) that the most profitable means of retaining human involvement in the forecast process would be to provide the primary focus on those weather situations that pose a significant threat to public safety, with a complementary effort toward an emphasis of nowcasting on the local and regional scale. Shifting the focus of forecasting from routine to critical tasks (public safety and/or specific user needs) would appear to be an effective means of maximizing forecast value in an era of shrinking scientific resources.

Acknowledgments. Partial support for this research was provided by National Science Foundation Grant ATM-9114598. Ideas for this research began to germinate through the continued involvement of the authors in the practical forecast problem, including most recently the STORM-FEST field program of 1992. The following individuals, representatives of the users of meteorological information discussed in this paper, were uniformly helpful and deserve our thanks: Jack Norton of Norton Home Builders; Don Myers, Superintendent of the Town of Colonie Recreation and Parks Department; Ray Joyce and Bob Logan of Niagara-Mohawk Gas. We also wish to thank Harold Brooks of NOAA/NSSL and the two anonymous reviewers for their comments on an earlier version of this paper.

REFERENCES

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.
——, and M. G. Landin, 1994: An assessment of thunderstorm probability forecasting skill. *Wea. Forecasting,* **9,** 522–531.
Boyle, J. S., 1981: Autocorrelations of moisture parameters on isentropic surfaces. *Mon. Wea. Rev.,* **109,** 2401–2404.
Carter, G. M., and P. D. Polger, 1986: A 20 year summary of National Weather Service verification results for temperature and precipitation. NOAA Tech. Memo. NWS FCST-31, National Weather Service, NOAA, U.S. Department of Commerce, 50 pp. [Available from U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Silver Spring, MD 20910.]
Clemen, R. T., A. H. Murphy, and R. L. Winkler, 1995: Screening probability forecasts: contrasts between choosing and combining. *Int. J. Forecasting,* **11,** 133–146.
Court, A., 1982: Wrap or not? News delivery dilemma. *J. Appl. Meteor.,* **21,** 537–539.
Hudlow, M. D., 1993: Briefing on modernized hydrologic forecasting services. Office of Hydrology, NWS/NOAA, U.S. Department of Commerce, 36 pp. [Available from U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Silver Spring, MD 20910.]
Kalnay, E., M. Kanamitsu, and W. E. Baker, 1990: Global numerical weather prediction at the National Meteorological Center. *Bull. Amer. Meteor. Soc.,* **71,** 1410–1428.
Katz, R. W., and A. H. Murphy, Eds., 1996: *Economic Value of Weather and Climate Forecasts.* Cambridge University Press, in press.
Krzysztofowicz, R., W. J. Drzal, T. R. Drake, J. C. Weyman, and L. A. Giordano, 1993: Probabilistic quantitative precipitation forecasts for river basins. *Wea. Forecasting,* **8,** 424–439.
Landis, R. C., 1994: Comments on ''Forecasting in meteorology.'' *Bull. Amer. Meteor. Soc.,* **75,** 823–827.
Murphy, A. H., 1969: Measures of the utility of probabilistic predictions in cost-loss ratio decision situations in which knowledge of the cost-loss ratios is incomplete. *J. Appl. Climatol.,* **8,** 863–873.

——, 1977: The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situation. *Mon. Wea. Rev.,* **105,** 803–816.

——, 1988: Skill scores based on their mean square error and their relationships to the correlation coefficient. *Mon. Wea. Rev.,* **116,** 2417–2424.

——, 1994: Assessing the economic value of weather forecasts: An overview of methods, results and issues. *Meteor. Applications,* **1,** 69–73.

——, and R. L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.,* **115,** 1330–1338.

Olson, D. A., N. W. Junker, and B. Korty, 1995: Evaluation of 33 years of quantitative precipitation forecasting at the NMC. *Wea. Forecasting,* **10,** 498–511.

Ostby, F. P., 1992: Operations of the National Severe Storms Forecast Center. *Wea. Forecasting,* **7,** 546–563.

Roebber, P. J., and L. F. Bosart, 1996: The contributions of education and experience to forecast skill. *Wea. Forecasting,* **11,** 21–40.

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

Sheets, R., 1990: The National Hurricane Center—Past, present and future. *Wea. Forecasting,* **5,** 185–232.

Shuman, F. G., 1989: History of numerical weather prediction at the National Meteorological Center. *Wea. Forecasting,* **4,** 286–296.

Suchman, D., B. A. Auvine, and B. H. Hinton, 1979: Some economic effects of private meteorological forecasting. *Bull. Amer. Meteor. Soc.,* **60,** 1148–1156.

Thiebaux, H. J., 1975: Experiments with correlation representations for objective analysis. *Mon. Wea. Rev.,* **103,** 617–627.

U.S. Department of Commerce, 1994: Natural disaster survey report, the Great Flood of 1993. NOAA, NWS, Silver Spring, MD, 181 pp. [Available from U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Silver Spring, MD 20910.]

Vislocky, R. L., and J. M. Fritsch, 1995: Improved model output statistics forecasts through model consensus. *Bull. Amer. Meteor. Soc.,* **76,** 1157–1164.