J6.2 Probability Forecasting, Quantitative Decision Modeling, and the Economic Value of Forecasts

Daniel S. Wilks
Cornell University, Ithaca NY
dsw5@cornell.edu

Introduction

"Using probability forecasts in decision making is what its all about, right? Of course, Ed knew this . . ." (Bob Glahn)

Why do we make weather forecasts? Why do we engage in research aimed ultimately at improving those forecasts? Because users of those forecasts need information about the future behavior of the atmosphere to improve their decision making.

Ultimately, the economic justification for both research and operations in the atmospheric sciences rests on the suitability and use of forecasts to support decision making by forecast users. Probability forecasts impart more value than do their nonprobabilistic counterparts, and the probabilistic format is essential for separating the role of the forecaster from that of the decision makers who use the forecasts. Both of these assertions can be demonstrated using quantitative decision-making models, which have roots in Bayesian statistics, and can be used to compute potential (assuming optimal use) economic value of forecasts in particular settings.

We use probability forecasts to quantify uncertainty about future atmospheric behavior. Uncertainty about future atmospheric behavior is fundamentally a consequence of "chaotic dynamics", or sensitivity to initial-condition uncertainty. Our best practical approach to dealing with this initial-condition uncertainty in dynamical models of the atmosphere is presently through use of ensemble forecasting, where the dynamical model is integrated multiple times from initial conditions that are (ideally) independent random draws from the probability distribution characterizing the initial-condition uncertainty. Although not strictly part of my remit in preparing this review, I recently re-read Epstein's (1969a) famous paper on stochastic-dynamic forecasting. To my surprise I discovered that he appears to have been the inventor of ensemble forecasting! In Section 4 of that paper he describes "Monte Carlo solutions," which he uses there only to provide verification data for his second-moment-closure analytical calculations. Yet the recipe he provides is essentially the method that is used operationally today:

"Discrete initial points in phase space are chosen by a random process such that the likelihood of selecting any given point is proportional to the given initial probability density. For each of these initial points (i.e. for each of the sample selected from the ensemble) deterministic trajectories in phase space are calculated by numerical integration . . . Means and variances are determined, corresponding to specific times, by averaging the appropriate quantities over the sample." (Epstein 1969a)

Ed Epstein clearly understood why uncertainty about the atmosphere will always be inescapable, the importance of using probabilities to quantitatively characterize that uncertainty, and the value to individuals and to society at large that would flow from their use. He was instrumental in advancing these issues, and much of that work was carried forward and extended by his Ph.D. student, Allan Murphy. As I was, in turn, Allan's student I am pleased to have been invited to review this portion of Ed's work. Epstein's contributions to probability forecasting and decision making also have close connections with other aspects of his work, notably stochastic-dynamic forecasting, but especially Bayesian statistics and forecast verification. I will also take note of these connections.

Probability Forecasts

"[P]robability statements are more useful than the conventional forecast to a broad segment of the users of the forecasts. The evidence for this is so clear that the wide use of probability statements would have occurred much sooner, were it not for uncertainties regarding the use and interpretation of these statements on the part of the public and especially on the part of the meteorologists who must issue them." (Epstein 1966b)

Although probabilistic weather forecasting in some form dates from perhaps the late eighteenth century (Murphy 1998), the first description of recognizably modern probabilistic forecasts, explicitly associating a numerical probability with occurrence of a clearly defined future event, may have been by Hallenbeck (1920). Experimental subjective precipitation probability forecasts for selected locations in the U.S. were initiated in the mid-1950's by the Travelers Weather Service (a private weather forecasting service) (Murphy and Winkler 1984), and by the U.S. Weather Bureau (predecessor to the current National Weather Service) (Root 1962). These probability-of-precipitation (PoP) forecasts became official, nationwide operational Weather Bureau forecast products in 1965 (Murphy 1998, Murphy and Winkler 1984).

The meaning of these new PoP forecasts was, and continues to be, the probability that at least 0.01" of (liquid-equivalent, in the case of frozen) precipitation occurs at a specific measurement location during the forecast valid period. Literature from that time (e.g., Epstein 1966a; Murphy and

Winkler 1971a,b; Murphy and Winkler 1974; Murphy et al. 1980; Rogell 1972, Scoggins and Vaughan 1971), and even from the present day (e.g., Williams et al. 2014) reveals some confusion and misunderstanding about the meaning of PoP forecasts on the part of not only the general public, but also among forecasters and statistically literate individuals as well. Epstein produced three papers that addressed these problems concerning interpretation, interpretability, and public acceptance of probability forecasts.

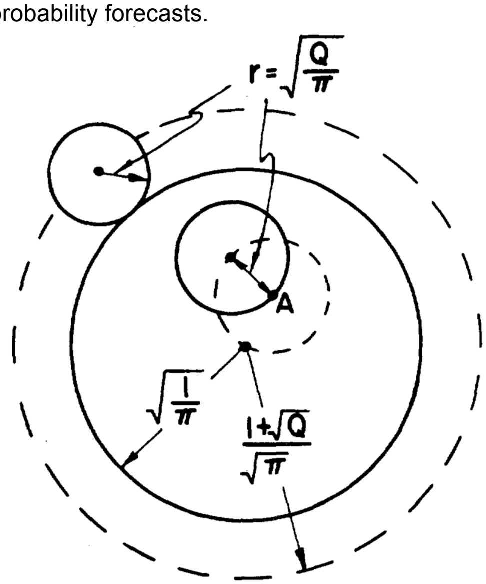

FIGURE 1.—Any cell located within the large dashed circle produces precipitation in the forecast area (large solid circle). Any cell located within the small dashed circle produces precipitation at station A. The small solid circles represent precipitation cells.

Even though format of the new PoPs were specifically for point locations, one aspect of the confusion about their meaning was whether they pertained instead to occurrences over larger areas, and the nature of the relationship of the point and areal probabilities. In the paper "Point and Area Precipitation Probabilities" (Epstein 1966a), Epstein analyzed an idealized model of random, circular rainfall areas in relation to a circular forecast area (Figure 1). In this model there are N precipitation cells per unit area, each having area Q, distributed over an area that is large compared to the unit forecast area. Using a Poisson distribution for the number of cells overlapping a particular point location within the forecast area, Epstein derived the relationship between the point probability P_p and the area probability P_a,

$$1 - P_a = (1 - P_p)^{[1+\sqrt{1/Q}]^2} , \qquad (1)$$

so that the point and area probabilities are nearly equal for large (synoptic-scale) rain areas Q, but the point probability is much smaller than the area probability for small (convective-scale) rain areas. As an aid to forecasters, who might be able to estimate the average m of a probability distribution for the number of precipitation cells per unit area, he then derived the expected values of the point and area probabilities, assuming known rain cell size and exponential probability distribution for the numbers of cells,

$$E[P_p] = \int_0^\infty (1 - e^{-NQ})(1/\mu)e^{-N/\mu}dN$$
$$= Q\mu/(Q\mu + 1) \qquad (2a)$$

and

$$E[P_a] = \int_0^\infty (1 - e^{-N[1+\sqrt{Q}]^2})(1/\mu)e^{-N/\mu}dN$$
$$= [1+\sqrt{Q}]^2\mu/([1+\sqrt{Q}]^2\mu + 1) \quad . \qquad (2b)$$

He also obtained somewhat more general expressions, assuming gamma distributions for the density of rain cells.

Reacting to logical inconsistencies in a worded forecast that included a precipitation probability statement, Curtiss (1968) proposed that precipitation probabilities should be formulated and interpreted in terms of expected areal coverage. In a formal response, Epstein (1968) concluded that point probabilities are better suited to user needs, that specifically forecasting areal coverage would not enhance the information content of the forecasts if it is implicit that the forecast probability applies uniformly throughout a forecast area of interest, and that explaining the meaning of expected areal coverage to the general public would introduce unnecessary complications.

Curtiss (1968) also questioned whether "the precipitation 'probabilities' which are being released to the public [are] really interpretable as action probabilities in day-to-day decision making . . .?". This question goes directly to the practical usefulness of probability forecasts, and therefore to the very rationale for forecasting in the first place. As noted in Epstein's other 1966 paper, "Quality Control of Probability Forecasts" (Epstein 1966b), whether forecast probabilities are "actionable" is equivalent to the question of whether, after transformation to betting odds, the uncertainty information in the forecast represents a fair bet. For example, if a forecaster believes the probability of precipitation tomorrow is 1/3, would that forecaster be equally comfortable taking either side of a low-stakes 2:1 odds bet against? Readers who may be uncomfortable about conceptually mapping their day-to-day decisions in terms of betting odds should realize that most decisions that are not trivially easy require that uncertainty in the outcome must be confronted, so that in effect there are many such "bets" in life.

In this "Quality Control" paper, Epstein (1966b) addresses the issue of quantitative interpretation of the new subjective probability forecasts, both on the part of the forecasters who would be recalibrating their judgment processes in light of forecast verification data (the "quality control" referred to in the title of the paper), and on the part of forecast users interested in the reliability of the forecasts (i.e., how literally the subjective forecast probabilities could be believed). He makes the statement that begins this section by way of motivating the analysis of these problems.

To address these problems, Epstein took an explicitly Bayesian view of probability, as a quantified degree of belief rather than as an expression of long-run relative frequency. Here he represents the forecaster's or forecast user's uncertainty about the (precipitation occurrence) event (equivalently, what would be fair betting odds), which he called *p*, given the forecaster's stated probability, which he called ρ, using standard beta distributions. Standard beta distributions have support on the unit interval, and are conventionally expressed in terms of the two distribution parameters α and β as

$$f(p) = \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right] p^{\alpha-1}(1-p)^{\beta-1} \quad . \qquad (3)$$

However, Epstein (1966b) reparameterized the standard beta distribution, assuming that the subjective probability forecasts were unbiased so that the distribution is defined by its mean $\alpha/(\alpha+\beta) = \rho$, and a parameter *L* indicating confidence in the stated probability, yielding the conditional beta probability density

$$f(p \mid \rho) = \left[\frac{\Gamma(L)}{\Gamma(L\rho)\Gamma(L-L\rho)}\right] p^{L\rho-1}(1-p)^{L(1-\rho)-1} \quad . \qquad (4)$$

Figure 2, taken from Epstein (1966b), shows some examples of these distributions. Larger values of the confidence parameter *L* yield distributions that are more concentrated, since the variances of these distributions are $\rho(1-\rho)/(L+1)$. Some perspective on the caution with which the new PoP forecasts were regarded can be had from Epstein's observation that his own prior (i.e., without having seen any verification data for the PoP forecasts) distributions "seem to correspond best to *L* = 10." So, for example, at the time and without the benefit of verification data, Epstein suggest that his interpretation a PoP forecast of ρ = 0.5 was near certainty with respect to outcome probability larger than 0.05 and smaller than 0.95, but with nontrivial probabilities anywhere in the range from 0.2 to 0.8!

The device of reparameterizing the beta distribution using the confidence parameter *L* is convenient within the context of Bayesian updating for a Bernoulli probability (e.g., Epstein 1985), in light of new information. In the present setting of recalibrating PoP forecasts, the dichotomous Bernoulli event is occurrence of at least 0.01" of precipitation in a forecast valid period, and the new information consists of verification data for *L* occasions on which the PoP forecast ρ had been issued. In addition to their capacity to represent a wide variety of functional forms, beta distributions are convenient in this context because they are conjugate to the Bernoulli likelihood, meaning that a beta posterior distribution results when updating a beta prior distribution.

Consider, then, prior distributions in the form of Equation 4, with parameters *L* and ρ, to will be

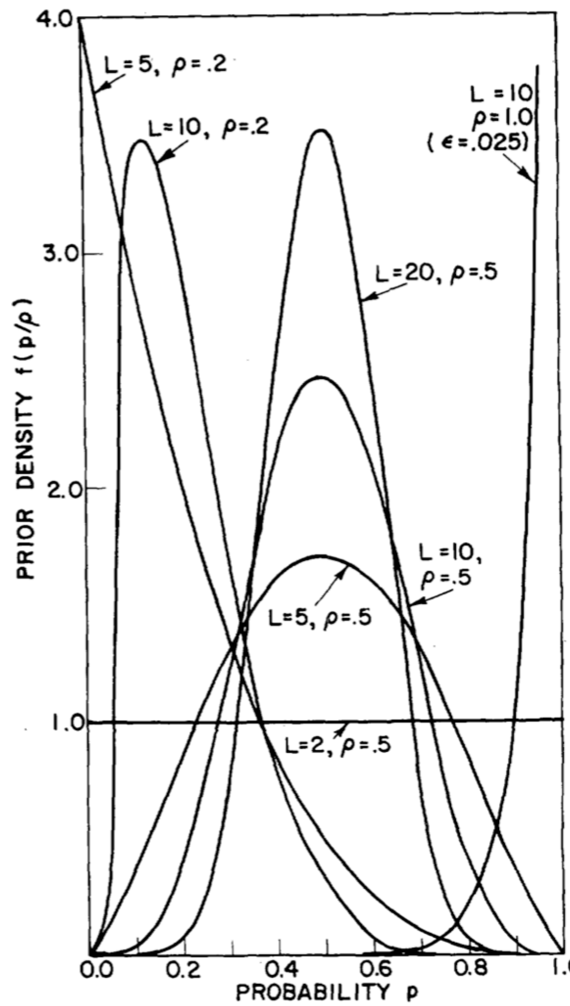

FIGURE 2.—Analytic form of prior degree of belief for several pairs of values of the parameters *L* and *ρ*.

updated after seeing verifications for *n* forecasts where PoP = ρ, and in which precipitation occurred *r* times. The result is another beta distribution, with conditional mean

$$\mu = (L\rho + r)/(L+n) \qquad (5)$$

and conditional variance

$$\sigma^2 = \frac{[L\rho + r][L(1-\rho)+n-r]}{(L+n)(L+n+1)} \quad . \qquad (6)$$

To pursue the example of Epstein's hypothetical mid-1960's prior distribution for *p*, given a forecast ρ = 0.5 and a confidence factor *L* = 10, if the results of *n* = 10 forecasts of ρ = 0.5 identify *r* = 5 instances of precipitation, the posterior mean μ is unchanged, but the posterior variance decreases from 0.0227 to σ^2 = 0.0119, which is the same distribution associated in Figure 2 with *L* = 20 and ρ = 0.5.

When large samples of forecast verification data are available, the posterior mean in Equation 5 converges to the empirical relative frequency, *r*/*n*, and the variance approaches (*r*/*n*)(1–*r*/*n*)/*n*, regardless of the parameters of the prior beta distribution. As a practical matter, very large verification data samples are available for nationally aggregated probability forecasts. For example Murphy (1985) shows sample sizes of $n \approx 12{,}000$ for a half-year of PoP forecasts across the U.S. in 1980, yielding standard deviations around 0.025 for the posterior distributions of ρ. The implication is that probability forecasts can be taken at face value for quantitative decision making given the large available verification data samples, when they indicate $\rho \approx r/n$ to good approximation for each possible probability forecast value ρ.

Decision Models: Overview

"Weather forecasts possess no intrinsic value in an economic sense. They acquire value by influencing the behaviour of individuals and organisations . . . whose activities are sensitive to weather conditions." (Murphy 1994)

I opened this paper with Bob Glahn's observation regarding the importance of, and connection between, probability forecasting and decision making. Having reviewed some of Ed Epstein's contributions to probability forecasting I will now move on to talk about the potential for their use in decision making. As meteorologists we tend to assume that our forecasts have value but as the quote from Murphy points out, any value economic must be mediated by the use to which they are put.

The analytical framework of decision analysis (e.g., Berger 1980, Clemen 1991, Winkler 1972) is a convenient and powerful approach to linking forecasts and their optimal use, and in the process to compute rational estimates of the economic value of forecasts. It also allows quantitative estimation of the expected economic value of potential future forecasts based on assumed improvements in accuracy. This framework, which has roots in Bayesian statistics, was apparently first introduced to the meteorological literature by Epstein's (1962) paper, "A Bayesian Approach to Decision Making in Applied Meteorology." Much of that paper is devoted to Bayesian calibration of forecasts given limited verification data, but since operational probability forecasts have demonstrated excellent calibration over the course of many years (e.g., Murphy 1985), in light of the results in Equations (5) and (6) I will assume in the following that available probability forecasts are well calibrated as issued and can be taken at face value.

Decision analyses are "prescriptive" in that they specify in advance what optimal decisions will be given particular forecasts and user contexts, rather than describing how decision makers actually use forecasts. In particular they assume rationality on the part of the decision maker, in the sense that optimal decisions will be those that maximize the statistical expectation of a relevant objective function. The basic structure of the analysis rests on the following elements:

1. Available Actions. The modeled decision maker must have at least two alternative actions to choose among: a_i, $i = 1, \ldots, I$. These are enumerated fully and explicitly.

2. Events, or "states of nature." Here these will be specific meteorological events x_j, $j = 1, \ldots, J$; of which there must again be at least two in order for the analysis to be meaningful.

3. Probabilities for the events. Here these are derived from the forecasts f_k, $k = 1, \ldots, K$; whose calibration yields the conditional event probabilities $p(x_j|f_k)$. In the case of operational PoP forecasts, which are well-calibrated probability forecasts, $p(x_j|f_k) = f_k$ (i.e., they "mean what they say") to good approximation. In the more general setting these conditional probabilities are Bayesian posterior probabilities.

4. Outcomes. Each possible combination of an action and an event leads to a specific quantified outcome, $O(a_i, x_j)$, and in order for the decision analysis to be meaningful there must be a preference order among these outcomes. Often, and most straightforwardly, the outcomes will be expressed in monetary terms. More generally the outcomes are expressed in units of "utility", which can represent the risk attitude of a decision maker with respect to possible outcomes. For example, if you would prefer receiving \$1000 for sure rather than the prospect of \$2000 on the flip of a coin, your utility for the \$1000 is more than half your utility for \$2000, $U(\$1000) > U(\$2000)/2$, and your utility function for money is said to be nonlinear. Expressing the outcomes in simple monetary terms is the special case of linear utility. Figure 3 schematically compares nonlinear (risk averse and risk seeking) and linear (risk-neutral) utility functions for money.

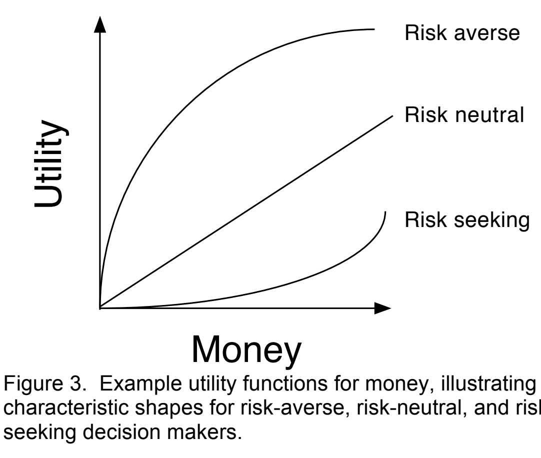

Figure 3. Example utility functions for money, illustrating characteristic shapes for risk-averse, risk-neutral, and risk-seeking decision makers.

The most straightforward application of decision analysis involves so-called "static" decisions, which are either one-off situations, or at least decisions that are independent of the results of other decisions. The objective function to be maximized in the decision making is the expected (probability-weighted average) outcome,

$$O^*(f_k) = \max_i \sum_{j=1}^{J} O(a_i, x_j)\, p(x_j \mid f_k) \quad . \qquad (7)$$

That is, there is an optimal action a_i* associated with each of the *K* possible forecasts. If the forecasts can make a difference to the decision maker, some of these optimal actions, associated with different forecasts, will be different from each other. Otherwise the forecasts will not have economic value, as indicated by the quote from Murphy at the beginning of this section.

The statistically expected (or, equivalently, long-run average) economic return to be derived from optimal use of the forecasts being considered can be computed as the probability-weighted average of the optimal outcomes in Equation (7),

$$ER = E_f[O^*(f_k)] = \sum_{k=1}^{K} O^*(f_k)\, p(f_k) \quad , \qquad (8)$$

where the probability weights come from the refinement distribution $p(f_k)$, which expresses the frequencies-of-use of each of the possible forecasts f_k. The refinement distribution together with the *K* calibration distributions $p(x_i|f_k)$ in Equation (7) contain the full information from the joint frequency distribution of the forecasts and observations, through its calibration-refinement factorization (Murphy and Winkler 1987):

$$p(f_k \cap x_j) = p(x_j \mid f_k)\; p(f_k) \quad . \qquad (9)$$

Thus, comprehensive forecast verification, and in particular what is called "diagnostic verification" (Murphy et al. 1989, Murphy and Winkler 1992), is central to optimal forecast use and to evaluation of the economic worth of forecasts.

The economic value of forecasts is a relative quantity, which requires specification of some baseline, possibly naive, information that would be available to the decision maker in the absence of the forecasts. In many analyses this will be an unchanging climatological forecast, f_0, which will have associated with it a single optimal action a^*_0 that maximizes

$$O^*(f_0) = \max_i \sum_{j=1}^{J} O(a_i, x_j)\, p(x_j \mid f_0) \quad . \qquad (10)$$

The expected value of the forecasts, relative to the constant baseline forecast f_0 is then simply the difference between the respective expected economic returns,

$$EV = ER - ER_0 = ER - O^*(f_0) \quad , \qquad (11)$$

where the second equality is consistent with Equation (8) because for the unchanging climatological forecast $p(f_0)$=1. In general the economic value to be derived from forecasts depends on the actions a_i available to the decision maker, the decision maker's outcome function $O(a_i, x_j)$, the information baseline against which the forecasts will be compared, and the quality of the forecasts as expressed in their joint distribution with the observations (Equation 9). Forecasts can achieve positive economic value only if some of the optimal decisions associated with them are different from the default baseline decision a^*_0, and therefore also not all equal to each other. Different decision makers having different outcome functions may well derive different levels of economic benefit from the same forecasts.

Example: The Simplest Possible Decision Model

"Neither has much attention been given to the added facility with which user requirements of weather forecasts can be met given specific information on uncertainties." (Epstein 1969a)

The general decision-analytic structure outlined in the previous section can perhaps be most easily illustrated using what has come to be known as the Cost/Loss ratio situation. This prototypical decision problem was apparently first proposed by Thompson (1950), who motivated it as a consequence of "the principle of calculated risk," and has been used many times since then as a simple, instructive, and analytically accessible prototype for more general and realistic decision settings (e.g., Roebber and Bosart 1996). An equivalent model was derived by Gringorten (1950) as part of an illustration of why the roles of forecaster and decision maker should be kept separated, and could be kept separated through the use of probability forecasting; and a similar structure was derived several decades earlier using nonprobabilistic forecasts by Angstrom (1922) and Bilham (1922).

The Cost-Loss ratio situation is the simplest possible decision problem, because it treats only the minimum numbers of actions (*I* = 2) and states of nature (*J* = 2). To fix ideas, it is usual to imagine that the two states of nature correspond to occurrence (x_1) or not (x_2) of a weather state that is adverse to an enterprise, and that a decision can be taken (a_1) or not (a_2) to protect against the adverse weather at a cost –*C*. If the protective action is taken the effects of the adverse weather are assumed to be completely negated, and if the adverse weather occurs without protection a loss –*L* is incurred. Accordingly there are four possible outcomes $O(a_i, x_j)$: $O(a_1, x_1) = O(a_1, x_2) = -C$, $O(a_2, x_1) = -L$, and $O(a_2, x_2) = 0$, which are illustrated graphically in Figure 4.

		Adverse Weather?	
		x_1 = Y	x_2 = N
Protect?	a_1 = Y	– C	– C
	a_2 = N	– L	0

Figure 4. Illustration of the outcome function for the Cost/Loss ratio decision problem.

Consider now probability forecasts f_k for the adverse weather event, which for simplicity (and consistent with real-world experience with operational

PoP forecasts) are assumed to be calibrated, so that the (respective Bernoulli) calibration distribution probabilities satisfy $p(x_1|f_k) = f_k$ and $p(x_2|f_k) = 1 - f_k$. The maximization in Equation (7) will then compare the two quantities

$$-C f_k - C\,(1 - f_k) = -C \tag{12a}$$

for the a_1 (protect) action, and

$$-L f_k - 0\,(1 - f_k) = -L f_k \tag{12b}$$

for the a_2 (do not protect) action. Thus $i = 1$ (protection) maximizes Equation (7) when $f_k > C/L$, yielding $O^*(f_k) = -C$ in these cases. Similarly $i = 2$ (no protection) maximizes Equation (7) when $f_k < C/L$, yielding $O^*(f_k) = -L\, f_k$ under this condition. Location of the decision threshold at C/L is the origin of the name "Cost/Loss ratio" problem.

Table 1. Refinement distribution for a set of hypothetical probability forecasts for a binary predictand. The example in the text assumes the forecasts are perfectly calibrated so that $p(x_1|f_k) = f_k$, $k = 1, \ldots, 11$.

k	f_k	$p(f_k)$
1	0.0	.290
2	0.1	.160
3	0.2	.100
4	0.3	.080
5	0.4	.070
6	0.5	.065
7	0.6	.060
8	0.7	.055
9	0.8	.050
10	0.9	.050
11	1.0	.020

To illustrate the computation of forecast value, consider the refinement distribution $p(f_k)$ for a hypothetical set of probability forecasts for a dichotomous outcome variable, shown in Table 1. Assuming that these forecasts are perfectly calibrated (conditionally unbiased) so that $p(x_1|f_k) = f_k$, the forecasts will also be unconditionally unbiased, yielding the climatological forecast

$$f_0 = \sum_{k=1}^{11} f_k\, p(f_k) = 0.300 \quad . \tag{13}$$

Thus Equation (10) yields the expected return per decision associated with the climatological information $O^*(f_0) = -C$ for decision makers whose problems involve $C/L < 0.3$, and $O^*(f_0) = -0.300\ L$ for decision makers having $C/L < 0.3$.

The solid curve in Figure 5 shows the expected forecast value (Equation 11) for the forecasts summarized in Table 1 relative to the constant climatological forecast (Equation 13), over the full meaningful range of the Cost/Loss ratio, $0 < C/L < 1$. Value approaches zero at the extremes of this range because the best decisions are increasingly obvious there without the forecast information: as C/L approaches zero the protective action will always be taken because it costs almost nothing, consistent with the best action associated with the climatological forecast for small C/L; and as C approaches L the protective action will never be taken because this will save no money even in the event of adverse weather for large C/L, again consistent with the climatological forecast. Forecast value is maximized for $C/L = f_0 = 0.3$ because the climatological information is least informative about the best action for this decision problem.

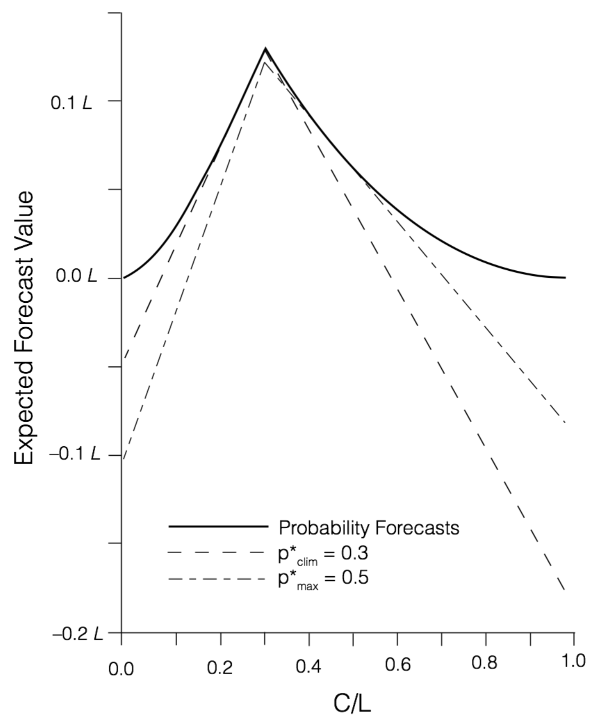

Figure 5. Expected economic value of calibrated probability forecasts from Table 1 (solid curve), and nonprobabilistic forecasts derived from them (dashed and dot-dashed lines), as functions of the Cost/Loss ratio.

Why Probability Forecasts Are More Valuable

"In any discussion of the role of weather prediction in the decision process, it must be kept clear that there is a duality of roles. The meteorologist analyzes and evaluates the present and past weather, and estimates the future state of the weather; the entrepreneur or other user of the meteorological service must be able to evaluate these predictions and analyses and translate them into the most favorable or most desirable course of action." (Epstein 1962).

Consider now the use of nonprobabilistic forecasts in the decision problem described in the previous section. Here the forecaster may only forecast "adverse weather" or "no adverse weather," without the ability to quantify the forecast uncertainty using probabilities. The forecaster will in that case issue the "adverse weather" forecast when the probability for that event is sufficiently high, i.e., when $f_k \geq p^*$, where p^* is some threshold probability. That is, to make these nonprobabilistic forecasts the forecaster must degrade the information content of his or her probabilistic judgments about the upcoming weather, and that is the ultimate source of the problem.

Ideally, p^* would be chosen to be consistent with the decision problems faced by the users of the forecasts. But forecasters rarely have a single client, or a homogeneous audience, for their forecasts. Necessarily then the choice of a p^* must be somewhat arbitrary, and whatever threshold is chosen will better serve some forecast users than others. Here two of the possible choices for p^* will be considered: $p^*_{clim} = 0.3$ yields "adverse weather" forecasts whenever in the forecaster's judgment the event probability is at least as large as its climatological probability, and $p^*_{max} = 0.5$ leads to "adverse weather" forecasts whenever this event is at least as likely as its complement.

Again using the hypothetical verification data in Table 1, 45% of forecasts will be for "adverse weather" when the forecaster uses p^*_{clim}, and this fraction will be 30% when using p^*_{max}. Decision makers following these forecasts will take protective action when warned of "adverse weather," incurring their cost C on those occasions. Similarly these decision makers will not take protective action when "no adverse weather" is forecast, and will incur their loss L with overall probability $(0.29)(0.0)+(0.16)(0.1)+(0.10)(0.2)=0.036$ when the forecaster uses p^*_{clim} and $(0.29)(0.0)+(0.16)(0.1)+(0.10)(0.2)+(0.08)(0.3)+(0.07)(0.4)=0.088$ when p^*_{max} is used.

The dashed and dash-dotted lines in Figure 5 show forecast value, again relative to optimal use of the climatological probabilities, for these nonprobabilistic forecasts based on p^*_{clim} and p^*_{max}, respectively. In both cases forecast value is never more than for the probabilistic forecasts shown by the solid line, and is equal only for $p^* = C/L$. That is, the thresholding process that produces the nonprobabilistic forecasts specializes optimal use of those forecasts only for users with C/L equal to the threshold probability. Outcomes for decision makers with Cost/Loss ratios different from p^* are accordingly suboptimal. *In effect the nonprobabilistic forecaster is making the decision, usurping the role of the decision makers without specific knowledge of any decision maker's economic situation.*

Forecast value in Figure 5 for all three of the forecast types achieves maxima for $C/L = f_0 = 0.3$ because the climatological information is least informative for this Cost/Loss ratio. The economic value for optimally used probability forecasts is everywhere nonnegative, but the economic value of the nonprobabilistic forecasts is negative for many decision problems. That is, for substantial ranges of the Cost/Loss ratio, decision makers are better served by following the naive strategy of always or never protecting, depending on whether their C/L is smaller or greater than f_0, respectively, than by using the nonprobabilistic forecasts.

The forecast value results here have been computed using the single illustrative refinement distribution in Table 1, two selected threshold probabilities, and under the assumption that the probability forecasts are perfectly calibrated. The Cost/Loss ratio situation has been used here to illustrate quantitative use of forecasts in decision problems because is simple and straightforward to analyze, but the result that optimally used probability forecasts dominate nonprobabilistic forecasts in terms of economic value is general (e.g., Alexandridis and Krzysztofowicz 1985, Krzysztofowicz 1983, Wilks et al. 1993), because degradation of the information content of probability forecasts to a single number prevents decision makers from optimizing their actions. Regardless of the accuracy or skill of nonprobabilistic forecasts, unless they are perfect (i.e., never wrong) there will be some forecast users (some range of C/L) for whom their value is negative relative to optimal use of the climatological probability (Thompson 1952). The superior value of probability forecasts is robust to the assumption of perfect calibration of the forecasts, unless the forecast quality is quite poor (Murphy 1977).

More Elaborate Decision Models

"Decision-making models have traditionally neglected the impact of potentially useful information contained in forecasts for periods beyond the initial period." (Epstein and Murphy 1988)

The idealized Cost/Loss decision framework can be a reasonable approximation to many real decision problems that are isolated in time, where decisions made and outcomes experienced on one occasion do not affect subsequent actions and outcomes. However, some decision problems are inherently sequential. For example, one example that has motivated Cost/Loss analysis is the decision to pour concrete or not, where adverse weather requires removal of a ruined pour and necessitates subsequent attempt(s) to complete the job (Thompson 1950, Roebber and Bosart 1996). In this setting there is a sequence of decisions that may continue for multiple periods, until the decision to pour concrete is followed by the occurrence of "no adverse weather." Alternatively, if the decision in question is whether to protect a frost-sensitive crop (Baquet et al. 1976, Katz et al. 1982, Thompson 1952), protective action may be contemplated on a sequence of nights, which may last until either the crop is harvested, all danger of frost has passed, or a complete loss is sustained. Such inherently sequential decision problems are known as "dynamic."

In the paper "Use and Value of Multiple-Period Forecasts in a Dynamic Model of the Cost-Loss Ratio Situation" Epstein and Murphy (1988) extended an earlier dynamic Cost/Loss decision analysis (Murphy et al. 1985) to include treatment of serially correlated weather forecasts and events. Even without consideration of serial correlation of the events and probabilities, the two-stage Cost/Loss decision situation (which is the simplest possible dynamic decision problem) is much more complicated. Figure 6, from Winkler and Murphy (1985) illustrates the structure of this problem with a decision tree, in which the square nodes represent the decisions (Protect or Not Protect), the circle nodes represent the events (adverse Weather or Not) where the adverse event occurs with probability *p*, and the terminal payoffs at the right of the figure indicate cumulative protection costs and (if applicable) loss. One branch of this tree has no decision node for Day 2 because it is assumed that the loss *L* can be incurred only once.

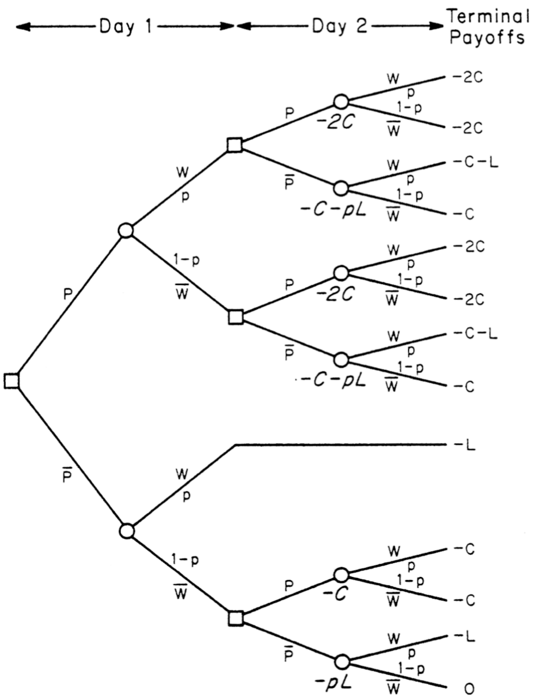

Figure 6. Decision tree for the 2-stage dynamic Cost/Loss decision problem.

The decision tree for this simplest possible dynamic decision problem is already fairly complicated, and its analysis is complicated further if the *ex ante* value of forecasts is to be computed because then refinement distributions for the forecasts must be specified, and the tree analyzed for all combinations of Day 1 and Day 2 forecasts. It is more complicated still when more than two decision periods are considered; and/or when the temporal dependence structure of the forecasts is included, necessitating the modeling of the joint distribution of the forecasts on all days in the sequence. Yet more complex analyses will also have more than two event nodes at each stage.

A seemingly odd feature of sequential decision problems is that they are solved backward, with the last decision period considered first, and the first decision period considered last. Consider the four terminal payoffs at the top of Figure 6. The expected return following the decision to Protect on Day 2 is $-2pC-2(1-p)C = -2C$, and the expected return following the no-protection action is $-p(C+L)-(1-p)C = -C-pL$, as indicated below the respective Day 2 event nodes. If the decision process has reached the decision node from which these four branches originate, the optimal action will be the one corresponding to the greater expected return, or $p>C/L$. Similar calculations are also performed for the other Day-2 decision nodes, and the larger of each of these pairs of expected returns become the "terminal" payoffs for the Day-1 decision.

Direct analysis of this kind is feasible only for very simple decision trees such as the one in Figure 6. More generally these problems are solved using a computation approach known as stochastic dynamic programming (e.g., Kennedy 1986). As before, the goal is to maximize expected return over the full sequence of decisions, but it is not necessary to draw or even imagine the decision tree. The tractability of the approach requires that the status of the decision problem at a given time can be specified by the values of a few state variables. As the solution procedure works backward in time, it is then not necessary to know the complete sequence of decision and event pairs in preceding periods (which is good, since these will yet to have been computed), but only their cumulative effects as reflected by the state variables. Problems amenable to solution in this way are accordingly sometimes called Markov decision problems. Thinking in terms of a decision tree, this structure means that there are as many decision nodes in a given decision period as there are possible combinations of state variables. For example, if a problem can be described using two state variables, each of which may take on one of ten values at each stage, then there are 100 branches on the decision tree at that decision node. In Figure 6 there is one state variable (weather state), which can take on one of two values (adverse or not).

The stochastic dynamic recursion can be expressed as

$$ER_t\{\lambda_t\} = E_f\left[\max_i \sum_{j=1}^{J}(ER_{t-1}\{O[a_i,x_j,\lambda_t]\}p[x_j \mid f_k])\right], \quad (14)$$

where λ_t is the (generally vector-valued) state variable at stage *t*, and time is counted backward so that the final decision in the sequence is for *t* = 1. The outcome function includes also the state vector as an additional argument, which specifies the particular node on the decision tree in which the process currently resides. Accordingly, the outcome function in effect points to the state in the next time period:

$$O[a_i,x_j,\lambda_t] = \lambda_{t-1} \quad . \quad (15)$$

The expected returns computed in Equation (14) pertain to the current *and all subsequent* decision periods, and the outer expectation pertains to the frequency-of-use of the possible forecasts expressed by their refinement distribution. The recursion in Equation (14) is initialized by setting the $ER_0(\lambda_0)$ equal to the terminal payoffs (e.g., outcome values on the right-hand side of Figure 6). During the calculation it is possible to save the optimal actions at each juncture for subsequent analysis or to provide guidance to decision makers.

Given a large enough number of state variables, complex and realistic decision problems can be modeled. For example, Wilks and Wolfe (1998) model irrigation decisions for a lettuce crop, over a 62-day decision period. Growth of the crop was represented using three state variables: plant dry matter, soil moisture, and a stress index. Serial correlation in the weather forecasts, and therefore also in the implicitly represented weather, was represented by including three state variables for the previous day's temperature forecast, the day-1 precipitation probability, the and day-2 precipitation probability. Accordingly the probabilities in the refinement distribution, with respect to which the outer expectation $E_f[\bullet]$ in Equation (14) is taken, also depend on the state vector.

Figure 7 shows thresholds for the day-1 PoP forecast below which an irrigation is optimal, as a function of date in the growing season (horizontal) and the available soil moisture status (vertical), when today's temperature forecast is 17.5˚C, and the day-2 PoP forecast is 0.0 (top panel), 0.4 (middle panel), and 0.8 (bottom panel). These threshold probabilities increase as the precipitation probability for day 2 increases. They are also smaller (meaning that irrigation is more sensitively needed) very early in the season when the plants are small and the absolute amount of soil moisture is small, and late in the season when good market size must be achieved within the growing period.

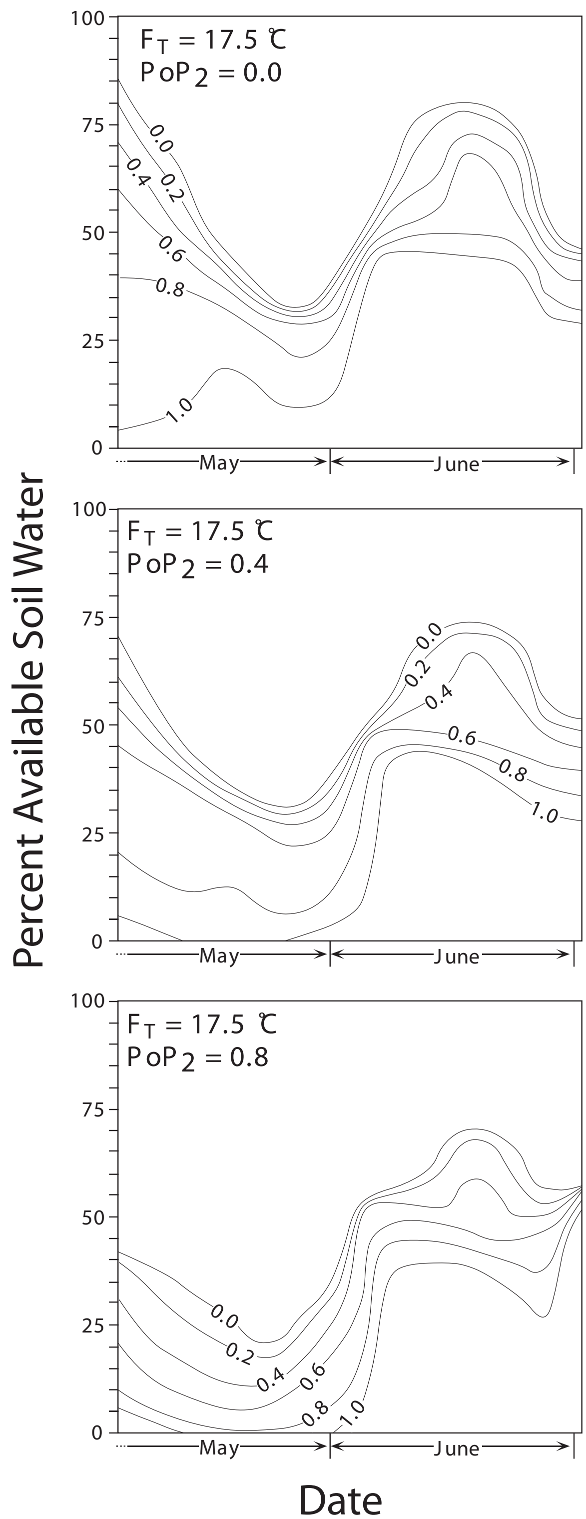

Figure 7. Probability thresholds triggering lettuce irrigation, as functions of day of the year (horizontal), and soil moisture (vertical) when the temperature forecast is 17.5˚C, and the PoP forecast for the following day is 0.0 (upper panel), 0.4 (middle panel), and 0.8 (lower panel).

Some Connections with Forecast Verification

"As a final note . . . I would introduce the general problem of forecast verification. Would not an optimum forecasting (or analysis) procedure provide maximum discrimination among, not necessarily the types of weather which may occur, but the utilities of the decisions from which one must choose?" (Epstein 1962)

In this paper I have reviewed some of Ed Epstein's contributions in probability forecasting and in optimal use and valuation of forecasts. The relationship of these topics to some of his work in forecast verification also merits some mention.

The quote at the beginning of this section from Epstein (1962) anticipates construction of a verification score using economic value as the specific metric. Murphy and Epstein (1967a) emphasized that one central purpose of forecast verification is economic — pertaining to the value of information — and that the appropriate metric is decision-theoretic utility. Murphy (1977, his Equation 11) appears to have been the first to define such a measure, although the approach seems not to have been used until rediscovered nearly a quarter-century later (Richardson 2000, Wilks 2001).

Another connection with Epstein's work in forecast verification is that his original derivation of the widely used Ranked Probability Score (RPS, Epstein 1969b) was in terms of decision theory, beginning with extension of the 2 x 2 Cost/Loss outcome matrix of Figure 4 to the more general K x K setting. The definition of RPS in terms of expected utility assured that the score is strictly proper, so that forecasters can receive best expected score only by forecasting their true beliefs (Murphy and Epstein 1967b).

Finally, availability of comprehensive (diagnostic) verification data is essential in order for users to derive full benefit from forecasts, by knowing that probability forecasts are well calibrated, or failing that having sufficient information to compute calibrated Bayesian posterior probabilities. Comprehensive forecast verification data is also necessary in order to compute expected economic value of forecasts *ex ante*. Similarly, having a good estimate of the joint distribution of forecasts and observations is a necessary starting point for modeling possible improved future forecasts (e.g., sharper refinement distributions for calibrated forecasts), leading to the ability to compute expected economic value to be derived from investments in those improvements.

Concluding Remarks

Uncertainty is inherent in the chaotic dynamics of the atmosphere, and probability forecasting is necessary to communicate the resulting uncertainty in its future behavior. Probability forecasts are essential for optimal, rational decision making, for realizing full potential economic value of forecasts, and to allow separation of the roles of forecaster and decision maker. Ed Epstein was well ahead of his time in understanding all of these things, and he made important contributions to their being understood and appreciated more widely.

References

Alexandridis, M.G., and R. Krzysztofowicz, 1985: Decision models for categorical and probabilistic forecasts. *Applied Mathematics and Computation*, **17**, 241-266.

Angstrom, A.K., 1922: On the effectivity of weather warnings. *Nord. Statistisk Tidskrift,* **1,** 394–408.

Baquet, A.E., A.N. Halter, and F.S. Conklin, 1976: The value of frost forecasting: a Bayesian appraisal. *American Journal of Agricultural Economics*, **58**, 511-520.

Berger, J.O., 1980: *Statistical Decision Theory*. Springer-Verlag, 425 pp.

Bilham, E.G., 1922: A problem in economics. *Nature*, **109**, 341-342.

Clemen, R.T., 1991: *Making Hard Decisions*. Duxbury, 557 pp.

Curtiss, J.H., 1968: An elementary model for the interpretation of precipitation probability forecasts. *Journal of Applied Meteorology*, **7**, 3-17.

Gringorten, I.I., 1950: Forecasting by statistical inferences. *Journal of Meteorology*, **7**, 388-394.

Epstein, E.S., 1962. A Bayesian approach to decision making in applied meteorology. *Journal of Applied Meteorology*, **1**, 169-177.

Epstein, E. S., 1966a: Point and area precipitation probabilities. *Monthly Weather Review*, **94**, 595-598.

Epstein, E. S., 1966b: Quality control of probability forecasts. *Monthly Weather Re*view, **94**, 487-494.

Epstein, E. S., 1968: On the correspondence between theory and practice in probability forecasts. *Proceedings of the First Statistical Meteorological Conference*, Hartford,, American Meteorological Society, 142-147.

Epstein, E.S., 1969a: Stochastic dynamic prediction. *Tellus*, **21**, 739-759.

Epstein, E.S., 1969b: A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology*, **8**, 985-987.

Epstein, E.S., 1985: *Statistical Inference and Prediction in Climatology: A Bayesian Approach*. American Meteorological Society, Meteorological Monographs No. 42, 199 pp.

Epstein, E.S., and A.H. Murphy, 1988: Use and value of multiple-period forecasts in a dynamic model of the cost-loss ratio situation. *Monthly Weather Review*, **116**, 746-761.

Gringorten, I.I., 1950: Forecasting by statistical inferences. *Journal of Meteorology*, **7**, 388-394.

Hallenbeck, C., 1920: Forecasting precipitation in percentages of probability. *Monthly Weather Review,* **48,** 645–647.

Katz, R.W., A.H. Murphy, and R.L. Winkler, 1982: Assessing the value of frost forecasts to orchardists a dynamic decision-making approach. *Journal of Applied Meteorology*, **21**, 518-531.

Kennedy, J.O.S., 1986: *Dynamic Programming, Applications to Agriculture and Natural Resources*. Elsevier, London, 341 pp. .

Krzysztofowicz, R., 1983: Why should a forecaster and a decision maker use Bayes Theorem. *Water Resources Research*, **19**, 327-336.

Murphy, A.H., 1977: The value of climatological, categorical, and probabilistic forecasts in the cost-loss ratio situation. *Monthly Weather Review*, **105**, 803-816.

Murphy, A.H., 1985: Probabilistic weather forecasting. In: A.H. Murphy and R.W. Katz, eds., *Probability, Statistics, and Decision Making in the Atmospheric Sciences*. Westview, Boulder, 337-377.

Murphy, A.H., 1994: Assessing the economic value of weather forecasts: an overview of methods, results and issues. *Meteorological Applications*, **1**, 69-73.

Murphy, A.H., 1998: The early history of probability forecasts: some extensions and clarifications. *Weather and Forecasting*, **13**, 5-15.

Murphy, A.H., B.G. Brown, and Y.-S. Chen, 1989: Diagnostic verification of temperature forecasts. *Weather and Forecasting*, **4**, 485-501.

Murphy, A.H., and E.S. Epstein, 1967a: Verification of probabilistic predictions: a brief review. *Journal of Applied Meteorology*, **6**, 748-755.

Murphy, A.H., and E.S. Epstein, 1967b: A note on probability forecasts and "hedging." *Journal of Applied Meteorology,* **6**, 1002-1004.

Murphy, A.H., R.W. Katz, R.L. Winkler, and W.-R. Hsu, 1985: Repetitive decision making and the value of forecasts in the cost-loss ratio situation: a dynamic model. *Monthly Weather Review*, **113**, 801-813.

Murphy, A.H., S. Lichtenstein, B. Fischoff, and R.L. Winkler, 1980: Misinterpretations of precipitation probability forecasts. *Bulletin of the American Meteorological Society*, **61**, 695-701.

Murphy, A.H., and R.L. Winkler, 1971a: Forecasters and probability forecasts: the responses to a questionnaire. *Bulletin of the American Meteorological Society*, **52**, 158-165.

Murphy, A.H., and R.L. Winkler, 1971b: Forecasters and probability forecasts: some current problems. *Bulletin of the American Meteorological Society*, **52**, 239-247.

Murphy, A.H., and R.L. Winkler, 1974: Probability forecasts: a survey of National Weather Service forecasters. *Bulletin of the American Meteorological Society*, **55**, 1449-1452.

Murphy, A.H., and R.L. Winkler, 1984: Probability forecasting in meteorology. *Journal of the American Statistical Association*, **79**, 489-500.

Murphy, A.H., and R.L. Winkler, 1987: A general framework for forecast verification. *Monthly Weather Review*, **115**, 1330-1338.

Murphy, A.H., and R.L. Winkler, 1992: Diagnostic verification of probability forecasts. *International Journal of Forecasting*, **7**, 435-455.

Richardson, D. S., 2000: Skill and economic value of the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society,* **126**, 649–667.

Roebber, P.J., and L.F. Bosart, 1996: The complex relationship between forecast skill and forecast value: a real-world analysis. *Weather and Forecasting*, **11**, 544-559.

Rogell, R.H., 1972: Weather terminology and the general public. *Weatherwise*, **25**, 126-132.

Scoggins, J.R., and W.W. Vaughan, 1971: How some nometeorological professionals view meteorology and weather forecasting. *Bulletin of the American Meteorological Society*, **52**, 974-979.

Root, H.E., 1962: Probability statements in weather forecasting. *Journal of Applied Meteorology*, **1**, 163-168.

Thompson, J.C., 1950: A numerical method for forecasting rainfall in the Los Angeles area. *Monthly Weather Review*, **78**, 113-124.

Thompson, J.C., 1952: On the operational deficiencies in categorical weather forecasts. *Bulletin of the American Meteorological Society*, **33**, 223-226.

Wilks, D.S., 2001: A skill score based on economic value for probability forecasts. *Meteorological Applications,* 8, 209-219.

Wilks, D.S., R.E. Pitt, and G.W. Fick, 1993: Modeling alfalfa harvest scheduling using short-range weather forecasts. *Agricultural Systems*, **42**, 277-305.

Wilks, D.S., and D.W. Wolfe, 1998: Optimal use and economic value of weather forecasts for lettuce irrigation in a humid climate. *Agricultural and Forest Meteorology*, **89**, 115-129.

Williams, C., E.A. Stewart, A. Horst, E. Knox, B. Brough, and J.A. Knox, 2014: Through the eyes of the experts: the perception of the probability of precipitation. 22nd Conference in Probability and Statistics in the Atmospheric Sciences, Atlanta GA, Paper TJ3.4.

Winkler, R.L., 1972: *Introduction to Bayesian Inference and Decision*. Holt, Rinehart and Winston, 563 pp.

Winkler, R.L., and A.H. Murphy, 1985: Decision Analysis. In: A.H. Murphy and R.W. Katz, eds., *Probability, Statistics, and Decision Making in the Atmospheric Sciences*. Westview, Boulder, 493-524.