Probability Forecasting in Meteorology

ALLAN H. MURPHY and ROBERT L. WINKLER*

Efforts to quantify the uncertainty in weather forecasts began more than 75 years ago, and many studies and experiments involving objective and subjective probability forecasting have been conducted in meteorology in the intervening period. Moreover, the U.S. National Weather Service (NWS) initiated a nationwide program in 1965 in which precipitation probability forecasts were formulated on an operational basis and routinely disseminated to the general public. In addition, the NWS now prepares objective probability forecasts for many variables, using statistical procedures. Hence probability forecasting in meteorology is unique in that very large sets of probability forecasts that have been subjected to detailed evaluation are available. This article has four objectives: (a) to review the history of probability forecasting in meteorology to acquaint statisticians with this body of literature; (b) to describe recent methodological, experimental, and operational activities in this field; (c) to examine current issues in probability forecasting; and (d) to discuss briefly the relationship between probability forecasting in meteorology and probability forecasting in other fields. Results of operational and experimental weather forecasting programs are presented, methods of verifying and evaluating probability forecasts in common use in meteorology are discussed, and an extensive list of references is provided.

KEY WORDS: Probability forecasts; Weather forecasting; Statistical forecasting; Objective and subjective forecasts; Subjective probability assessment; Forecast verification/evaluation; Scoring rules; Use and value of probability forecasts.

1. INTRODUCTION

Forecasts generally consist of point or interval estimates of the future values of the relevant variables. The uncertainty inherent in such forecasts, however, is seldom expressed in a quantitative manner. Categorical or deterministic forecasts suffer from two serious deficiencies. First, a forecaster or forecast system is seldom, if ever, certain which event will occur. Second, categorical forecasts do not provide users of the forecasts with the information that they need to make rational decisions in uncertain situations.

It is convenient to distinguish between objective and subjective methods of forecasting, both of which are used extensively in meteorology. So-called objective methods are objective in the sense that for a particular procedure and set of relevant data, the specific forecasts produced do not depend on a forecaster's judgment, although subjectivity *is* involved in the choice of a procedure and a set of data. Subjective methods, on the other hand, are methods in which the formulation of the forecasts is based at least in part on the judgments of one or more forecasters.

Meteorologists have long recognized the existence of uncertainty in their forecasts, and efforts to quantify this uncertainty began more than 75 years ago. Many studies and experiments involving objective and subjective probability forecasting have been conducted in meteorology. Moreover, the U.S. Weather Bureau—since 1970, the National Weather Service (NWS)—in 1965 initiated a nationwide program in which subjective precipitation probability forecasts were formulated on an operational basis and routinely disseminated to the general public. In addition to this program, the NWS now prepares objective probability forecasts for many variables, using statistical and numerical-statistical procedures. These forecasts are provided to the forecasters as guidance and are available to individuals who subscribe to the NWS communications system. Hence probability forecasting in meteorology is unique in that very large sets of probability forecasts that have been subjected to detailed evaluation are available.

The objectives of this article are fourfold. First, we want to review the history of probability forecasting in meteorology to familiarize statisticians with this body of literature. This review is presented in Section 2. Second, in Section 3 we describe recent methodological, experimental, and operational activities. Third, we examine some current issues in probability forecasting in Section 4, emphasizing issues of particular interest in meteorology but also including issues of more general interest. Fourth, we discuss briefly in Section 5 the relationship between probability forecasting in meteorology and probability forecasting in other fields.

2. HISTORICAL DEVELOPMENTS

2.1 Early Work (1900–1925)

The term *probability* has been used informally to describe anticipated meteorological conditions since the beginnings of operational forecasting more than a century

* Allan H. Murphy is Professor and Director of the Statistics and Climatic Impacts Laboratory, Department of Atmospheric Sciences, Oregon State University, Corvallis, OR 97331. Robert L. Winkler is IBM Research Professor, The Fuqua School of Business, Duke University, Durham, NC 27706. The research was supported in part by the National Science Foundation, Division of Atmospheric Sciences, Grants ATM-8004680 and ATM-8209713.

© Journal of the American Statistical Association
September 1984, Volume 79, Number 387
Applications Section

ago. However, the need to *quantify* the uncertainty inherent in weather forecasts was first recognized by Cooke (1906a,b) who advocated that each meteorological prediction be accompanied by a single number that would "indicate, approximately, the weight or degree of probability which the forecaster himself attaches to that particular prediction" (1906b, p. 274). Numerical weights were assigned to 2,036 statements in public forecasts for Western Australia in 1905, and the results are presented in Table 1. In defense of his procedure, which was under attack for being impractical and confusing, Cooke (1906b) made a statement that was very perceptive for its time:

> Now this is the point I wish to make clear. Those forecasts which were marked "doubtful" were the *best I could frame* under the circumstances. I could see no way of improving them at the time, and they would not have been expressed differently whether I weighted them or not. If I make no distinction between these and others, I degrade the whole. (p. 275)

Von Myrbach (1913), apparently without knowledge of Cooke's work, advocated that meteorologists express the "approximate degree of probability of success" of forecasts to increase their reliability. His proposal was adopted by the Signal Corps Meteorological Service of the American Expeditionary Forces in France in World War I (U.S. Army Signal Corps 1919).

In response to requests by local farmers for "the chances of rain," Hallenbeck (1920) developed a subjective procedure for forecasting precipitation in the Pecos Valley in New Mexico in terms of "percentages of probability." The procedure was based on a series of composite weather maps. Some results of this program are presented in Table 2.

During this early period, several meteorologists formulated objective procedures that were capable of producing probabilistic forecasts (e.g., Besson 1905, Rolf 1917, and Blair 1921,1924). These techniques were based on simple contingency tables and scatter diagrams and generally involved forecasting the occurrence or nonoccurrence of an event with at most two predictors. Unfortunately the resulting probability forecasts were almost always immediately translated into categorical forecasts by selecting the event with the largest probability.

Table 1. Weighted Forecasts for Western Australia

Weight*	No. of Forecasts	No. of Correct Forecasts	% Correct
5	685	675	98.5
4	970	910	93.8
3	296	233	78.7
2	66	37	56.1
1	19	11	57.9
Total/average	2,036	1,866	91.7

* Cooke (1906a) gave the weights the following definitions: 5 = almost absolute certainty; 4 = tolerable certainty (wrong about 1 in 10 times); 3 = very doubtful (probably wrong about 4 times out of 10); 2 = just possible, but not likely; and 1 = the barest possibility (not at all likely).
NOTE: Data from Cooke (1906a).

Table 2. Precipitation Probability Forecasts for the Pecos Valley in New Mexico

Forcast Probability (%)	No. of Forecasts	No. of Precipitation Occurrences	Relative Frequency of Precipitation (%)
0–9	0	0	—
10–19	22	4	18.2
20–29	31	7	22.6
30–39	18	6	33.3
40–49	15	8	53.3
50–59	13	8	61.5
60–69	15	11	73.3
70–79	7	6	85.7
80–89	1	1	100.0
90–100	1	1	100.0
Total/average	123	52	42.3

NOTE: Data from Hallenbeck (1920).

2.2 Period of Renewed Interest (1940–1955)

Interest among meteorologists in probability forecasting lay dormant for several years but increased significantly in the 1940–1955 period, as they began to investigate a variety of problems related to the formulation, evaluation, and use of such forecasts. Although the experiments in probability forecasting conducted during this period were generally small, they yielded encouraging results concerning the ability of forecasters and objective forecasting systems to quantify the uncertainty in forecasts of a variety of weather variables.

In a key paper, Brier (1944) discussed the need for and use of probabilistic forecasts, presented the results of several small probability forecasting experiments, and considered some aspects of the problem of evaluating probabilistic forecasts. Williams (1951) described an experiment in which Weather Bureau forecasters assigned confidence factors of 6, 8, or 10 to their regular 12-hour forecasts of precipitation or no precipitation. These factors were intended to indicate a 60%, 80%, or 100% chance that the (categorical) forecast would actually verify. Williams also briefly discussed an experiment in which a forecaster assigned similar confidence factors to forecasts of maximum and minimum temperatures.

An experiment in which probabilities were assigned to five temperature categories (much below normal, below normal, near normal, above normal, and much above normal) in five-day, extended-range forecasting was described by Leight (1953). In addition, Schroeder (1954) discussed an experiment in which fire-weather forecasters used terms such as "chance," "possible," and "probable" in connection with their categorical forecasts of precipitation occurrence. These terms were assigned specific definitions in terms of a range of probability values.

Renewed interest in objective probability forecasting resulted primarily from a paper by Brier (1946), in which he presented a general framework for a scatter diagram,

or graphical regression, approach. This framework enabled the meteorologist to consider a relatively large number of predictor variables and to incorporate these variables into the forecasting procedure in a systematic manner. Brier himself used this technique to make probabilistic forecasts of precipitation amount for the TVA Basin, and it was used successfully by others to formulate probabilistic forecasts for a variety of variables in various geographic locations. For a review of objective probabilistic forecasting at this time, see Gringorten (1955).

In the early probability forecasting studies, forecast evaluation consisted primarily of comparing the forecast probabilities and the observed relative frequencies. Moreover, objective probabilistic forecasts were generally translated into categorical forecasts, particularly when it came to evaluation. A notable exception is the paper by Thompson (1950), in which he compared objective forecasts with experimental subjective forecasts by using the first quantitative measure designed specifically to evaluate probabilistic forecasts, namely, the so-called Brier score formulated by Brier (1950). This measure is simply the mean squared error of the forecasts. It is of interest to note that Brier stated that this measure "cannot influence the forecaster in any undesirable way" (p. 1). Although Brier offered no formal proof of this assertion, the Brier score *is* the first known example of a strictly proper scoring rule (i.e., a scoring rule that discourages hedging on the part of the forecaster). For reviews of forecast evaluation at this time, refer to Brier and Allen (1951) and Gringorten (1951).

Early studies of the value and use of different types of forecasts involved simple decision-making situations (e.g., Angstrom 1922 and Bilham 1922). Thompson (1952) determined the ex post economic value of conventional categorical forecasts within the framework of a two-action (protect/do not protect), two-state (adverse weather/no adverse weather) situation. In this so-called cost–loss ratio situation, he found that the value of forecasts based solely on the climatological probability exceeds the value of categorical forecasts for significant ranges of values of the cost–loss ratio. Subsequently Thompson and Brier (1955) described more general procedures for assessing the economic value of categorical and climatological forecasts in this situation, and they examined the relationships between these measures of value and conventional measures of the accuracy of (categorical) forecasts.

2.3 Developmental/Experimental Period (1955–1965)

In the mid-1950's, several subjective probability forecasting programs more extensive than previous activities were initiated. Travelers Weather Service (TWS) began to issue precipitation probability forecasts to the general public in Hartford, Connecticut, on a regular basis in 1954. These forecasts were prepared subjectively by TWS forecasters. To the best of our knowledge, this program represented the first time that probability statements were included in public weather forecasts.

At about the same time, students and staff at the Massachusetts Institute of Technology (MIT) began to make probability forecasts of several variables, including temperature and precipitation, in conjunction with their synoptic laboratory exercises. Some early results of this ongoing activity are described by Sanders (1958,1963), who concluded that "forecasters are capable of skillful statements about the probability of occurrence of a wide variety of meteorological events over ranges up to seventy-two hours" (Sanders 1963, p. 200).

In 1956 the Weather Bureau Forecast Center in San Francisco added precipitation probabilities to their public weather forecasts (Root 1962). These probability forecasts were subjective in nature. The Forecast Center in Los Angeles adopted this practice the following year.

These experimental and limited operational programs, together with advances in computer technology and increased interest in applying statistical techniques to make weather forecasts, stimulated further methodological work in objective probability forecasting. Several statistical procedures were used to generate experimental probability forecasts at the Travelers Research Center (TRC) beginning in 1959. For example, a nonparametric approach to discriminant analysis developed by Fix and Hodges (1951) was used to estimate probabilities for multiple categories in ceiling height and visibility forecasting (Miller 1962). Moreover, drawing on earlier work by Lund (1955), Miller (1964) developed a technique called REEP (regression estimation of event probabilities) in which the predictand is treated as a set of two or more Boolean variables. The objective probability forecasts produced in these studies were comparable in quality to forecasts made by Weather Bureau forecasters. In addition to the work at TRC, Glahn (1964a,b) investigated the use of contingency tables and adaptive logic as procedures to produce probability forecasts of ceiling height. A useful review of objective statistical weather forecasting at this time was provided by Glahn (1965).

With an increasing number of probability forecasts came a greater interest in methods of evaluating such forecasts. Sanders (1958,1963) partitioned the Brier score into two terms—one measuring reliability (the correspondence between forecast probabilities and observed relative frequencies) and one measuring resolution (the ability of a forecaster (or forecasting method) to separate, in advance, the set of situations for which the event of interest (e.g., precipitation) will occur and the set of situations for which it will not occur). Other evaluation measures for probability forecasts formulated in a meteorological context during this period included the information ratio (Holloway and Woodbury 1955), the validity measure (Miller 1962), and the matrix skill score (Gringorten 1965), but the use of these measures has been limited.

Meteorologists (and others) explored the value and use of probabilistic forecasts, and weather forecasts in general, in a variety of decision-making situations during this period. For example, Thompson (1962) studied the ex

post value of probability forecasts within the context of the cost–loss ratio situation. The use of probability forecasts within the framework of a Bayesian approach to decision making was investigated by Epstein (1962). Other studies concerned with the value and/or use of such forecasts in general or in specific decision-making situations included the work of Gringorten (1959), Borgman (1960), Lave (1963), Glahn (1964a), and Nelson and Winter (1964).

2.4 Initiation of Nationwide Program in 1965 (1965–1970)

After almost 10 years of experience with operational precipitation probability forecasting at San Francisco and Los Angeles and some experience at other locations (e.g., see Hughes 1965), the Weather Bureau initiated a nationwide program in 1965 in which precipitation probabilities were appended to all public weather forecasts on a regular basis. This program represented the first time that probabilities in any field of application were issued on such a large scale, and the program continues, largely unchanged, to this day.

The forecast, which represents the probability of measurable precipitation (≥.01 inches) at a specific point (the official rain gauge) during a given period, is called a PoP (probability of precipitation) forecast. The PoP forecasts disseminated to the public are prepared by Weather Bureau forecasters at local forecast offices. Beginning in 1968, the National Meteorological Center (NMC) of the Weather Bureau provided forecasters with PoP guidance forecasts. Initially these centrally-prepared forecasts were formulated subjectively by NMC meteorologists based primarily on information related to the large-scale features of the atmosphere. The forecasters considered the guidance forecasts, together with a variety of other types of data (e.g., analyses and prognostic charts, satellite and radar data, and recent local surface observations), in arriving at their official PoP forecasts.

In terms of methodological developments, experimental work concerning applications of regression and discriminant analysis continued during this period. Another method introduced at this time and subsequently used to generate probability forecasts of precipitation type was the logit model (Brelsford and Jones 1967 and Jones 1968). It was also during this period (1965–1970) that efforts were first made to use the output of numerical (i.e., dynamical) models of the atmosphere in combination with statistical techniques to generate probability forecasts (e.g., Glahn and Lowry 1969).

The Brier score was used by the Weather Bureau to evaluate the PoP forecasts in its nationwide program. In addition, a skill score, which measures the percentage improvement in the Brier score for the PoP forecasts over the Brier score for climatological forecasts, was computed. The skill scores indicated that the forecasters' performance was clearly better than climatology, with the amount of improvement over climatology decreasing as a function of lead time (e.g., Hughes 1968 and Derouin and Cobb 1970). Winkler and Murphy (1968a) obtained similar results for probability forecasts formulated by TWS forecasters, as did Sanders (1967) for the MIT forecasting program.

It is interesting to note that several contributions concerning strictly proper scoring rules appeared in the meteorological literature during this period. Two general papers on this topic are Murphy and Epstein (1967) and Winkler and Murphy (1968b). In addition, Murphy (1966) gave the Brier score an economic interpretation, and Epstein (1969b) generalized this economic model to develop the ranked probability score.

3. RECENT ACTIVITIES

3.1 Objective Probability Forecasting

In the classical approach to objective weather forecasting, a prediction equation relating y_t (the value of the predictand at some future time t) directly to x_0 (the values of the predictor variables at the initial time) is derived from past observational data and then applied to the forecasting situation at hand. The classical method is most useful for short-range forecasts (0–12 hours). For example, short-range probability forecasts of severe local storms (i.e., tornadoes or severe thunderstorms) are of considerable importance, and Charba (1979) developed a system based on the combined use of the classical method and the model output statistics (MOS) technique to generate such forecasts. Recently Miller (1981) used a classical approach involving regression analysis to obtain probability forecasts for a large number of variables one hour in advance.

For forecasting one or more days in advance, meteorologists believe that numerical prediction models are needed for effective weather forecasting. Thus objective weather forecasts generally are obtained by using statistical techniques (usually regression analysis) in *combination* with numerical models of the atmosphere. This approach, which is often called numerical-statistical forecasting (e.g., Glahn 1984), has been used to generate probability forecasts as well as categorical forecasts.

Meteorologists distinguish between two approaches to numerical-statistical weather forecasting: the perfect prog procedure and the MOS technique. In the perfect prog procedure, the prediction equation based on past data relates y_t to x_t (the set of values of the predictors at time t). In application, however, x_t is not known when the forecast is being made. Numerical models are used to predict x_t, given x_0 (thus the potential predictors must be variables predicted by the numerical models). The prediction, $\hat{x}_t$, is then treated as a perfect prediction (hence the name perfect prog) and is used in place of x_t in the statistical model that generates a forecast $\hat{y}_t$. Thus the perfect prog technique uses numerical models but ignores the uncertainty in the predictions generated by these models. As the lead time increases, this uncertainty becomes greater, thereby limiting the range of applicability of this procedure. The perfect prog technique was used

extensively in the 1960's and early 1970's to produce categorical forecasts, and Klein (1971) employed this technique to generate precipitation probability forecasts on an experimental basis.

The numerical-statistical technique that has received the greatest attention in the past decade is the MOS technique (Glahn and Lowry 1969,1972 and Klein and Glahn 1974). In the MOS approach, the prediction equation based on past data relates y_t to $\hat{x}_t$, with the latter taken from a numerical model. MOS allows for the uncertainty in the predictions generated by a numerical model, but it requires a substantial number of forecasts from the model to derive the prediction equation (thus numerical model forecasts must be archived). Moreover, since the prediction equations developed in the MOS technique are associated with particular numerical models, these equations may need to be redeveloped if the models undergo major changes. Fortunately the evidence available to date indicates that the decrease in skill is minimal when a MOS equation is used in conjunction with a model other than that on which it was developed (Glahn 1984).

MOS has been used to generate probability forecasts for a variety of different events and variables (Glahn 1984 and Murphy 1984). Skill scores are presented for four different samples of PoP forecasts in Table 3. The Glahn et al. (1971) data include over 5,200 forecasts for each period from 86 stations for October–December 1970; the Klein and Glahn (1974) data consist of PoP forecasts from 86 stations for January–May 1972; the Glahn (1976) data include a sample for the 1974–1975 cool season and the 1975 warm season; and the Schwartz et al. (1981) and Carter et al. (1982) data contain approximately 12,000 forecasts for each period from 89 stations for the 1980–1981 cool season and the 1981 warm season. In all cases, MOS improves upon climatology, with the percentage improvement decreasing as the forecast lead time increases. Note that although perfect prog also improves upon climatology, the improvement is much less than that of MOS. A comparison of the skill scores for centralized guidance precipitation probability forecasts and local subjective PoP forecasts for 50 forecast offices from 1966 to 1978 is presented in Figure 1. As indicated previously, the guidance forecasts were prepared subjectively through 1971 and objectively by the MOS technique thereafter.

With regard to other variables, Bermowitz and Zurndorfer (1979) discuss MOS probability forecasts of precipitation amount. For conditional probability forecasts of frozen precipitation (snow or sleet) given the occur-

Table 3. Skill Scores (%) for Objective and Subjective PoP Forecasts

Forecast Period	MOS	Perfect Prog	Subjective
Glahn et al. (1971)			
First	30.7	18.6	41.3
Second	30.9	13.3	23.4
Third	10.5	8.9	13.7
Klein and Glahn (1974)			
First	30.4	19.7	39.3
Second	24.1	14.4	23.8
Third	15.0	9.1	15.3
Fourth	6.4	.3	—
Glahn (1976)			
Cool season			
First	32.0		39.8
Second	25.4		28.1
Third	18.4		18.1
Warm season			
First	16.5		21.3
Second	12.1		14.4
Third	8.2		8.8
Schwartz et al. (1981) and Carter et al. (1982)			
Cool season			
First	46.8		49.5
Second	34.2		33.4
Third	33.2		34.2
Warm season			
First	27.7		30.6
Second	21.0		22.0
Third	16.2		16.2

NOTE: MOS = model output statistics. PoP = probability of precipitation.

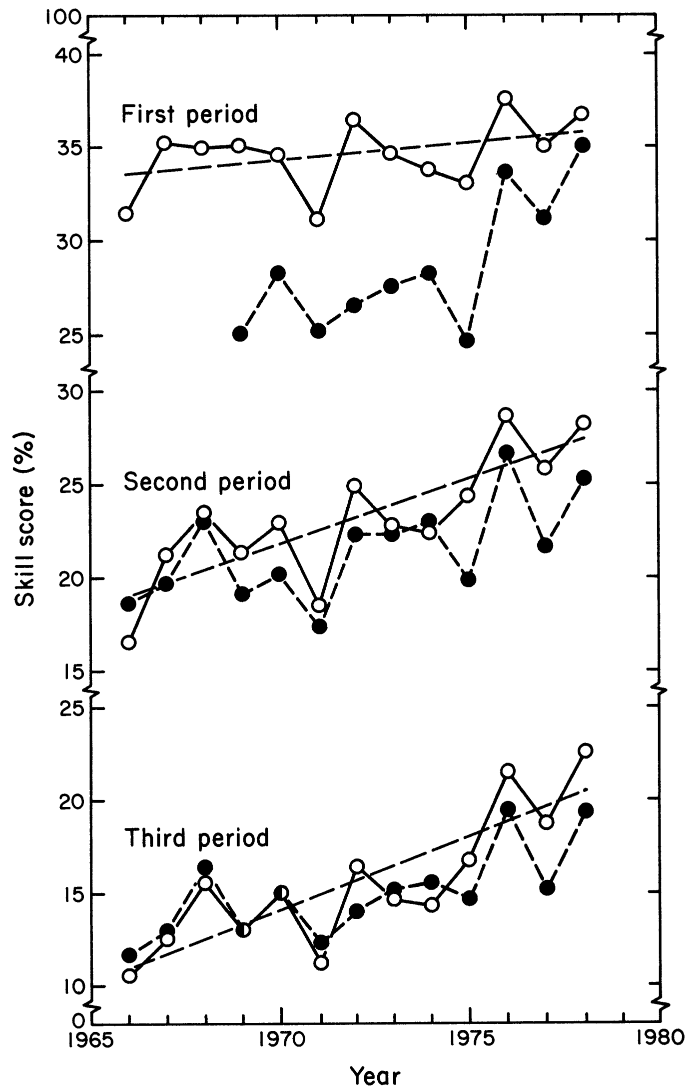

Figure 1. Skill of Centralized Guidance (– –) and Local Subjective (—) Precipitation Probability Forecasts for the Period 1966–1978 (Charba and Klein 1980).

rence of precipitation, the MOS approach has been used in conjunction with a logit model, which fits an S-shaped curve to a dichotomous predictand as a function of continuous predictors (Bocchieri 1979). Short-range (2–6 hours) and medium-range (6–24 hours) forecasts of the probability of thunderstorms and the conditional probability of severe storms given the occurrence of thunderstorms were described by Charba (1979) and Reap and Foster (1979), respectively. Cloud amount, ceiling height, and visibility have also been forecast probabilistically with the MOS approach, using regression analysis (Carter and Glahn 1976 and Bocchieri et al. 1974).

In recognition of the desirability of incorporating all uncertainty into the numerical forecasting model itself, an approach called stochastic-dynamic prediction (Epstein 1969a and Fleming 1971a,b) was developed. This approach is inherently probabilistic, and it involves equations for the evolution of second- and higher-order moments of the probability distributions of the variables of interest as well as of the first moment or mean. In the studies undertaken to date, this approach has been limited to equations for the first and second moments of these distributions for relatively simple dynamical models. Moreover, the large number of degrees of freedom involved in using the stochastic-dynamic approach in conjunction with realistic numerical weather prediction models has precluded its application in practice. Leith (1978,1980) proposed an alternative approach to stochastic-dynamic prediction based on the use of Monte Carlo methods. This approach, which involves repeated application of the first moment equation (starting from different initial conditions), substantially reduces the computational burden but still provides an estimate of the uncertainty inherent in the forecasts. For the forseeable future, however, stochastic-dynamic prediction is not likely to represent a practical method of producing probability forecasts.

3.2 Subjective Probability Forecasting

Historically most weather forecasts have been subjective in nature, whether the forecasts were expressed in probabilistic or categorical terms. Analyses of subjective PoP forecasts have indicated that the forecasters who formulate them generally perform very well. First, these forecasts are reliable in the sense that the relative frequency of occurrence of precipitation on days with a probability forecast of p tends to be very close to p. For example, a reliability diagram for 154,799 PoP forecasts formulated at 87 NWS offices during the two-year period from April 1977 through March 1979 is presented in Figure 2.

Of course reliability is only one aspect of probability forecasts. Weather forecasters are also concerned with Brier scores (i.e., measures of accuracy) and comparisons, in terms of Brier scores, with climatology and with objective forecasting schemes such as MOS (e.g., measures of skill). Analyses of precipitation probability forecasts, as illustrated in Table 3 and Figure 1, indicate that

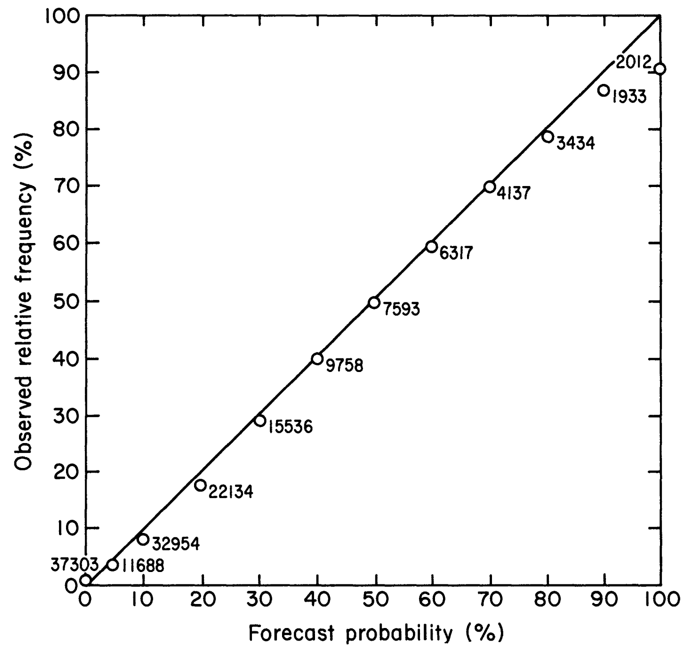

Figure 2. Reliability Diagram for Subjective Precipitation Probability Forecasts Formulated at 87 National Weather Service Offices During the Period April 1977–March 1979 (Charba and Klein 1980). The number adjacent to each point indicates the number of forecasts with that forecast probability.

the improvement in Brier scores for subjective forecasts over MOS forecasts is most pronounced for short-range (i.e., first-period) forecasts, although even this gap has narrowed in recent years.

The amount of information available to a weather forecaster has increased greatly over the years, to the point where a forecaster is virtually inundated with global and regional analyses and prognostic charts, objective weather forecasts, and satellite and radar data. Figure 1 illustrates that both objective and subjective PoP forecasts have improved with time. It should be noted that the MOS forecasts are available to the forecasters as guidance in the preparation of their subjective PoP forecasts and that the forecasters have access to data that are not available at the time that the objective forecasts are prepared.

Experiments were conducted to study the ability of forecasters to differentiate among different points in a forecast area in terms of precipitation probabilities and to make other types of probability forecasts related to precipitation (Winkler and Murphy 1976 and Murphy and Winkler 1977). The forecasters were able to differentiate with moderate success among different points in the forecast area, and the different types of forecasts exhibited a high degree of consistency. In other experiments (Sanders 1979 and Murphy et al. 1982), forecasters provided subjective probability distributions for precipitation amount instead of probabilities of precipitation occurrence. Although some overforecasting was observed, the forecasts showed modest skill for short-term forecasts and for moderate precipitation amounts.

Another important variable in weather forecasts is tem-

perature, and several experiments concerning subjective probability forecasts of minimum and/or maximum temperature have been conducted. These experiments provide evidence that forecasters can make such forecasts in a reliable and skillful manner. In terms of reliability, observed temperatures in two experiments (Murphy and Winkler 1974 and Winkler and Murphy 1979a) fell within the 50%-credible intervals 51.9% of the time and within the 75%-credible intervals 78.0% of the time. In terms of skill, the probability distributions discussed by Sanders (1973) improved upon climatology by about 55%, 35%, 20%, and 10% for one-, two-, three-, and four-day forecasts, respectively. It should be noted that no objective probability forecasts were available as guidance for the forecasters who participated in these experimental programs. In other experiments, Bosart (1975) considered the probability that the minimum temperature would be below the climatological mean, and Gregg (1977) forecast the probability of the temperature falling below 28°F (a level below which frost damage could occur to fruit trees).

Subjective probability forecasts of other variables, including tornadoes (Murphy and Winkler 1982) and wind speed and visibility (Daan and Murphy 1982), were formulated on an experimental basis. The results were encouraging but exhibited a moderate amount of overforecasting. Recently feedback and experience have been shown to lead to reductions in the amount of overforecasting (Murphy and Daan 1984b).

Occasionally subjective forecasts for the same event are made by more than one forecaster. In such cases, each individual forecaster can be evaluated, but a consensus forecast can also be found by combining the forecasts in some manner. For example, many of the results reported by Sanders (1979) and Bosart (1975) involved consensus forecasts formulated by simply averaging the individual forecasts. Winkler, Murphy, and Katz (1977) tried various weighting schemes in a study involving precipitation probability forecasts. In all of these experiments, the consensus schemes performed better than almost all of the individual forecasters.

3.3 Evaluation of Probability Forecasts

During the last 10 years, work in the area of the evaluation of probability forecasts in meteorology has focused primarily on (a) properties and partitions of scoring rules, including the formulation of new measures, and (b) various topics and issues related to the value of such forecasts. With regard to the former, Murphy (1972a,b) generalized the work of Sanders (1958,1963) and obtained both scalar and vector partitions of the Brier score. In addition, a new vector partition of the Brier score was defined (Murphy 1973) that consisted of three terms—namely, a measure of the uncertainty (or variance) of the observations, the usual measure of reliability, and a new measure of resolution. This partition provided the basis for the definition of a new scoring rule, the "sample skill score" (Murphy 1974).

In the area of the value of forecasts, studies were conducted in which the value of probability forecasts was compared with the value of other types of forecasts (e.g., climatological and/or categorical forecasts) in simple decision-making situations from both ex ante and ex post viewpoints (Murphy 1977 and Winkler and Murphy 1979b). These studies revealed that the value of reliable probabilistic forecasts exceeds the value of other types of (imperfect) forecasts. Moreover, studies with real sets of probability forecasts indicated that these (ex post) value relationships are robust, in the sense that they also generally hold even for moderately unreliable sets of probability forecasts (see Murphy 1977). For a recent review of the evaluation of probability forecasts in meteorology, see Murphy and Daan (1984a).

3.4 Operational Developments

As indicated in Section 2, the Weather Bureau, after several years of experimentation, initiated a nationwide PoP forecasting program in 1965. Extensive evaluations of these subjective forecasts were conducted, and the results of such studies indicated that PoP forecasts were both reliable and skillful (e.g., Cummings 1974, Sadowski and Cobb 1974, Hughes 1980, and Murphy 1981). Moreover, although the nationwide PoP forecasting program initially encountered some resistance on the part of both the forecasters and the public, it is now generally agreed that these probabilities are an important and integral part of the NWS's public weather forecasts.

The NWS produces about 600,000 objective guidance forecasts daily (including categorical as well as probability forecasts), and these forecasts are disseminated by various communications systems throughout the United States (Glahn 1984). Such guidance forecasts, however, are seldom issued directly to the general public.

The NWS program of subjective PoP forecasts and objective guidance forecasts represents the largest and oldest program involving probability forecasts on an operational basis. In addition, probability forecasts related to the location of hurricane landfall have been provided by the NWS to local government officials and the general public since 1983 (Sheets 1984). The NWS also now routinely expresses its monthly and seasonal forecasts of temperature and precipitation anomalies in probabilistic terms (Gilman 1983). Moreover, some probability forecasts have been issued on an operational basis by other organizations. For example, the Atmospheric Environment Service in Canada initiated a nationwide PoP forecasting program in July 1982, and the Royal Netherlands Meteorological Institute has issued probabilistic dry-period forecasts on a limited basis for several years (Daan and Murphy 1982).

4. CURRENT ISSUES AND FUTURE PROSPECTS

4.1 Objective Probability Forecasting

In the MOS approach, the output of numerical models (and other data) is related statistically, usually via linear regression analysis, to the weather variables or events of

primary interest. These models consist of highly nonlinear sets of differential equations that are solved numerically over a global, hemispheric, or regional grid of points. The basic inputs to these models are analyses of observations of meteorological variables at the earth's surface and at various levels in the atmosphere. Thus this forecasting system consists of three components: analyses of basic observations, numerical models, and statistical procedures.

The first step in this approach to weather forecasting involves the basic observations and their analysis or interpolation. These observations are available only at a widely separated set of points in the three-dimensional space representing the atmosphere (moreover, the density of observations varies greatly from one region to another, with the average distance between observations being much greater over the oceans and the Southern Hemisphere than over North America and Europe). Obviously a finer grid would provide more information on the current (or initial) state of the atmosphere and produce a better description of the smaller scales of motion that influence local weather conditions. Although it is unlikely that the number of observations will increase significantly in the near future, it may be possible to increase the density of available observations through more effective use of satellite and radar data. Regardless of the density of observational data, the observations themselves are subject to error (or may not be representative of local conditions). Reductions in uncertainty associated with such errors would lead to improved forecasts. A related topic is the problem of interpolating or analyzing the observational data to a regular grid of points. This problem, generally referred to as *objective analysis*, was discussed by Schlatter (1975), Bergman (1979), and Wahba and Wendelberger (1980).

The next step in this procedure involves the use of a numerical model of the atmosphere to predict, conditional on the initial state of the atmosphere, the state of the atmosphere at some future time. Of course, models are only approximations to the true physical and dynamical processes operating in the atmosphere. As noted in Section 3, different models have been used in conjunction with the MOS approach. Moreover, some changes in these models have led to noticeable improvements in the quality of MOS forecasts (e.g., see Charba and Klein 1980), and additional work on such models undoubtedly will lead to further improvements. A major problem of current interest is the development and implementation of smaller-scale forecasting models that can be used on an operational basis (Pielke 1977).

The last step in the so-called MOS approach, relating the output of the numerical models to the event or variable of primary concern, usually involves linear regression with ordinary least-squares estimates and a forward stepwise variable selection process, although a logit model is used in some applications. Perhaps the development of additional predictor variables and the use of other variable selection methods, nonlinear regression models, some sort of weighted least-squares estimates, or some time-series models would lead to improved forecasts.

The stochastic-dynamic approach to probability forecasting attempts to recognize and model some of the uncertainties noted above. In relation to MOS, however, the stochastic-dynamic approach poses many more difficulties of a computational nature. Dynamic models of the atmosphere that attempt to include initial uncertainty are difficult to work with and impose a severe computational burden. Nevertheless, the stochastic-dynamic approach is conceptually appealing; and if the problems of the specification of initial uncertainty and computational burden can be dealt with to make the approach feasible to apply, then it may produce very useful forecasts. Another possibility is to employ stochastic-dynamic predictions in place of the numerical model predictions that are currently used in MOS.

4.2 Subjective Probability Forecasting

Subjective probability forecasts are generally reliable and skillful, as indicated in Section 3. However, considerable room for improvement still exists. The question is how to improve the *process* used by the forecaster in preparing subjective probability forecasts. Aspects of this issue include more effective use of the many information sources that are available to the forecaster; motivation to encourage forecasters to improve their performance; provision of formal procedures to assist forecasters in quantifying their uncertainty in terms of probabilities; and quick and extensive feedback concerning performance. Training and motivating forecasters to improve their probability forecasts is a promising approach, and some modest efforts have been made in this direction (e.g., Air Weather Service 1978 and Hughes 1980). Even relatively simple steps such as training forecasters to understand the relationships among certain probabilities (e.g., 6-hour precipitation probabilities vs. 12-hour precipitation probabilities, or point probabilities vs. area probabilities of precipitation) should be helpful.

To improve the process of preparing subjective probability forecasts, it should be useful to model the process. This activity may help to isolate certain steps that tend to be beneficial and certain other steps that tend to be detrimental (perhaps because they may mislead the forecaster, for instance). Such an approach may help to identify what differentiates a good forecaster from a mediocre forecaster. A recent step in this direction was taken by Allen (1982).

Forecasts can always be adjusted. If we are aware of certain biases of a particular forecaster, we might choose to adjust his or her forecasts accordingly (e.g., see Roberts 1968). Of course, with feedback these biases might be reduced via a self-calibration approach on the part of the forecaster. It also often seems worthwhile to obtain forecasts from different forecasters (or objective methods) and to aggregate these forecasts to form a single

forecast. Further work on aggregation procedures, particularly those that incorporate aspects such as dependence among forecasts, is of interest.

4.3 Evaluation of Probability Forecasts

Ideally comparisons of different forecasting methods or different forecasters should involve sets of forecasts that are matched in the sense of being prepared at the same times for the same forecast periods at the same locations. For instance, scores obtained by two forecasters at different locations are not directly comparable. Matched sets of forecasts are often unavailable, and methods to compare unmatched sets of forecasts are needed. Such comparisons might involve scoring rules with adjustments made for differences in the forecasting situations, or they might involve a completely different type of evaluation measure. A related topic is the comparison of different forecasting situations (e.g., PoP forecasts at Los Angeles vs. PoP forecasts at Chicago) in terms of some concept of "difficulty."

In the evaluation of probability forecasts, the forecasts should be viewed as information and the evaluation scheme should measure, in some manner, how informative the forecast is to the recipient. For instance, if the forecast consists of probabilities for multiple categories for an ordered variable (e.g., temperature or amount of precipitation), the ranked probability score seems preferable to the Brier score because the former considers how close the forecast was to what actually happened (e.g., if the temperature is in the 10°–15°F range, which was assigned a probability of .20, a forecast assigning probabilities of .10 and .70 to 0°–5°F and 5°–10°F is closer than a forecast assigning probabilities of .70 and .10 to these respective intervals). It may also be more informative to use different definitions of events in different situations. For example, perhaps it is helpful to count traces (precipitation totaling less than .01 in. in a forecast period) as precipitation in some locations.

Viewing the forecasts as information leads naturally to a consideration of the impact of the forecasts on users. More work on user-oriented evaluation schemes could include the relation of scoring rules (or other evaluation measures) to weather-related decision-making problems and the investigation of the value of information or different forecasts for such problems. Of course many decision-making problems are specific to certain subclasses of users, and it would be desirable to have evaluation schemes that are as general as possible in the sense of relating to a wide range of decision-making problems.

4.4 Operational Aspects

An issue of interest concerns the inclusion of probability forecasts in weather forecasts that are disseminated to the general public or to specific users. Although objective and subjective probability forecasts have been formulated for a variety of weather-related events and variables, until recently only precipitation probabilities were issued to the general public. Occasionally other probabilities may have been provided to specific users (e.g., probabilities of frost to orchardists), but this practice has not been widespread. The empirical evidence regarding probability forecasts of temperature that have been formulated on an experimental basis suggests that such forecasts are reliable and accurate. Thus probability forecasts of temperature are good candidates for inclusion in weather forecasts (see Murphy and Winkler 1979), both for specific users and for the general public. Hughes (1980) recommended that probabilities of various precipitation amounts be included in public weather forecasts, and Pielke (1977) suggested the dissemination of several types of probability forecasts.

One aspect of the nationwide PoP program that has been lacking is the education of the public and specific users (and perhaps even some forecasters) concerning the interpretation of such forecasts. Training of recipients of PoP forecasts should include seemingly simple matters, such as the exact definition of a precipitation probability. Some individuals interpret a precipitation probability in terms of an area probability, an expected proportion of time that precipitation will occur, or some other erroneous interpretation. Furthermore, it appears that such misinterpretations tend to be caused by misunderstanding of the precipitation event rather than by misunderstanding of the notion of probability (Murphy et al. 1980). Although some modest attempts have been made to educate users with respect to the purposes, meaning, and use of probability forecasts (e.g., Bennett et al. 1969), this sort of educational program needs to be undertaken on a much larger scale.

5. PROBABILITY FORECASTING: METEOROLOGY VIS-A-VIS OTHER FIELDS

Interest in probability forecasting in fields other than meteorology has increased considerably in recent years. For example, probability forecasts are used in portfolio analysis by some large banks, formal recognition of uncertainty in terms of probabilities is becoming more common in economic forecasting, and probabilistic diagnosis is receiving increasing attention in the medical field. Probabilities represent an important input in applications of decision analysis, and the burgeoning area called risk analysis depends heavily on probability forecasts, especially for low-probability events such as an earthquake or a serious accident at a nuclear power plant. Compared with the empirical evidence regarding probability forecasting in meteorology, however, relatively little information is available concerning probability forecasts in other fields. In some cases the forecasts and any subsequent evaluations are considered to be proprietary information (e.g., portfolio analysis and decision analysis). In other cases, the time horizons involved do not permit quick evaluation (e.g., risk analysis). Perhaps the most common circumstance is the consideration of probability forecasts for "one-shot" situations, in which the small number of forecasts prevents effective evaluation (e.g.,

decision analysis). To the best of our knowledge, meteorology is still the only field providing a large number of probability forecasts and evaluations in the public domain on a regular, continuing basis.

Of particular interest is the high quality of probability forecasts in meteorology. In terms of objective forecasts, some work involving economic and demographic series suggested that simple, naive forecasting schemes such as persistence tend to perform as well as, or even better than, more sophisticated forecasting models (e.g., Makridakis and Hibon 1979 and Makridakis et al. 1982). In contrast, techniques developed and used in meteorology, such as MOS, provide forecasts that are considerably better than simple procedures such as climatology (see Table 3) or persistence. In terms of subjective forecasts, psychological studies cast doubt on the abilities of human forecasters to formulate reliable and accurate probabilities (e.g., Hogarth 1975 and Lichtenstein, Fischhoff, and Phillips 1977). Yet it appears that weather forecasters can formulate probability forecasts that are reliable and skillful (see Table 3 and Figures 1 and 2).

Perhaps the high quality of probability forecasts in meteorology can be explained in part by the fact that weather forecasting rests on a sound physical and dynamical basis. Moreover, forecasting has always been considered a major function in meteorology. As a result, much time and effort have been devoted to developing, evaluating, and improving forecasts. Forecasts are made frequently, event feedback is obtained quickly, and evaluations are provided routinely. Thus expertise in forecasting, including probability forecasting, has been developed. Furthermore, similar meteorological situations recur from time to time, and past data, therefore, may be quite useful (as opposed to applications in other fields, where the current situation may differ sufficiently to reduce the value of information about previous situations). Finally, the eventual outcomes are independent of the forecasts, which is not always the case in other fields (e.g., in economics, forecasts may influence decisions, which may in turn affect outcomes). These factors may all contribute to the high quality of probability forecasts in meteorology, and other fields may share some but not all of these factors. Nonetheless, it would seem that a careful study of probability forecasting in meteorology could be of some benefit to those interested in probability forecasting in other fields (Murphy and Brown in press). In turn, we believe that future improvements in probability forecasting in meteorology will draw on work from various fields, such as statistics, psychology, and operations research, and interdisciplinary work will be valuable in this regard.

[*Received August 1983. Revised March 1984.*]

REFERENCES

AIR WEATHER SERVICE (1978), "Probability Forecasting: A Guide for Forecasters and Staff Weather Officers," Pamphlet No. 105-51, U.S. Air Force, Headquarters Air Weather Service, Scott Air Force Base, Il.

ALLEN, G. (1982), "Probability Judgement in Weather Forecasting," in *Preprints of the Ninth Conference on Weather Forecasting and Analysis*, Boston: American Meteorological Society, 1–6.

ANGSTROM, A. (1922), "On the Effectivity of Weather Warnings," *Nordisk Statistisk Tidskrift*, 1, 394–408.

BENNETT, E.S., NEWMAN, J.E., DAILEY, P.W., Jr., PETZHOLD, C.E., and SMITH, L.H. (1969), "Precipitation Probabilities in Weather Forecasts," Publication No. AY-174, Purdue University, Cooperative Extension Service.

BERGMAN, K.H. (1979), "Multivariate Analysis of Temperatures and Winds Using Optimum Interpolation," *Monthly Weather Review*, 107, 1423–1444.

BERMOWITZ, R.J., and ZURNDORFER, E.A. (1979), "Automated Guidance for Predicting Quantitative Precipitation," *Monthly Weather Review*, 107, 122–128.

BESSON, L. (1905), "Essai de Prevision Methodique de Temps" [Attempts at Methodical Weather Forecasting], *Observatoire Municipal de Montsouris, Annales*, 6, 473–495.

BILHAM, E.G. (1922), "A Problem in Economics," *Nature*, 109, 341–342.

BLAIR, T.A. (1921), "Local Forecast Studies—Summer Rainfall," *Monthly Weather Review*, 49, 183–190.

——— (1924), "Local Forecast Studies—Winter Precipitation," *Monthly Weather Review*, 52, 79–85.

BOCCHIERI, J.R. (1979), "A New Operational System for Forecasting Precipitation Type," *Monthly Weather Review*, 107, 637–649.

BOCCHIERI, J.R., CRISCI, R.L., GLAHN, H.R., LEWIS, F., and GLOBOKAR, F.T. (1974), "Recent Developments in Automated Prediction of Ceiling and Visibility," *Journal of Applied Meteorology*, 13, 277–288.

BORGMAN, L.E. (1960), "Weather Forecast Profitability From a Client's Viewpoint," *Bulletin of the American Meteorological Society*, 41, 347–356.

BOSART, L.F. (1975), "SUNYA Experimental Results in Forecasting Daily Temperature and Precipitation," *Monthly Weather Review*, 103, 1013–1020.

BRELSFORD, W.M., and JONES, R.H. (1967), "Estimating Probabilities," *Monthly Weather Review*, 95, 750–756.

BRIER, G.W. (1944), "Verification of a Forecaster's Confidence and the Use of Probability Statements in Weather Forecasting," Research Paper No. 16, U.S. Weather Bureau, Washington, D.C.

——— (1946), "A Study of Quantitative Precipitation Forecasting in the T.V.A. Basin," Research Paper No. 26, U.S. Weather Bureau, Washington, D.C.

——— (1950), "Verification of Forecasts Expressed in Terms of Probability," *Monthly Weather Review*, 78, 1–3.

BRIER, G.W., and ALLEN, R.A. (1951), "Verification of Weather Forecasts," in *Compendium of Meteorology*, ed. T.F. Malone, Boston: American Meteorological Society, 841–848.

CARTER, G.M., and GLAHN, H.R. (1976), "Objective Prediction of Cloud Amounts Based on Model Output Statistics," *Monthly Weather Review*, 104, 1565–1572.

CARTER, G.M., BOCCHIERI, J.R., DALLAVALLE, J.P., HOLLENBAUGH, G.H., MAGLARAS, G.J., and SCHWARTZ, B.E. (1982), "Comparative Verification of Guidance and Local Aviation/Public Weather Forecasts—No. 12 (April 1981–September 1981)," TDL Office Note 82-8, National Weather Service, National Oceanic and Atmospheric Administration, Silver Spring, Md.

CHARBA, J.P. (1979), "Two to Six Hour Severe Local Storm Probabilities: An Operational Forecasting System," *Monthly Weather Review*, 107, 268–282.

CHARBA, J.P., and KLEIN, W.H. (1980), "Skill in Precipitation Forecasting in the National Weather Service," *Bulletin of the American Meteorological Society*, 61, 1546–1555.

COOKE, W.E. (1906a), "Forecasts and Verifications in Western Australia," *Monthly Weather Review*, 34, 23–24.

——— (1906b), "Weighting Forecasts," *Monthly Weather Review*, 34, 274–275.

CUMMINGS, A.D. (1974), "Some Aspects of PoP Forecast Performance by Southern Region WSFOs—Revisited," Technical Memorandum No. SR-77, National Weather Service, National Oceanic and Atmospheric Administration, Fort Worth, Tex.

DAAN, H., and MURPHY, A.H. (1982), "Subjective Probability Forecasting in the Netherlands: Some Operational and Experimental Results," *Meteorologische Rundschau*, 35, 99–112.

DEROUIN, R.G., and COBB, G.F. (1970), "Weather Bureau Forecast Verification Scores 1968–69 and Some Performance Trends From

1966," Technical Memorandum No. WBTM FCST-15, Weather Bureau, Environmental Science Services Administration, Silver Spring, Md.

EPSTEIN, E.S. (1962), "A Bayesian Approach to Decision Making in Applied Meteorology," *Journal of Applied Meteorology*, 1, 169–177.

——— (1969a), "Stochastic Dynamic Prediction," *Tellus*, 21, 739–759.

——— (1969b), "A Scoring System for Probabilities of Ranked Categories," *Journal of Applied Meteorology*, 8, 985–987.

FIX, E., and HODGES, J.L., Jr. (1951), "Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties," Report No. 4, U.S. Air Force School of Aviation Medicine, Randolph Field, Tex.

FLEMING, R.J. (1971a), "On Stochastic Dynamic Prediction: I. The Energetics of Uncertainty and the Question of Closure," *Monthly Weather Review*, 99, 851–872.

——— (1971b), "On Stochastic Dynamic Prediction: II. Predictability and Utility," *Monthly Weather Review*, 99, 927–938.

GILMAN, D.L. (1983), "Predicting the Weather for the Long Term," *Weatherwise*, 36, 290–297.

GLAHN, H.R. (1964a), "The Use of Decision Theory in Meteorology," *Monthly Weather Review*, 92, 383–388.

——— (1964b), "An Application of Adaptive Logic to Meteorological Prediction," *Journal of Applied Meteorology*, 3, 718–725.

——— (1965), "Objective Weather Forecasting by Statistical Methods," *The Statistician*, 15, 111–142.

——— (1976), "Numerical—Statistical Forecasting in the U.S. National Weather Service," in *Weather Forecasting and Weather Forecasts: Models, Systems, and Users* (Vol. 2), Boulder, Colo.: National Center for Atmospheric Research, 448–522.

——— (1984), "Statistical Weather Forecasting," in *Probability, Statistics, and Decision Making in the Atmospheric Sciences*, eds. A.H. Murphy and R.W. Katz, Boulder, Colo.: Westview Press.

GLAHN, H.R., and LOWRY, D.A. (1969), "An Operational Method for Objectively Forecasting Probability of Precipitation," Technical Memorandum No. WBTM TDL 27, Weather Bureau, Environmental Science Services Administration, Silver Spring, Md.

——— (1972), "The Use of Model Output Statistics (MOS) in Objective Weather Forecasting," *Journal of Applied Meteorology*, 11, 1203–1211.

GLAHN, H.R., LOWRY, D.A., HOLLENBAUGH, G.W., and ANNETT, J.R. (1971), "Statistics of Numerical Prediction Models," in *Preprints of the International Symposium on Probability and Statistics in the Atmospheric Sciences*, Boston: American Meteorological Society, 15–20.

GREGG, G.T. (1977), "Probability Forecasts of a Temperature Event," *National Weather Digest*, 2 (2), 33–34.

GRINGORTEN, I.I. (1951), "The Verification and Scoring of Weather Forecasts," *Journal of the American Statistical Association*, 46, 279–296.

——— (1955), "Methods of Objective Weather Forecasting," in *Advances in Geophysics* (Vol. 2), New York: Academic Press, 57–92.

——— (1959), "Probability Estimates of the Weather in Relation to Operational Decisions," *Journal of Meteorology*, 16, 663–671.

——— (1965), "A Measure of Skill in Forecasting a Continuous Variable," *Journal of Applied Meteorology*, 4, 47–53.

HALLENBECK, C. (1920), "Forecasting Precipitation in Percentages of Probability," *Monthly Weather Review*, 48, 645–647.

HOGARTH, R.M. (1975), "Cognitive Processes and the Assessment of Subjective Probability Distributions," *Journal of the American Statistical Association*, 70, 271–294.

HOLLOWAY, J.L., and WOODBURY, M.A. (1955), "Application of Information Theory and Discriminant Function Analysis to Weather Forecasting and Forecast Verification," Technical Report No. 1, University of Pennsylvania, Philadelphia.

HUGHES, L.A. (1965), "On the Probability Forecasting of the Occurrence of Precipitation," Technical Note No. 20-CR-3, Weather Bureau, Environmental Science Services Administration, Kansas City, Mo.

——— (1968), "Probability Verification Results (24 Months)," Technical Memorandum No. WBTM CR-19, Weather Bureau, Environmental Science Services Administration, Kansas City, Mo.

——— (1980), "Probability Forecasting: Reasons, Procedures, Problems," Technical Memorandum No. NWS FCST-24, National Weather Service, National Oceanic and Atmospheric Administration, Silver Spring, Md.

JONES, R.H. (1968), "A Nonlinear Model for Estimating Probabilities of K Events," *Monthly Weather Review*, 96, 383–384.

KLEIN, W.H. (1971), "Computer Prediction of Precipitation Probability in the United States," *Journal of Applied Meteorology*, 10, 903–915.

KLEIN, W.H., and GLAHN, H.R. (1974), "Forecasting Local Weather by Means of Model Output Statistics," *Bulletin of the American Meteorological Society*, 55, 1217–1227.

LAVE, L. (1963), "The Value of Better Weather Information to the Raisin Industry," *Econometrica*, 31, 151–164.

LEIGHT, W.G. (1953), "The Use of Probability Statements in Extended Forecasting," *Monthly Weather Review*, 81, 349–356.

LEITH, C.E. (1978), "Objective Methods for Weather Prediction," *Annual Review of Fluid Mechanics*, 10, 107–128.

——— (1980), "Stochastic-Dynamic Prediction," in *Preprints of WMO Symposium on Probabilistic and Statistical Methods in Weather Forecasting*, Geneva: World Meteorological Organization, 39–43.

LICHTENSTEIN, S., FISCHHOFF, B., and PHILLIPS, L.D. (1977), "Calibration of Probabilities: The State of the Art," in *Decision Making and Change in Human Affairs*, eds. H. Jungermann and G. de Zeeuw, Dordrecht, Holland: D. Reidel, 45–56.

LUND, I.A. (1955), "Estimating the Probability of a Future Event From Dichotomously Classified Predictors," *Bulletin of the American Meteorological Society*, 36, 325–328.

MAKRIDAKIS, S., ANDERSEN, A., CARBONE, R., FILDES, R., HIBON, M., LEWANDOWSKI, R., NEWTON, J., PARZEN, E., and WINKLER, R.L. (1982), "The Accuracy of Extrapolation (Time Series) Methods," *Journal of Forecasting*, 1, 111–153.

MAKRIDAKIS, S., and HIBON, M. (1979), "Accuracy of Forecasting: An Empirical Investigation," *Journal of the Royal Statistical Society*, Ser. A, 142, 97–145.

MILLER, R.G. (1962), "Statistical Prediction by Discriminant Analysis," *Meteorological Monographs*, 4 (25), 1–54.

——— (1964), "Regression Estimation of Event Probabilities," Technical Report No. 1, Travelers Research Center, Inc., Hartford, Conn.

——— (1981), "GEM: A Statistical Weather Forecasting Procedure," Technical Report No. NWS-28, National Weather Service, National Oceanic and Atmospheric Administration, Silver Spring, Md.

MURPHY, A.H. (1966), "A Note on the Utility of Probabilistic Predictions and the Probability Score in the Cost–Loss Ratio Decision Situation," *Journal of Applied Meteorology*, 5, 534–537.

——— (1972a), "Scalar and Vector Partitions of the Probability Score: Part I. Two-State Situation," *Journal of Applied Meteorology*, 11, 273–282.

——— (1972b), "Scalar and Vector Partitions of the Probability Score: Part II. *N*-State Situation," *Journal of Applied Meteorology*, 11, 1183–1192.

——— (1973), "A New Vector Partition of the Probability Score," *Journal of Applied Meteorology*, 12, 595–600.

——— (1974), "A Sample Skill Score for Probability Forecasts," *Monthly Weather Review*, 102, 48–55.

——— (1977), "The Value of Climatological, Categorical and Probabilistic Forecasts in the Cost–Loss Ratio Situation," *Monthly Weather Review*, 105, 803–816.

——— (1981), "Subjective Quantification of Uncertainty in Weather Forecasts in the United States," *Meteorologische Rundschau*, 34, 65–77.

——— (1984), "Probabilistic Weather Forecasting," in *Probability, Statistics, and Decision Making in the Atmospheric Sciences*, eds. A.H. Murphy and R.W. Katz, Boulder, Colo.: Westview Press.

MURPHY, A.H., and BROWN, B.G. (in press), "A Comparative Evaluation of Objective and Subjective Weather Forecasts in the United States," *Journal of Forecasting*, 3.

MURPHY, A.H., and DAAN, H. (1984a), "Forecast Evaluation," in *Probability, Statistics, and Decision Making in the Atmospheric Sciences*, eds. A.H. Murphy and R. W. Katz, Boulder, Colo.: Westview Press.

——— (1984b), "Impacts of Feedback and Experience on the Quality of Subjective Probability Forecasts: Comparison of Results From the First and Second Years of the Zierikzee Experiment," *Monthly Weather Review*, 112, 413–423.

MURPHY, A.H., and EPSTEIN, E.S. (1967), "A Note on Probability Forecasts and 'Hedging'," *Journal of Applied Meteorology*, 6, 1002–1004.

MURPHY, A.H., HSU, W.-R., and WINKLER, R.L. (1982), "Subjective Probabilistic Quantitative Precipitation Forecasts: Some Experimental Results," in *Preprints of the Ninth Conference on Weather Forecasting and Analysis*, Boston: American Meteorological Society, 94–100.

MURPHY, A.H., LICHTENSTEIN, S., FISCHHOFF, B., and

WINKLER, R.L. (1980), "Misinterpretations of Precipitation Probability Forecasts," *Bulletin of the American Meteorological Society*, 61, 695–701.

MURPHY, A.H., and WINKLER, R.L. (1974), "Credible Interval Temperature Forecasting: Some Experimental Results," *Monthly Weather Review*, 102, 784–794.

——— (1977), "Experimental Point and Area Precipitation Probability Forecasts for a Forecast Area With Significant Local Effects," *Atmosphere*, 15, 61–78.

——— (1979), "Probabilistic Temperature Forecasts: The Case for an Operational Program," *Bulletin of the American Meteorological Society*, 60, 12–19.

——— (1982), "Subjective Probabilistic Tornado Forecasts: Some Experimental Results," *Monthly Weather Review*, 110, 1288–1297.

NELSON, R.R., and WINTER, S.G. (1964), "A Case Study in the Economics of Information and Coordination: The Weather Forecasting System," *Quarterly Journal of Economics*, 78, 420–441.

PIELKE, R.A. (1977), "An Overview of Recent Work in Weather Forecasting and Suggestions for Future Work," *Bulletin of the American Meteorological Society*, 58, 506–519.

REAP, R.M., and FOSTER, D.S. (1979), "Automated 12–36 Hour Probability Forecasts of Thunderstorms and Severe Local Storms," *Journal of Applied Meteorology*, 18, 1304–1315.

ROBERTS, H.V. (1968), "On the Meaning of the Probability of Rain," in *Proceedings of the First National Conference on Statistical Meteorology*, Boston: American Meteorological Society, 133–141.

ROLF, B. (1917), "Probabilite et Prognostics des Pluies d'Ete" [Probability and Forecasting of Summer Rains], unpublished Ph.D. thesis, University of Uppsala, Sweden.

ROOT, H.E. (1962), "Probability Statements in Weather Forecasting," *Journal of Applied Meteorology*, 1, 163–168.

SADOWSKI, A.F., and COBB, G.F. (1974), "National Weather Service April 1972 to March 1973 Public Forecast Verification Summary," Technical Memorandum No. NWS FCST-21, National Weather Service, National Oceanic and Atmospheric Administration, Silver Spring, Md.

SANDERS, F. (1958), "The Evaluation of Subjective Probability Forecasts," Scientific Report No. 5, Massachusetts Institute of Technology, Dept. of Meteorology.

——— (1963), "On Subjective Probability Forecasting," *Journal of Applied Meteorology*, 2, 191–201.

——— (1967), "The Verification of Probability Forecasts," *Journal of Applied Meteorology*, 6, 756–761.

——— (1973), "Skill in Forecasting Daily Temperature and Precipitation: Some Experimental Results," *Bulletin of the American Meteorological Society*, 54, 1171–1179.

——— (1979), "Trends in Skill of Daily Forecasts of Temperature and Precipitation, 1966–78," *Bulletin of the American Meteorological Society*, 60, 763–769.

SCHLATTER, T.W. (1975), "Some Experiments With a Multivariate Statistical Objective Analysis Scheme," *Monthly Weather Review*, 103, 246–257.

SCHROEDER, M.J. (1954), "Verification of 'Probability' Fire-Weather Forecasts," *Monthly Weather Review*, 82, 257–260.

SCHWARTZ, B.E., BOCCHIERI, J.R., CARTER, G.M., DALLAVALLE, J.P., HOLLENBAUGH, G.H., MAGLARAS, G.J., and VERCELLI, D.J. (1981), "Comparative Verification of Guidance and Local Aviation/Public Weather Forecasts—No. 11 (October 1980–March 1981)," TDL Office Note No. 81-10, National Weather Service, National Oceanic and Atmospheric Administration, Silver Spring, Md.

SHEETS, R.C. (1984), "The National Weather Service Hurricane Probability Program," in *Postprints of the Fifteenth Conference on Hurricanes and Tropical Meteorology*, Boston: American Meteorological Society, J12–J19.

THOMPSON, J.C. (1950), "A Numerical Method for Forecasting Rainfall in the Los Angeles Area," *Monthly Weather Review*, 78, 113–124.

——— (1952), "On the Operational Deficiencies in Categorical Weather Forecasts," *Bulletin of the American Meteorological Society*, 88, 223–226.

——— (1962), "Economic Gains From Scientific Advances and Operational Improvements in Meteorological Prediction," *Journal of Applied Meteorology*, 1, 13–17.

THOMPSON, J.C., and BRIER, G.W. (1955), "The Economic Utility of Weather Forecasts," *Monthly Weather Review*, 83, 249–254.

U.S. ARMY SIGNAL CORPS (1919), "The Signal Corps Meteorological Service, A.E.F. (Excerpts From the Annual Report of the Chief Signal Officer)," *Monthly Weather Review*, 47, 870–876.

VON MYRBACH, O. (1913), "Verlasslichkeit der Wetterprognosen (Ein Vorschlag)" [Reliability of Weather Forecasts (A Proposal)], *Meteorologische Zeitschrift*, 30, 496–497.

WAHBA, G., and WENDELBERGER, J. (1980), "Some New Mathematical Methods for Variational Objective Analysis Using Splines and Cross Validation," *Monthly Weather Review*, 108, 1122–1143.

WILLIAMS, P. (1951), "The Use of Confidence Factors in Forecasting," *Bulletin of the American Meteorological Society*, 32, 279–281.

WINKLER, R.L., and MURPHY, A.H. (1968a), "Evaluation of Subjective Precipitation Probability Forecasts," in *Proceedings of the First National Conference on Statistical Meteorology*, Boston: American Meteorological Society, 148–157.

——— (1968b), "'Good' Probability Assessors," *Journal of Applied Meteorology*, 7, 751–758.

——— (1976), "Point and Area Precipitation Probability Forecasts: Some Experimental Results," *Monthly Weather Review*, 104, 86–95.

——— (1979a), "The Use of Probabilities in Forecasts of Maximum and Minimum Temperatures," *Meteorological Magazine*, 108, 317–329.

——— (1979b), "The Value of Weather Forecasts in the Cost–Loss Ratio Situation: An Ex Ante Approach," in *Preprints of the Sixth Conference on Probability and Statistics in Atmospheric Sciences*, Boston: American Meteorological Society, 134–138.

WINKLER, R.L., MURPHY, A.H., and KATZ, R.W. (1977), "The Consensus of Subjective Probability Forecasts: Are Two, Three, . . . Heads Better Than One?," in *Preprints of the Fifth Conference on Probability and Statistics in Atmospheric Sciences*, Boston: American Meteorological Society, 57–62.